2.8 The Derivative of a Function Between Normed Vector Spaces

In most cases, $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$. However, to define infinite dimensional manifolds, it is necessary to allow E and F to be infinite dimensional.

Let E and F be two *normed vector spaces*, let $A \subseteq E$ be some open subset of A, and let $a \in A$ be some element of A. Even though a is a vector, we may also call it a point.

The idea behind the derivative of the function f at a is that it is a *linear approximation* of f in a small open set around a.

The difficulty is to make sense of the quotient

$$\frac{f(a+h)-f(a)}{h}$$

where h is a vector. We circumvent this difficulty in two stages.

A first possibility is to consider the *directional derivative* with respect to a vector $u \neq 0$ in E.

We can consider the vector $f(a+tu)-f(a)$, where $t \in \mathbb{R}$ (or $t \in \mathbb{C}$). Now,

$$\frac{f(a+tu)-f(a)}{t}$$

makes sense.

The idea is that in E, the points of the form $a+tu$, for t in some small closed interval $[-\epsilon, \epsilon]$ in $\mathbb{R}$ form a line segment $[r, s]$ in A, and that the image of this line segment defines a small curve segment on $f(A)$.

This curve (segment) is defined by the map $t \mapsto f(a + tu)$, from $[r, s]$ to F, and the directional derivative $\mathrm{D}_u f(a)$ defines the direction of the tangent line at a to this curve.

Definition 2.8.1 Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, for any $u \neq 0$ in E, the *directional derivative of f at a w.r.t. the vector u*, denoted by $\mathrm{D}_u f(a)$, is the limit (if it exists)

$$\lim_{t \to 0,\, t \in U} \frac{f(a + tu) - f(a)}{t},$$

where $U = \{t \in \mathbb{R} \mid a + tu \in A,\ t \neq 0\}$
(or $U = \{t \in \mathbb{C} \mid a + tu \in A,\ t \neq 0\}$).

Since the map $t \mapsto a + tu$ is continuous, and since $A - \{a\}$ is open, the inverse image U of $A - \{a\}$ under the above map is open, and the definition of the limit in Definition 2.8.1 makes sense.

Remark: Since the notion of limit is purely topological, the existence and value of a directional derivative is independent of the choice of norms in E and F, as long as they are equivalent norms.

The directional derivative is sometimes called the *Gâteaux derivative*.

In the special case where $E = \mathbb{R}$ and $F = \mathbb{R}$, and we let $u = 1$ (i.e., the real number 1, viewed as a vector), it is immediately verified that $\mathrm{D}_1 f(a) = f'(a)$.

When $E = \mathbb{R}$ (or $E = \mathbb{C}$) and F is any normed vector space, the derivative $\mathrm{D}_1 f(a)$, also denoted by $f'(a)$, provides a suitable generalization of the notion of derivative.

However, when E has dimension ≥ 2, directional derivatives present a serious problem, which is that their definition is not sufficiently uniform.

Indeed, there is no reason to believe that the directional derivatives w.r.t. all nonzero vectors u share something in common.

As a consequence, a function can have all directional derivatives at a, and yet not be continuous at a. Two functions may have all directional derivatives in some open sets, and yet their composition may not. Thus, we introduce a more uniform notion.

Definition 2.8.2 Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, we say that f is *differentiable at $a \in A$* if there is a *linear continuous* map, $L\colon E \to F$, and a function, $\epsilon(h)$, such that

$$f(a+h) = f(a) + L(h) + \epsilon(h)\|h\|$$

for every $a + h \in A$, where

$$\lim_{h\to 0,\, h\in U} \epsilon(h) = 0,$$

with $U = \{h \in E \mid a + h \in A,\ h \neq 0\}$.

The linear map L is denoted by $\mathrm{D}f(a)$, or $\mathrm{D}f_a$, or $df(a)$, or df_a, or $f'(a)$, and it is called the *Fréchet derivative*, or *derivative*, or *total derivative*, or *total differential*, or *differential*, of f at a.

Since the map $h \mapsto a + h$ from E to E is continuous, and since A is open in E, the inverse image U of $A - \{a\}$ under the above map is open in E, and it makes sense to say that

$$\lim_{h \to 0,\, h \in U} \epsilon(h) = 0.$$

Note that for every $h \in U$, since $h \neq 0$, $\epsilon(h)$ is uniquely determined since

$$\epsilon(h) = \frac{f(a+h) - f(a) - L(h)}{\|h\|},$$

and the value $\epsilon(0)$ plays absolutely no role in this definition.

It does no harm to assume that $\epsilon(0) = 0$, and we will assume this from now on.

Remark: Since the notion of limit is purely topological, the existence and value of a derivative is independent of the choice of norms in E and F, as long as they are equivalent norms.

Note that *the continuous linear map L is unique*, if it exists. In fact, the next proposition implies this as a corollary.

The following proposition shows that our new definition is consistent with the definition of the directional derivative.

Proposition 2.8.3 *Let E and F be two normed spaces, let A be a nonempty open subset of E, and let $f\colon A \to F$ be any function. For any $a \in A$, if $\mathrm{D}f(a)$ is defined, then f is continuous at a and f has a directional derivative $\mathrm{D}_u f(a)$ for every $u \neq 0$ in E. Furthermore,*

$$\mathrm{D}_u f(a) = \mathrm{D}f(a)(u).$$

The uniqueness of L follows from Proposition 2.8.3. Also, when E is of finite dimension, it is easily shown that every linear map is continuous and this assumption is then redundant.

If $\mathrm{D}f(a)$ exists for every $a \in A$, we get a map

$$\mathrm{D}f\colon A \to \mathcal{L}(E; F),$$

called the *derivative of f on A*, and also denoted by df. Here, $\mathcal{L}(E; F)$ denotes the vector space of continuous linear maps from E to F.

When E is of finite dimension n, for any basis $(u_1, \ldots, u_n)$ of E, we can define the directional derivatives with respect to the vectors in the basis $(u_1, \ldots, u_n)$ (actually, we can also do it for an infinite basis). This way, we obtain the definition of partial derivatives, as follows.

Definition 2.8.4 For any two normed spaces E and F, if E is of finite dimension n, for every basis $(u_1, \ldots, u_n)$ for E, for every $a \in E$, for every function $f \colon E \to F$, the directional derivatives $\mathrm{D}_{u_j} f(a)$ (if they exist) are called the *partial derivatives of f with respect to the basis $(u_1, \ldots, u_n)$*. The partial derivative $\mathrm{D}_{u_j} f(a)$ is also denoted by $\partial_j f(a)$, or $\dfrac{\partial f}{\partial x_j}(a)$.

The notation $\dfrac{\partial f}{\partial x_j}(a)$ for a partial derivative, although customary and going back to Leibniz, is a "logical obscenity." Indeed, the variable x_j really has nothing to do with the formal definition. This is just another of these situations where tradition is just too hard to overthrow!

We now consider a number of standard results about derivatives.

Proposition 2.8.5 *Given two normed spaces E and F, if $f\colon E \to F$ is a constant function, then $\mathrm{D}f(a) = 0$, for every $a \in E$. If $f\colon E \to F$ is a continuous affine map, then $\mathrm{D}f(a) = \overrightarrow{f}$, for every $a \in E$, the linear map associated with f.*

Proposition 2.8.6 *Given a normed space E and a normed vector space F, for any two functions $f, g\colon E \to F$, for every $a \in E$, if $\mathrm{D}f(a)$ and $\mathrm{D}g(a)$ exist, then $\mathrm{D}(f+g)(a)$ and $\mathrm{D}(\lambda f)(a)$ exist, and*

$$\begin{aligned} \mathrm{D}(f+g)(a) &= \mathrm{D}f(a) + \mathrm{D}g(a), \\ \mathrm{D}(\lambda f)(a) &= \lambda \mathrm{D}f(a). \end{aligned}$$

Proposition 2.8.7 *Given three normed vector spaces E_1, E_2, and F, for any continuous bilinear map $f\colon E_1 \times E_2 \to F$, for every $(a, b) \in E_1 \times E_2$, $\mathrm{D}f(a, b)$ exists, and for every $u \in E_1$ and $v \in E_2$,*

$$\mathrm{D}f(a, b)(u, v) = f(u, b) + f(a, v).$$

We now state the very useful *chain rule*.

Theorem 2.8.8 *Given three normed spaces E, F, and G, let A be an open set in E, and let B an open set in F. For any functions $f \colon A \to F$ and $g \colon B \to G$, such that $f(A) \subseteq B$, for any $a \in A$, if $\mathrm{D}f(a)$ exists and $\mathrm{D}g(f(a))$ exists, then $\mathrm{D}(g \circ f)(a)$ exists, and*

$$\mathrm{D}(g \circ f)(a) = \mathrm{D}g(f(a)) \circ \mathrm{D}f(a).$$

Theorem 2.8.8 has many interesting consequences. We mention two corollaries.

Proposition 2.8.9 *Given two normed spaces E and F, let A be some open subset in E, let B be some open subset in F, let $f \colon A \to B$ be a bijection from A to B, and assume that $\mathrm{D}f$ exists on A and that $\mathrm{D}f^{-1}$ exists on B. Then, for every $a \in A$,*

$$\mathrm{D}f^{-1}(f(a)) = (\mathrm{D}f(a))^{-1}.$$

Proposition 2.8.9 has the remarkable consequence that the two vector spaces E and F have the same dimension.

In other words, a local property, the existence of a bijection f between an open set A of E and an open set B of F, such that f is differentiable on A and f^{-1} is differentiable on B, implies a global property, that the two vector spaces E and F have the same dimension.

If both E and F are of finite dimension, for any basis $(u_1, \ldots, u_n)$ of E and any basis $(v_1, \ldots, v_m)$ of F, every function $f\colon E \to F$ is determined by m functions $f_i\colon E \to \mathbb{R}$ (or $f_i\colon E \to \mathbb{C}$), where

$$f(x) = f_1(x)v_1 + \cdots + f_m(x)v_m,$$

for every $x \in E$.

Then, we get

$$\mathrm{D}f(a)(u_j) = \mathrm{D}f_1(a)(u_j)v_1 + \cdots + \mathrm{D}f_i(a)(u_j)v_i + \cdots + \mathrm{D}f_m(a)(u_j)v_m,$$

that is,

$$\mathrm{D}f(a)(u_j) = \partial_j f_1(a)v_1 + \cdots + \partial_j f_i(a)v_i + \cdots + \partial_j f_m(a)v_m.$$

Since the j-th column of the $m \times n$-matrix $J(f)(a)$ w.r.t. the bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_m)$ representing $\mathrm{D}f(a)$ is equal to the components of the vector $\mathrm{D}f(a)(u_j)$ over the basis $(v_1, \ldots, v_m)$, the linear map $\mathrm{D}f(a)$ is determined by the $m \times n$-matrix $J(f)(a) = (\partial_j f_i(a))$, or $J(f)(a) = \left(\frac{\partial f_i}{\partial x_j}(a)\right)$:

$$J(f)(a) = \begin{pmatrix} \partial_1 f_1(a) & \partial_2 f_1(a) & \ldots & \partial_n f_1(a) \\ \partial_1 f_2(a) & \partial_2 f_2(a) & \ldots & \partial_n f_2(a) \\ \vdots & \vdots & \ddots & \vdots \\ \partial_1 f_m(a) & \partial_2 f_m(a) & \ldots & \partial_n f_m(a) \end{pmatrix}$$

or

$$J(f)(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \ldots & \frac{\partial f_1}{\partial x_n}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \ldots & \frac{\partial f_2}{\partial x_n}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1}(a) & \frac{\partial f_m}{\partial x_2}(a) & \ldots & \frac{\partial f_m}{\partial x_n}(a) \end{pmatrix}$$

This matrix is called the *Jacobian matrix* of $\mathrm{D}f$ at a. When $m = n$, the determinant, $\det(J(f)(a))$, of $J(f)(a)$ is called the *Jacobian* of $\mathrm{D}f(a)$.

We know that this determinant only depends on $\mathrm{D}f(a)$, and not on specific bases. However, partial derivatives give a means for computing it.

When $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$, for any function $f\colon \mathbb{R}^n \to \mathbb{R}^m$, it is easy to compute the partial derivatives $\dfrac{\partial f_i}{\partial x_j}(a)$. We simply treat the function $f_i\colon \mathbb{R}^n \to \mathbb{R}$ as a function of its j-th argument, leaving the others fixed, and compute the derivative as the usual derivative.

Example 2.1 For example, consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$f(r, \theta) = (r\cos\theta, r\sin\theta).$$

Then, we have

$$J(f)(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

and the Jacobian (determinant) has value $\det(J(f)(r,\theta)) = r$.

In the case where $E = \mathbb{R}$ (or $E = \mathbb{C}$), for any function $f\colon \mathbb{R} \to F$ (or $f\colon \mathbb{C} \to F$), the Jacobian matrix of $\mathrm{D}f(a)$ is a column vector. In fact, this column vector is just $\mathrm{D}_1 f(a)$. Then, for every $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$),

$$\mathrm{D}f(a)(\lambda) = \lambda \mathrm{D}_1 f(a).$$

This case is sufficiently important to warrant a definition.

Definition 2.8.10 Given a function $f\colon \mathbb{R} \to F$ (or $f\colon \mathbb{C} \to F$), where F is a normed space, the vector

$$\mathrm{D}f(a)(1) = \mathrm{D}_1 f(a)$$

is called the *vector derivative or velocity vector (in the real case)* at a. We usually identify $\mathrm{D}f(a)$ with its Jacobian matrix $\mathrm{D}_1 f(a)$, which is the column vector corresponding to $\mathrm{D}_1 f(a)$.

By abuse of notation, we also let $\mathrm{D}f(a)$ denote the vector $\mathrm{D}f(a)(1) = \mathrm{D}_1 f(a)$.

When $E = \mathbb{R}$, the physical interpretation is that f defines a (parametric) curve that is the trajectory of some particle moving in $\mathbb{R}^m$ as a function of time, and the vector $\mathrm{D}_1 f(a)$ is the *velocity* of the moving particle $f(t)$ at $t = a$.

Example 2.2

1. When $A = (0, 1)$, and $F = \mathbb{R}^3$, a function $f\colon (0, 1) \to \mathbb{R}^3$ defines a (parametric) curve in $\mathbb{R}^3$. If $f = (f_1, f_2, f_3)$, its Jacobian matrix at $a \in \mathbb{R}$ is

$$J(f)(a) = \begin{pmatrix} \dfrac{\partial f_1}{\partial t}(a) \\ \dfrac{\partial f_2}{\partial t}(a) \\ \dfrac{\partial f_3}{\partial t}(a) \end{pmatrix}$$

2. When $E = \mathbb{R}^2$, and $F = \mathbb{R}^3$, a function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$ defines a parametric surface. Letting $\varphi = (f, g, h)$, its Jacobian matrix at $a \in \mathbb{R}^2$ is

$$J(\varphi)(a) = \begin{pmatrix} \dfrac{\partial f}{\partial u}(a) & \dfrac{\partial f}{\partial v}(a) \\ \dfrac{\partial g}{\partial u}(a) & \dfrac{\partial g}{\partial v}(a) \\ \dfrac{\partial h}{\partial u}(a) & \dfrac{\partial h}{\partial v}(a) \end{pmatrix}$$

3. When $E = \mathbb{R}^3$, and $F = \mathbb{R}$, for a function $f\colon \mathbb{R}^3 \to \mathbb{R}$, the Jacobian matrix at $a \in \mathbb{R}^3$ is

$$J(f)(a) = \left(\frac{\partial f}{\partial x}(a) \;\; \frac{\partial f}{\partial y}(a) \;\; \frac{\partial f}{\partial z}(a) \right).$$

More generally, when $f\colon \mathbb{R}^n \to \mathbb{R}$, the Jacobian matrix at $a \in \mathbb{R}^n$ is the row vector

$$J(f)(a) = \left(\frac{\partial f}{\partial x_1}(a) \;\cdots\; \frac{\partial f}{\partial x_n}(a) \right).$$

Its transpose is a column vector called the *gradient* of f at a, denoted by $\mathrm{grad} f(a)$ or $\nabla f(a)$. Then, given any $v \in \mathbb{R}^n$, note that

$$\mathrm{D}f(a)(v) = \frac{\partial f}{\partial x_1}(a)\, v_1 + \cdots + \frac{\partial f}{\partial x_n}(a)\, v_n = \mathrm{grad} f(a) \cdot v,$$

the scalar product of $\mathrm{grad} f(a)$ and v.

When E, F, and G have finite dimensions, $(u_1, \ldots, u_p)$ is a basis for E, $(v_1, \ldots, v_n)$ is a basis for F, and $(w_1, \ldots, w_m)$ is a basis for G, if A is an open subset of E, B is an open subset of F, for any functions $f \colon A \to F$ and $g \colon B \to G$, such that $f(A) \subseteq B$, for any $a \in A$, letting $b = f(a)$, and $h = g \circ f$, if $\mathrm{D}f(a)$ exists and $\mathrm{D}g(b)$ exists, by Theorem 2.8.8, the Jacobian matrix $J(h)(a) = J(g \circ f)(a)$ w.r.t. the bases $(u_1, \ldots, u_p)$ and $(w_1, \ldots, w_m)$ is the product of the Jacobian matrices $J(g)(b)$ w.r.t. the bases $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$, and $J(f)(a)$ w.r.t. the bases $(u_1, \ldots, u_p)$ and $(v_1, \ldots, v_n)$:

$$J(h)(a) = \begin{pmatrix} \frac{\partial g_1}{\partial y_1}(b) & \frac{\partial g_1}{\partial y_2}(b) & \ldots & \frac{\partial g_1}{\partial y_n}(b) \\ \frac{\partial g_2}{\partial y_1}(b) & \frac{\partial g_2}{\partial y_2}(b) & \ldots & \frac{\partial g_2}{\partial y_n}(b) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m}{\partial y_1}(b) & \frac{\partial g_m}{\partial y_2}(b) & \ldots & \frac{\partial g_m}{\partial y_n}(b) \end{pmatrix} \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \ldots & \frac{\partial f_1}{\partial x_p}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \ldots & \frac{\partial f_2}{\partial x_p}(a) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1}(a) & \frac{\partial f_n}{\partial x_2}(a) & \ldots & \frac{\partial f_n}{\partial x_p}(a) \end{pmatrix}$$

Thus, we have the familiar formula

$$\frac{\partial h_i}{\partial x_j}(a) = \sum_{k=1}^{k=n} \frac{\partial g_i}{\partial y_k}(b) \frac{\partial f_k}{\partial x_j}(a).$$

Given two normed spaces E and F of finite dimension, given an open subset A of E, if a function $f\colon A \to F$ is differentiable at $a \in A$, then its Jacobian matrix is well defined.

One should be warned that the converse is false. There are functions such that all the partial derivatives exist at some $a \in A$, but yet, the function is not differentiable at a, and not even continuous at a.

However, there are sufficient conditions on the partial derivatives for $\mathrm{D}f(a)$ to exist, namely, continuity of the partial derivatives.

If f is differentiable on A, then f defines a function $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$.

It turns out that the continuity of the partial derivatives on A is a necessary and sufficient condition for $\mathrm{D}f$ to exist and to be continuous on A.

Theorem 2.8.11 *Given two normed affine spaces E and F, where E is of finite dimension n and where $(u_1, \ldots, u_n)$ is a basis of E, given any open subset A of E, given any function $f\colon A \to F$, the derivative $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$ is defined and continuous on A iff every partial derivative $\partial_j f$ (or $\frac{\partial f}{\partial x_j}$) is defined and continuous on A, for all j, $1 \leq j \leq n$. As a corollary, if F is of finite dimension m, and $(v_1, \ldots, v_m)$ is a basis of F, the derivative $\mathrm{D}f\colon A \to \mathcal{L}(E; F)$ is defined and continuous on A iff every partial derivative $\partial_j f_i$ (or $\frac{\partial f_i}{\partial x_j}$) is defined and continuous on A, for all i, j, $1 \leq i \leq m$, $1 \leq j \leq n$.*

Definition 2.8.12 Given two normed affine spaces E and F, and an open subset A of E, we say that a function $f\colon A \to F$ is a *C^0-function on A* if f is continuous on A. We say that $f\colon A \to F$ is a *C^1-function on A* if $\mathrm{D}f$ exists and is continuous on A.

Let E and F be two normed affine spaces, let $U \subseteq E$ be an open subset of E and let $f \colon E \to F$ be a function such that $Df(a)$ exists for all $a \in U$. If $Df(a)$ is injective for all $a \in U$, we say that f is an *immersion* (on U) and if $Df(a)$ is surjective for all $a \in U$, we say that f is a *submersion* (on U).

When E and F are finite dimensional with $\dim(E) = n$ and $\dim(F) = m$, if $m \geq n$, then f is an immersion iff the Jacobian matrix, $J(f)(a)$, has full rank (n) for all $a \in E$ and if $n \geq m$, then then f is a submersion iff the Jacobian matrix, $J(f)(a)$, has full rank (m) for all $a \in E$.

A very important theorem is the inverse function theorem. In order for this theorem to hold for infinite dimensional spaces, it is necessary to assume that our normed spaces are complete.

Given a normed vector space, E, we say that a sequence, $(u_n)_n$, with $u_n \in E$, is a *Cauchy sequence* iff for every $\epsilon > 0$, there is some $N > 0$ so that for all $m, n \geq N$,

$$\|u_n - u_m\| < \epsilon.$$

A normed vector space, E, is *complete* iff every Cauchy sequence converges.

A complete normed vector space is also called a *Banach space*, after Stefan Banach (1892-1945).

Fortunately, $\mathbb{R}, \mathbb{C}$, and every finite dimensional (real or complex) normed vector space is complete.

A real (resp. complex) vector space, E, is a real (resp. complex) *Hilbert space* if it is complete as a normed space with the norm $\|u\| = \sqrt{\langle u, u \rangle}$ induced by its Euclidean (resp. Hermitian) inner product (of course, positive, definite).

Definition 2.8.13 Given two topological spaces E and F and an open subset A of E, we say that a function $f\colon A \to F$ is a *local homeomorphism from A to F* if for every $a \in A$, there is an open set $U \subseteq A$ containing a and an open set V containing $f(a)$ such that f is a homeomorphism from U to $V = f(U)$. If B is an open subset of F, we say that $f\colon A \to F$ is a *(global) homeomorphism from A to B* if f is a homeomorphism from A to $B = f(A)$.

If E and F are normed spaces, we say that $f\colon A \to F$ is a *local diffeomorphism from A to F* if for every $a \in A$, there is an open set $U \subseteq A$ containing a and an open set V containing $f(a)$ such that f is a bijection from U to V, f is a C^1-function on U, and f^{-1} is a C^1-function on $V = f(U)$. We say that $f\colon A \to F$ is a *(global) diffeomorphism from A to B* if f is a homeomorphism from A to $B = f(A)$, f is a C^1-function on A, and f^{-1} is a C^1-function on B.

Note that a local diffeomorphism is a local homeomorphism. Also, as a consequence of Proposition 2.8.9, if f is a diffeomorphism on A, then $\mathrm{D}f(a)$ is a bijection for every $a \in A$.

Theorem 2.8.14 *(Inverse Function Theorem) Let E and F be complete normed spaces, let A be an open subset of E, and let $f\colon A \to F$ be a C^1-function on A. The following properties hold:*

(1) For every $a \in A$, if $\mathrm{D}f(a)$ is invertible, then there exist some open subset $U \subseteq A$ containing a, and some open subset V of F containing $f(a)$, such that f is a diffeomorphism from U to $V = f(U)$. Furthermore,

$$\mathrm{D}f^{-1}(f(a)) = (\mathrm{D}f(a))^{-1}.$$

For every neighborhood N of a, the image $f(N)$ of N is a neighborhood of $f(a)$, and for every open ball $U \subseteq A$ of center a, the image $f(U)$ of U contains some open ball of center $f(a)$.

(2) If $\mathrm{D}f(a)$ *is invertible for every* $a \in A$*, then* $B = f(A)$ *is an open subset of* F*, and* f *is a local diffeomorphism from* A *to* B*. Furthermore, if* f *is injective, then* f *is a diffeomorphism from* A *to* B*.*

Part (1) of Theorem 2.8.14 is often referred to as the "(local) inverse function theorem." It plays an important role in the study of manifolds and (ordinary) differential equations.

If E and F are both of finite dimension, the case where $\mathrm{D}f(a)$ is just injective or just surjective is also important for defining manifolds, using implicit definitions.

2.9 Manifolds, Lie Groups, and Lie Algebras, the "Baby Case"

In this section we define precisely *embedded submanifolds*, *matrix Lie groups*, and their *Lie algebras*.

One of the reasons that Lie groups are nice is that they have a *differential structure*, which means that the notion of tangent space makes sense at any point of the group.

Furthermore, the tangent space at the identity happens to have some algebraic structure, that of a *Lie algebra*.

Roughly, the tangent space at the identity provides a "linearization" of the Lie group, and it turns out that many properties of a Lie group are reflected in its Lie algebra.

The challenge that we are facing is that a certain amount of basic differential geometry is required to define Lie groups and Lie algebras in full generality.

Fortunately, most of the Lie groups that we need to consider are subspaces of $\mathbb{R}^N$ for some sufficiently large N.

In fact, they are all isomorphic to subgroups of $\mathbf{GL}(N, \mathbb{R})$ for some suitable N, even $\mathbf{SE}(n)$, which is isomorphic to a subgroup of $\mathbf{SL}(n+1)$.

Such groups are called *linear Lie groups* (or *matrix groups*).

Since the groups under consideration are subspaces of $\mathbb{R}^N$, we do not need immediately the definition of an abstract manifold.

We just have to define *embedded submanifolds* (also called submanifolds) of $\mathbb{R}^N$ (in the case of $\mathbf{GL}(n, \mathbb{R})$, $N = n^2$).

In general, the difficult part in proving that a subgroup of $\mathbf{GL}(n, \mathbb{R})$ is a Lie group is to prove that it is a manifold.

Fortunately, there is simple a characterization of the linear groups.

This characterization rests on two theorems. First, a Lie subgroup H of a Lie group G (where H is an embedded submanifold of G) is closed in G.

Second, a theorem of Von Neumann and Cartan asserts that a closed subgroup of $\mathbf{GL}(n, \mathbb{R})$ is an embedded submanifold, and thus, a Lie group.

Thus, *a linear Lie group is a closed subgroup of* $\mathbf{GL}(n, \mathbb{R})$.

Since our Lie groups are subgroups (or isomorphic to subgroups) of $\mathbf{GL}(n, \mathbb{R})$ for some suitable n, it is easy to define the Lie algebra of a Lie group using *curves*.

A small annoying technical arises in our approach, the problem with discrete subgroups.

If A is a subset of $\mathbb{R}^N$, recall that A inherits a topology from $\mathbb{R}^N$ called the *subspace topology*, and defined such that a subset V of A is open if

$$V = A \cap U$$

for some open subset U of $\mathbb{R}^N$. A point $a \in A$ is said to be *isolated* if there is there is some open subset U of $\mathbb{R}^N$ such that

$$\{a\} = A \cap U,$$

in other words, if $\{a\}$ is an open set in A.

The group $\mathbf{GL}(n, \mathbb{R})$ of real invertible $n \times n$ matrices can be viewed as a subset of $\mathbb{R}^{n^2}$, and as such, it is a topological space under the subspace topology (in fact, a dense open subset of $\mathbb{R}^{n^2}$).

One can easily check that multiplication and the inverse operation are continuous, and in fact smooth (i.e., C^∞-continuously differentiable).

This makes $\mathbf{GL}(n,\mathbb{R})$ a *topological group*.

Any subgroup G of $\mathbf{GL}(n,\mathbb{R})$ is also a topological space under the subspace topology.

A subgroup G is called a *discrete subgroup* if it has some isolated point.

This turns out to be equivalent to the fact that *every point of G is isolated*, and thus, G has the discrete topology (every subset of G is open).

Now, because $\mathbf{GL}(n,\mathbb{R})$ is Hausdorff, it can be shown that *every discrete subgroup of $\mathbf{GL}(n,\mathbb{R})$ is closed* and in fact, countable.

Thus, *discrete subgroups of $\mathbf{GL}(n,\mathbb{R})$ are Lie groups!*

But these are not very interesting Lie groups so we will consider only closed subgroups of $\mathbf{GL}(n, \mathbb{R})$ that are not discrete.

Let us now review the definition of an embedded submanifold. For simplicity, we restrict our attention to smooth manifolds.

For the sake of brevity, we use the terminology *manifold* (but other authors would say *embedded submanifolds*, or something like that).

The intuition behind the notion of a smooth manifold in $\mathbb{R}^N$ is that a subspace M is a manifold of dimension m if every point $p \in M$ is contained in some open subset U of M (in the subspace topology) that can be parametrized by some function $\varphi\colon \Omega \to U$ from some open subset Ω of the origin in $\mathbb{R}^m$, and that φ has some nice properties: (1) the definition of *smooth functions on M* and (2) the definition of the *tangent space at p*.

For this, φ has to be at least a homeomorphism, but more is needed: φ must be smooth, and the derivative $\varphi'(0_m)$ at the origin must be injective (letting $0_m = \underbrace{(0, \ldots, 0)}_{m}$).

Definition 2.9.1 Given any integers N, m, with $N \geq m \geq 1$, an *m-dimensional smooth manifold in $\mathbb{R}^N$, for short a manifold*, is a nonempty subset M of $\mathbb{R}^N$ such that for every point $p \in M$ there are two open subsets $\Omega \subseteq \mathbb{R}^m$ and $U \subseteq M$, with $p \in U$, and a smooth function $\varphi\colon \Omega \to \mathbb{R}^N$ such that φ is a homeomorphism between Ω and $U = \varphi(\Omega)$, and $\varphi'(t_0)$ is injective, where $t_0 = \varphi^{-1}(p)$.

The function $\varphi\colon \Omega \to U$ is called a *(local) parametrization of M at p*. If $0_m \in \Omega$ and $\varphi(0_m) = p$, we say that $\varphi\colon \Omega \to U$ is *centered at p*.

Recall that $M \subseteq \mathbb{R}^N$ is a topological space under the subspace topology, and U is some open subset of M in the subspace topology, which means that $U = M \cap W$ for some open subset W of $\mathbb{R}^N$.

Since $\varphi\colon \Omega \to U$ is a homeomorphism, it has an inverse $\varphi^{-1}\colon U \to \Omega$ that is also a homeomorphism, called a *(local) chart*.

Since $\Omega \subseteq \mathbb{R}^m$, for every $p \in M$ and every parametrization $\varphi\colon \Omega \to U$ of M at p, we have $\varphi^{-1}(p) = (z_1, \ldots, z_m)$ for some $z_i \in \mathbb{R}$, and we call $z_1, \ldots, z_m$ the *local coordinates of p (w.r.t. φ^{-1})*.

We often refer to a manifold M without explicitly specifying its dimension (the integer m).

Intuitively, a chart provides a "flattened" local map of a region on a manifold.

For instance, in the case of surfaces (2-dimensional manifolds), a chart is analogous to a planar map of a region on the surface.

For a concrete example, consider a map giving a planar representation of a country, a region on the earth, a curved surface.

Remark: We could allow $m = 0$ in definition 2.9.1. If so, a manifold of dimension 0 is just a set of isolated points, and thus it has the discrete topology.

In fact, it can be shown that a discrete subset of $\mathbb{R}^N$ is countable. Such manifolds are not very exciting, but they do correspond to discrete subgroups.

Example 2.3 The unit sphere S^2 in $\mathbb{R}^3$ defined such that

$$S^2 = \left\{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\right\}$$

is a smooth 2-manifold, because it can be parametrized using the following two maps φ_1 and φ_2:

$$\varphi_1\colon (u, v) \mapsto \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right)$$

and

$$\varphi_2\colon (u, v) \mapsto \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{1 - u^2 - v^2}{u^2 + v^2 + 1}\right).$$

The map φ_1 corresponds to the inverse of the stereographic projection from the north pole $N = (0, 0, 1)$ onto the plane $z = 0$, and the map φ_2 corresponds to the inverse of the stereographic projection from the south pole $S = (0, 0, -1)$ onto the plane $z = 0$, as illustrated in Figure 2.1.

The reader should check that the map φ_1 parametrizes $S^2 - \{N\}$ and that the map φ_2 parametrizes $S^2 - \{S\}$ (and that they are smooth, homeomorphisms, etc.).

Using φ_1, the open lower hemisphere is parametrized by the open disk of center O and radius 1 contained in the plane $z = 0$.

The chart φ_1^{-1} assigns local coordinates to the points in the open lower hemisphere.

If we draw a grid of coordinate lines parallel to the x and y axes inside the open unit disk and map these lines onto the lower hemisphere using φ_1, we get curved lines on the lower hemisphere.

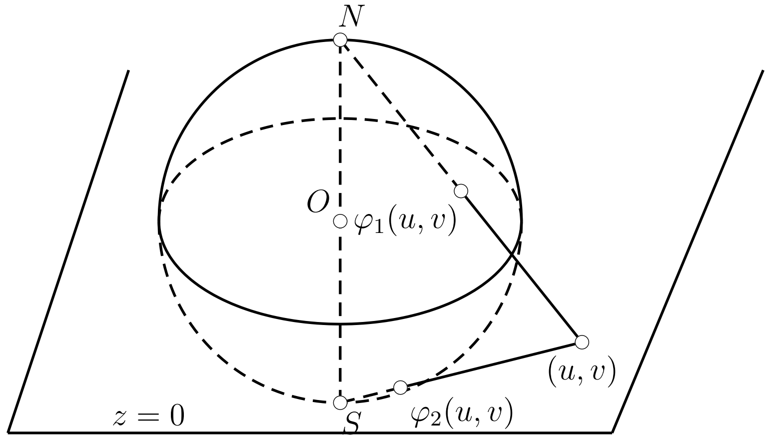

Figure 2.1: Inverse stereographic projections

These "coordinate lines" on the lower hemisphere provide local coordinates for every point on the lower hemisphere.

For this reason, older books often talk about *curvilinear coordinate systems* to mean the coordinate lines on a surface induced by a chart.

We urge our readers to define a manifold structure on a torus. This can be done using four charts.

Every open subset of $\mathbb{R}^N$ is a manifold in a trivial way. Indeed, we can use the inclusion map as a parametrization.

In particular, $\mathbf{GL}(n,\mathbb{R})$ is an open subset of $\mathbb{R}^{n^2}$, since its complement is closed (the set of invertible matrices is the inverse image of the determinant function, which is continuous).

Thus, $\mathbf{GL}(n,\mathbb{R})$ is a manifold. We can view $\mathbf{GL}(n,\mathbb{C})$ as a subset of $\mathbb{R}^{(2n)^2}$ using the embedding defined as follows:

For every complex $n \times n$ matrix A, construct the real $2n \times 2n$ matrix such that every entry $a + ib$ in A is replaced by the 2×2 block

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$$

where $a, b \in \mathbb{R}$.

It is immediately verified that this map is in fact a group isomorphism.

Thus, we can view $\mathbf{GL}(n, \mathbb{C})$ as a subgroup of $\mathbf{GL}(2n, \mathbb{R})$, and as a manifold in $\mathbb{R}^{(2n)^2}$.

A 1-manifold is called a *(smooth) curve*, and a 2-manifold is called a *(smooth) surface* (although some authors require that they also be connected).

The following two lemmas provide the link with the definition of an abstract manifold.

The first lemma is easily shown using the inverse function theorem.

Lemma 2.9.2 *Given an m-dimensional manifold M in $\mathbb{R}^N$, for every $p \in M$ there are two open sets $O, W \subseteq \mathbb{R}^N$ with $0_N \in O$ and $p \in M \cap W$, and a smooth diffeomorphism $\varphi\colon O \to W$, such that $\varphi(0_N) = p$ and*

$$\varphi(O \cap (\mathbb{R}^m \times \{0_{N-m}\})) = M \cap W.$$

The next lemma is easily shown from Lemma 2.9.2. It is a key technical result used to show that interesting properties of maps between manifolds do not depend on parametrizations.

Lemma 2.9.3 *Given an m-dimensional manifold M in $\mathbb{R}^N$, for every $p \in M$ and any two parametrizations $\varphi_1\colon \Omega_1 \to U_1$ and $\varphi_2\colon \Omega_2 \to U_2$ of M at p, if $U_1 \cap U_2 \neq \emptyset$, the map $\varphi_2^{-1} \circ \varphi_1\colon \varphi_1^{-1}(U_1 \cap U_2) \to \varphi_2^{-1}(U_1 \cap U_2)$ is a smooth diffeomorphism.*

The maps $\varphi_2^{-1} \circ \varphi_1\colon \varphi_1^{-1}(U_1 \cap U_2) \to \varphi_2^{-1}(U_1 \cap U_2)$ are called *transition maps*. Lemma 2.9.3 is illustrated in Figure 2.2.

Using Definition 2.9.1, it may be quite hard to prove that a space is a manifold. Therefore, it is handy to have alternate characterizations such as those given in the next Proposition:

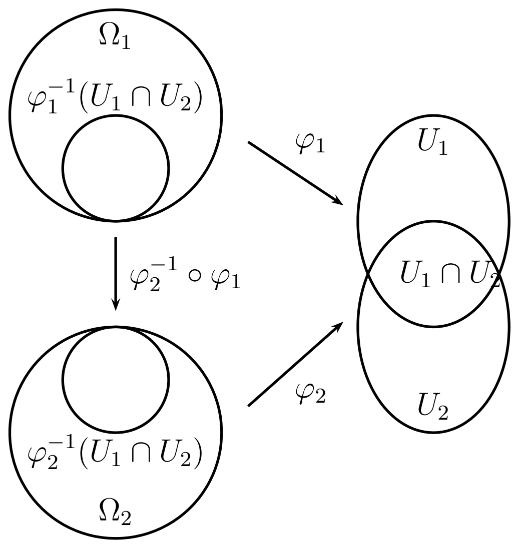

Figure 2.2: Parametrizations and transition functions

Proposition 2.9.4 *A subset, $M \subseteq \mathbb{R}^{m+k}$, is an m-dimensional manifold iff either*

(1) For every $p \in M$, there is some open subset, $W \subseteq \mathbb{R}^{m+k}$, with $p \in W$ and a (smooth) submersion, $f \colon W \to \mathbb{R}^k$, so that $W \cap M = f^{-1}(0)$,
or

(2) For every $p \in M$, there is some open subset, $W \subseteq \mathbb{R}^{m+k}$, with $p \in W$ and a (smooth) map, $f \colon W \to \mathbb{R}^k$, so that $f'(p)$ is surjective and $W \cap M = f^{-1}(0)$.

Observe that condition (2), although apparently weaker than condition (1), is in fact equivalent to it, but more convenient in practice.

This is because to say that $f'(p)$ is surjective means that the Jacobian matrix of $f'(p)$ has rank m, which means that some determinant is nonzero, and because the determinant function is continuous this must hold in some open subset $W_1 \subseteq W$ containing p.

Consequenly, the restriction, f_1, of f to W_1 is indeed a submersion and $f_1^{-1}(0) = W_1 \cap f^{-1}(0) = W_1 \cap W \cap M = W_1 \cap M$.

The proof, which is somewhat illuminating, is based on two technical lemmas that are proved using the inverse function theorem.

Lemma 2.9.5 *Let $U \subseteq \mathbb{R}^m$ be an open subset of $\mathbb{R}^m$ and pick some $a \in U$. If $f: U \to \mathbb{R}^n$ is a smooth immersion at a, i.e., df_a is injective (so, $m \leq n$), then there is an open set, $V \subseteq \mathbb{R}^n$, with $f(a) \in V$, an open subset, $U' \subseteq U$, with $a \in U'$ and $f(U') \subseteq V$, an open subset $O \subseteq \mathbb{R}^{n-m}$, and a diffeomorphism, $\theta: V \to U' \times O$, so that*

$$\theta(f(x_1, \ldots, x_m)) = (x_1, \ldots, x_m, 0, \ldots, 0),$$

for all $(x_1, \ldots, x_m) \in U'$.

Lemma 2.9.6 *Let $W \subseteq \mathbb{R}^m$ be an open subset of $\mathbb{R}^m$ and pick some $a \in W$. If $f: W \to \mathbb{R}^n$ is a smooth submersion at a, i.e., df_a is surjective (so, $m \geq n$), then there is an open set, $V \subseteq W \subseteq \mathbb{R}^m$, with $a \in V$, and a diffeomorphism, ψ, with domain $O \subseteq \mathbb{R}^m$, so that $\psi(O) = V$ and*

$$f(\psi(x_1, \ldots, x_m)) = (x_1, \ldots, x_n),$$

for all $(x_1, \ldots, x_m) \in O$.

Using Lemmas 2.9.5 and 2.9.6, we can prove the following theorem which confirms that all our characterizations of a manifold are equivalent.

Theorem 2.9.7 *A nonempty subset, $M \subseteq \mathbb{R}^N$, is an m-manifold (with $1 \leq m \leq N$) iff any of the following conditions hold:*

(1) For every $p \in M$, there are two open subsets $\Omega \subseteq \mathbb{R}^m$ and $U \subseteq M$, with $p \in U$, and a smooth function $\varphi\colon \Omega \to \mathbb{R}^N$ such that φ is a homeomorphism between Ω and $U = \varphi(\Omega)$, and $\varphi'(0)$ is injective, where $p = \varphi(0)$.

(2) For every $p \in M$, there are two open sets $O, W \subseteq \mathbb{R}^N$ with $0_N \in O$ and $p \in M \cap W$, and a smooth diffeomorphism $\varphi\colon O \to W$, such that $\varphi(0_N) = p$ and

$$\varphi(O \cap (\mathbb{R}^m \times \{0_{N-m}\})) = M \cap W.$$

(3) For every $p \in M$, there is some open subset, $W \subseteq \mathbb{R}^N$, with $p \in W$ and a smooth submersion, $f\colon W \to \mathbb{R}^{N-m}$, so that $W \cap M = f^{-1}(0)$.

(4) For every $p \in M$, there is some open subset, $W \subseteq \mathbb{R}^N$, and $N - m$ smooth functions, $f_i \colon W \to \mathbb{R}$, so that the linear forms $df_1(p), \ldots, df_{N-m}(p)$ are linearly independent and

$$W \cap M = f_1^{-1}(0) \cap \cdots \cap f_{N-m}^{-1}(0).$$

Condition (4) says that locally (that is, in a small open set of M containing $p \in M$), M is "cut out" by $N - m$ smooth functions, $f_i \colon W \to \mathbb{R}$, in the sense that the portion of the manifold $M \cap W$ is the intersection of the $N - m$ hypersurfaces, $f_i^{-1}(0)$, (the zero-level sets of the f_i) and that this intersection is "clean", which means that the linear forms $df_1(p), \ldots, df_{N-m}(p)$ are linearly independent.

As an illustration of Theorem 2.9.7, the sphere

$$S^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_2^2 - 1 = 0\}$$

is an n-dimensional manifold in $\mathbb{R}^{n+1}$. Indeed, the map $f\colon \mathbb{R}^{n+1} \to \mathbb{R}$ given by $f(x) = \|x\|_2^2 - 1$ is a submersion, since

$$df(x)(y) = 2\sum_{k=1}^{n+1} x_k y_k.$$

The rotation group, $\mathbf{SO}(n)$, is an $\frac{n(n-1)}{2}$-dimensional manifold in $\mathbb{R}^{n^2}$.

Indeed, $\mathbf{GL}^+(n)$ is an open subset of $\mathbb{R}^{n^2}$ (recall, $\mathbf{GL}^+(n) = \{A \in \mathbf{GL}(n) \mid \det(A) > 0\}$) and if f is defined by

$$f(A) = A^\top A - I,$$

where $A \in \mathbf{GL}^+(n)$, then $f(A)$ is symmetric, so $f(A) \in \mathbf{S}(n) = \mathbb{R}^{\frac{n(n+1)}{2}}$.

It is easy to show (using directional derivatives) that

$$df(A)(H) = A^\top H + H^\top A.$$

But then, $df(A)$ is surjective for all $A \in \mathbf{SO}(n)$, because if S is any symmetric matrix, we see that

$$df(A)\left(\frac{AS}{2}\right) = S.$$

As $\mathbf{SO}(n) = f^{-1}(0)$, we conclude that $\mathbf{SO}(n)$ is indeed a manifold.

A similar argument proves that $\mathbf{O}(n)$ is an $\frac{n(n-1)}{2}$-dimensional manifold.

Using the map, $f\colon \mathbf{GL}(n) \to \mathbb{R}$, given by $A \mapsto \det(A)$, we can prove that $\mathbf{SL}(n)$ is a manifold of dimension n^2-1.

Remark: We have $df(A)(B) = \det(A)\mathrm{tr}(A^{-1}B)$, for every $A \in \mathbf{GL}(n)$.

The third characterization of Theorem 2.9.7 suggests the following definition.

Definition 2.9.8 Let $f\colon \mathbb{R}^{m+k} \to \mathbb{R}^k$ be a smooth function. A point, $p \in \mathbb{R}^{m+k}$, is called a *critical point (of f)* iff df_p is *not* surjective and a point $q \in \mathbb{R}^k$ is called a *critical value (of f)* iff $q = f(p)$, for some critical point, $p \in \mathbb{R}^{m+k}$. A point $p \in \mathbb{R}^{m+k}$ is a *regular point (of f)* iff p is not critical, i.e., df_p is surjective, and a point $q \in \mathbb{R}^k$ is a *regular value (of f)* iff it is not a critical value. In particular, any $q \in \mathbb{R}^k - f(\mathbb{R}^{m+k})$ is a regular value and $q \in f(\mathbb{R}^{m+k})$ is a regular value iff *every* $p \in f^{-1}(q)$ is a regular point (but, in contrast, q is a critical value iff *some* $p \in f^{-1}(q)$ is critical).

Part (3) of Theorem 2.9.7 implies the following useful proposition:

Proposition 2.9.9 *Given any smooth function, $f\colon \mathbb{R}^{m+k} \to \mathbb{R}^k$, for every regular value, $q \in f(\mathbb{R}^{m+k})$, the preimage, $Z = f^{-1}(q)$, is a manifold of dimension m.*

Definition 2.9.8 and Proposition 2.9.9 can be generalized to manifolds

Regular and critical values of smooth maps play an important role in differential topology.

Firstly, given a smooth map, $f\colon \mathbb{R}^{m+k} \to \mathbb{R}^k$, almost every point of $\mathbb{R}^k$ is a regular value of f. To make this statement precise, one needs the notion of a *set of measure zero*.

Then, *Sard's theorem* says that the set of critical values of a smooth map has measure zero.

Secondly, if we consider smooth functions, $f\colon \mathbb{R}^{m+1} \to \mathbb{R}$, a point $p \in \mathbb{R}^{m+1}$ is critical iff $df_p = 0$.

Then, we can use second order derivatives to further classify critical points. The *Hessian matrix* of f (at p) is the matrix of second-order partials

$$H_f(p) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(p)\right)$$

and a critical point p is a *nondegenerate critical point* if $H_f(p)$ is a nonsingular matrix.

The remarkable fact is that, at a nondegenerate critical point, p, the local behavior of f is completely determined, in the sense that after a suitable change of coordinates (given by a smooth diffeomorphism)

$$f(x) = f(p) - x_1^2 - \cdots - x_\lambda^2 + x_{\lambda+1}^2 + \cdots + x_{m+1}^2$$

near p, where λ called the *index of f at p* is an integer which depends only on p (in fact, λ is the number of negative eigenvalues of $H_f(p)$).

This result is known as *Morse lemma* (after Marston Morse, 1892-1977).

Smooth functions whose critical points are all nondegenerate are called *Morse functions*.

It turns out that every smooth function, $f\colon \mathbb{R}^{m+1} \to \mathbb{R}$, gives rise to a large supply of Morse functions by adding a linear function to it.

More precisely, the set of $a \in \mathbb{R}^{m+1}$ for which the function f_a given by

$$f_a(x) = f(x) + a_1x_1 + \cdots + a_{m+1}x_{m+1}$$

is not a Morse function has measure zero.

Morse functions can be used to study topological properties of manifolds.

In a sense to be made precise and under certain technical conditions, a Morse function can be used to reconstuct a manifold by attaching cells, up to homotopy equivalence.

However, these results are way beyond the scope of these notes.

Let us now review the definitions of a smooth curve in a manifold and the tangent vector at a point of a curve.

Definition 2.9.10 Let M be an m-dimensional manifold in $\mathbb{R}^N$. A *smooth curve γ in M* is any function $\gamma: I \to M$ where I is an open interval in $\mathbb{R}$ and such that for every $t \in I$, letting $p = \gamma(t)$, there is some parametrization $\varphi: \Omega \to U$ of M at p and some open interval $]t - \epsilon,\ t + \epsilon[\ \subseteq I$ such that the curve $\varphi^{-1} \circ \gamma:]t - \epsilon, t + \epsilon[\to \mathbb{R}^m$ is smooth.

Using Lemma 2.9.3, it is easily shown that Definition 2.9.10 does not depend on the choice of the parametrization $\varphi: \Omega \to U$ at p.

Lemma 2.9.3 also implies that γ viewed as a curve $\gamma \colon I \to \mathbb{R}^N$ is smooth.

Then the *tangent vector to the curve $\gamma \colon I \to \mathbb{R}^N$ at t*, denoted by $\gamma'(t)$, is the value of the derivative of γ at t (a vector in $\mathbb{R}^N$) computed as usual:

$$\gamma'(t) = \lim_{h \mapsto 0} \frac{\gamma(t+h) - \gamma(t)}{h}.$$

Given any point $p \in M$, we will show that the set of tangent vectors to all smooth curves in M through p is a vector space isomorphic to the vector space $\mathbb{R}^m$.

The tangent vector at p to a curve γ on a manifold M is illustrated in Figure 2.3.

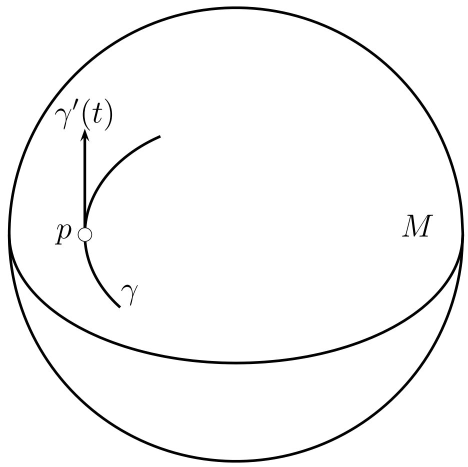

Figure 2.3: Tangent vector to a curve on a manifold

Given a smooth curve $\gamma\colon I \to M$, for any $t \in I$, letting $p = \gamma(t)$, since M is a manifold, there is a parametrization $\varphi\colon \Omega \to U$ such that $\varphi(0_m) = p \in U$ and some open interval $J \subseteq I$ with $t \in J$ and such that the function

$$\varphi^{-1} \circ \gamma\colon J \to \mathbb{R}^m$$

is a smooth curve, since γ is a smooth curve.

Letting $\alpha = \varphi^{-1} \circ \gamma$, the derivative $\alpha'(t)$ is well-defined, and it is a vector in $\mathbb{R}^m$.

But $\varphi \circ \alpha \colon J \to M$ is also a smooth curve, which agrees with γ on J, and by the chain rule,

$$\gamma'(t) = \varphi'(0_m)(\alpha'(t)),$$

since $\alpha(t) = 0_m$ (because $\varphi(0_m) = p$ and $\gamma(t) = p$).

Observe that $\gamma'(t)$ is a vector in $\mathbb{R}^N$.

Now, for every vector $v \in \mathbb{R}^m$, the curve $\alpha \colon J \to \mathbb{R}^m$ defined such that

$$\alpha(u) = (u - t)v$$

for all $u \in J$ is clearly smooth, and $\alpha'(t) = v$.

This shows that the set of tangent vectors at t to all smooth curves (in $\mathbb{R}^m$) passing through 0_m is the entire vector space $\mathbb{R}^m$.

Since every smooth curve $\gamma \colon I \to M$ agrees with a curve of the form $\varphi \circ \alpha \colon J \to M$ for some smooth curve $\alpha \colon J \to \mathbb{R}^m$ (with $J \subseteq I$) as explained above, and since it is assumed that $\varphi'(0_m)$ is injective, $\varphi'(0_m)$ maps the vector space $\mathbb{R}^m$ injectively to the set of tangent vectors to γ at p, as claimed.

All this is summarized in the following definition.

Definition 2.9.11 Let M be an m-dimensional manifold in $\mathbb{R}^N$. For every point $p \in M$, the *tangent space T_pM at p* is the set of all vectors in $\mathbb{R}^N$ of the form $\gamma'(0)$, where $\gamma: I \to M$ is any smooth curve in M such that $p = \gamma(0)$.

The set T_pM is a vector space isomorphic to $\mathbb{R}^m$. Every vector $v \in T_pM$ is called a *tangent vector to M at p*.

We can now define Lie groups.

Definition 2.9.12 A *Lie group* is a nonempty subset G of $\mathbb{R}^N$ $(N \geq 1)$ satisfying the following conditions:

(a) G is a group.

(b) G is a manifold in $\mathbb{R}^N$.

(c) The group operation $\cdot : G \times G \to G$ and the inverse map ${}^{-1}: G \to G$ are smooth.

Actually, we haven't defined yet what a smooth map between manifolds is (in clause (c)).

This notion is explained in Definition 2.9.16, but we feel that most readers will appreciate seeing the formal definition a Lie group, as early as possible.

It is immediately verified that $\mathbf{GL}(n, \mathbb{R})$ is a Lie group. Since all the Lie groups that we are considering are subgroups of $\mathbf{GL}(n, \mathbb{R})$, the following definition is in order.

Definition 2.9.13 A *linear Lie group* is a subgroup G of $\mathbf{GL}(n, \mathbb{R})$ (for some $n \geq 1$) which is a smooth manifold in $\mathbb{R}^{n^2}$.

Let $\mathbf{M}(n, \mathbb{R})$ denote the set of all real $n \times n$ matrices (invertible or not). If we recall that the exponential map

$$\exp\colon A \mapsto e^A$$

is well defined on $\mathbf{M}(n, \mathbb{R})$, we have the following crucial theorem due to Von Neumann and Cartan:

Theorem 2.9.14 *A closed subgroup G of $\mathbf{GL}(n,\mathbb{R})$ is a linear Lie group. Furthermore, the set $\mathfrak{g}$ defined such that*

$$\mathfrak{g} = \{X \in \mathbf{M}(n,\mathbb{R}) \mid e^{tX} \in G \textit{ for all } t \in \mathbb{R}\}$$

is a vector space equal to the tangent space $T_I G$ at the identity I, and $\mathfrak{g}$ is closed under the Lie bracket $[-,-]$ defined such that $[A,B] = AB - BA$ for all $A, B \in \mathbf{M}(n,\mathbb{R})$.

Theorem 2.9.14 applies even when G is a discrete subgroup, but in this case, $\mathfrak{g}$ is trivial (i.e., $\mathfrak{g} = \{0\}$). For example, the set of nonzero reals $\mathbb{R}^* = \mathbb{R} - \{0\} = \mathbf{GL}(1,\mathbb{R})$ is a Lie group under multiplication, and the subgroup

$$H = \{2^n \mid n \in \mathbb{Z}\}$$

is a discrete subgroup of $\mathbb{R}^*$. Thus, H is a Lie group. On the other hand, the set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ of nonzero rational numbers is a multiplicative subgroup of $\mathbb{R}^*$, but it is not closed, since $\mathbb{Q}$ is dense in $\mathbb{R}$.

If G is closed and not discrete, we must have $m \geq 1$, and $\mathfrak{g}$ has dimension m.

With the help of Theorem 2.9.14 it is now very easy to prove that $\mathbf{SL}(n)$, $\mathbf{O}(n)$, $\mathbf{SO}(n)$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$, and $\mathbf{SU}(n)$ are Lie groups. We can also prove that $\mathbf{SE}(n)$ is a Lie group as follows. Recall that we can view every element of $\mathbf{SE}(n)$ as a real $(n+1) \times (n+1)$ matrix

$$\begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}$$

where $R \in \mathbf{SO}(n)$ and $U \in \mathbb{R}^n$.

In fact, such matrices belong to $\mathbf{SL}(n+1)$.

This embedding of $\mathbf{SE}(n)$ into $\mathbf{SL}(n+1)$ is a group homomorphism, since the group operation on $\mathbf{SE}(n)$ corresponds to multiplication in $\mathbf{SL}(n+1)$:

$$\begin{pmatrix} RS & RV+U \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix} \begin{pmatrix} S & V \\ 0 & 1 \end{pmatrix}.$$

Note that the inverse is given by

$$\begin{pmatrix} R^{-1} & -R^{-1}U \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} R^{\top} & -R^{\top}U \\ 0 & 1 \end{pmatrix}.$$

Also note that the embedding shows that as a manifold, $\mathbf{SE}(n)$ is diffeomorphic to $\mathbf{SO}(n) \times \mathbb{R}^n$ (given a manifold M_1 of dimension m_1 and a manifold M_2 of dimension m_2, the product $M_1 \times M_2$ can be given the structure of a manifold of dimension $m_1 + m_2$ in a natural way).

Thus, $\mathbf{SE}(n)$ is a Lie group with underlying manifold $\mathbf{SO}(n) \times \mathbb{R}^n$, and in fact, a subgroup of $\mathbf{SL}(n+1)$.

Even though $\mathbf{SE}(n)$ is diffeomorphic to $\mathbf{SO}(n)\times\mathbb{R}^n$ as a manifold, it is *not* isomorphic to $\mathbf{SO}(n)\times\mathbb{R}^n$ as a group, because the group multiplication on $\mathbf{SE}(n)$ is not the multiplication on $\mathbf{SO}(n)\times\mathbb{R}^n$. Instead, $\mathbf{SE}(n)$ is a *semidirect product* of $\mathbf{SO}(n)$ and $\mathbb{R}^n$.

Returning to Theorem 2.9.14, the vector space $\mathfrak{g}$ is called the *Lie algebra* of the Lie group G.

Lie algebras are defined as follows.

Definition 2.9.15 A *(real) Lie algebra* $\mathcal{A}$ is a real vector space together with a bilinear map $[\cdot, \cdot] \colon \mathcal{A} \times \mathcal{A} \to \mathcal{A}$ called the *Lie bracket* on $\mathcal{A}$ such that the following two identities hold for all $a, b, c \in \mathcal{A}$:

$$[a,\, a] = 0,$$

and the so-called *Jacobi identity*

$$[a,\, [b,\, c]] + [c,\, [a,\, b]] + [b,\, [c,\, a]] = 0.$$

It is immediately verified that $[b, a] = -[a, b]$.

In view of Theorem 2.9.14, the vector space $\mathfrak{g} = T_I G$ associated with a Lie group G is indeed a Lie algebra. Furthermore, the exponential map $\exp \colon \mathfrak{g} \to G$ is well-defined.

In general, exp is neither injective nor surjective, as we observed earlier.

Theorem 2.9.14 also provides a kind of recipe for "computing" the Lie algebra $\mathfrak{g} = T_I G$ of a Lie group G.

Indeed, $\mathfrak{g}$ is the tangent space to G at I, and thus we can use curves to compute tangent vectors.

Actually, for every $X \in T_I G$, the map

$$\gamma_X \colon t \mapsto e^{tX}$$

is a smooth curve in G, and it is easily shown that $\gamma_X'(0) = X$. Thus, we can use these curves.

As an illustration, we show that the Lie algebras of $\mathbf{SL}(n)$ and $\mathbf{SO}(n)$ are the matrices with null trace and the skew symmetric matrices.

Let $t \mapsto R(t)$ be a smooth curve in $\mathbf{SL}(n)$ such that $R(0) = I$. We have $\det(R(t)) = 1$ for all $t \in]-\epsilon, \epsilon\,[$.

Using the chain rule, we can compute the derivative of the function

$$t \mapsto \det(R(t))$$

at $t = 0$, and we get

$$\det{}'_I(R'(0)) = 0.$$

It is an easy exercise to prove that

$$\det{}'_I(X) = \mathrm{tr}(X),$$

and thus $\mathrm{tr}(R'(0)) = 0$, which says that the tangent vector $X = R'(0)$ has null trace.

Another proof consists in observing that $X \in \mathfrak{sl}(n, \mathbb{R})$ iff

$$\det(e^{tX}) = 1$$

for all $t \in \mathbb{R}$. Since $\det(e^{tX}) = e^{\mathrm{tr}(tX)}$, for $t = 1$, we get $\mathrm{tr}(X) = 0$, as claimed.

Clearly, $\mathfrak{sl}(n, \mathbb{R})$ has dimension $n^2 - 1$.

Let $t \mapsto R(t)$ be a smooth curve in $\mathbf{SO}(n)$ such that $R(0) = I$. Since each $R(t)$ is orthogonal, we have

$$R(t)\, R(t)^\top = I$$

for all $t \in]-\epsilon, \epsilon\, [$.

Taking the derivative at $t = 0$, we get

$$R'(0)\, R(0)^\top + R(0)\, R'(0)^\top = 0,$$

but since $R(0) = I = R(0)^\top$, we get

$$R'(0) + R'(0)^\top = 0,$$

which says that the tangent vector $X = R'(0)$ is skew symmetric.

Since the diagonal elements of a skew symmetric matrix are null, the trace is automatically null, and the condition $\det(R) = 1$ yields nothing new.

This shows that $\mathfrak{o}(n) = \mathfrak{so}(n)$. It is easily shown that $\mathfrak{so}(n)$ has dimension $n(n-1)/2$.

As a concrete example, the Lie algebra $\mathfrak{so}(3)$ of $\mathbf{SO}(3)$ is the real vector space consisting of all 3×3 real skew symmetric matrices. Every such matrix is of the form

$$\begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}$$

where $b, c, d \in \mathbb{R}$. The Lie bracket $[A, B]$ in $\mathfrak{so}(3)$ is also given by the usual commutator, $[A, B] = AB - BA$.

We can define an isomorphism of Lie algebras $\psi \colon (\mathbb{R}^3, \times) \to \mathfrak{so}(3)$ by the formula

$$\psi(b, c, d) = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix}.$$

It is indeed easy to verify that

$$\psi(u \times v) = [\psi(u),\ \psi(v)].$$

It is also easily verified that for any two vectors $u = (b, c, d)$ and $v = (b', c', d')$ in $\mathbb{R}^3$,

$$\psi(u)(v) = u \times v.$$

The exponential map $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ is given by Rodrigues's formula (see Lemma 2.3.3):

$$e^A = \cos\theta\, I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} B,$$

or equivalently by

$$e^A = I_3 + \frac{\sin\theta}{\theta} A + \frac{(1-\cos\theta)}{\theta^2} A^2$$

if $\theta \neq 0$, where

$$A = \begin{pmatrix} 0 & -d & c \\ d & 0 & -b \\ -c & b & 0 \end{pmatrix},$$

$\theta = \sqrt{b^2 + c^2 + d^2}$, $B = A^2 + \theta^2 I_3$, and with $e^{0_3} = I_3$.

Using the above methods, it is easy to verify that the Lie algebras $\mathfrak{gl}(n, \mathbb{R})$, $\mathfrak{sl}(n, \mathbb{R})$, $\mathfrak{o}(n)$, and $\mathfrak{so}(n)$, are respectively $\mathbf{M}(n, \mathbb{R})$, the set of matrices with null trace, and the set of skew symmetric matrices (in the last two cases).

A similar computation can be done for $\mathfrak{gl}(n, \mathbb{C})$, $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$, confirming the claims of Section 2.5.

It is easy to show that $\mathfrak{gl}(n, \mathbb{C})$ has dimension $2n^2$, $\mathfrak{sl}(n, \mathbb{C})$ has dimension $2(n^2 - 1)$, $\mathfrak{u}(n)$ has dimension n^2, and $\mathfrak{su}(n)$ has dimension $n^2 - 1$.

For example, the Lie algebra $\mathfrak{su}(2)$ of $\mathbf{SU}(2)$ (or S^3) is the real vector space consisting of all 2×2 (complex) skew Hermitian matrices of null trace.

As we showed, $\mathbf{SE}(n)$ is a Lie group, and its lie algebra $\mathfrak{se}(n)$ described in Section 2.7 is easily determined as the subalgebra of $\mathfrak{sl}(n+1)$ consisting of all matrices of the form

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix}$$

where $B \in \mathfrak{so}(n)$ and $U \in \mathbb{R}^n$.

Thus, $\mathfrak{se}(n)$ has dimension $n(n+1)/2$. The Lie bracket is given by

$$\begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix} \begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix} - \begin{pmatrix} C & V \\ 0 & 0 \end{pmatrix} \begin{pmatrix} B & U \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} BC - CB & BV - CU \\ 0 & 0 \end{pmatrix}.$$

We conclude by indicating the relationship between homomorphisms of Lie groups and homomorphisms of Lie algebras.

First, we need to explain what is meant by a smooth map between manifolds.

Definition 2.9.16 Let M_1 (m_1-dimensional) and M_2 (m_2-dimensional) be manifolds in $\mathbb{R}^N$. A function $f\colon M_1 \to M_2$ is *smooth* if for every $p \in M_1$ there are parametrizations $\varphi\colon \Omega_1 \to U_1$ of M_1 at p and $\psi\colon \Omega_2 \to U_2$ of M_2 at $f(p)$ such that $f(U_1) \subseteq U_2$ and

$$\psi^{-1} \circ f \circ \varphi\colon \Omega_1 \to \mathbb{R}^{m_2}$$

is smooth.

Using Lemma 2.9.3, it is easily shown that Definition 2.9.16 does not depend on the choice of the parametrizations $\varphi\colon \Omega_1 \to U_1$ and $\psi\colon \Omega_2 \to U_2$.

A smooth map f between manifolds is a *smooth diffeomorphism* if f is bijective and both f and f^{-1} are smooth maps.

We now define the derivative of a smooth map between manifolds.

Definition 2.9.17 Let M_1 (m_1-dimensional) and M_2 (m_2-dimensional) be manifolds in $\mathbb{R}^N$. For any smooth function $f\colon M_1 \to M_2$ and any $p \in M_1$, the function $f'_p\colon T_pM_1 \to T_{f(p)}M_2$, called the *tangent map of f at p, or derivative of f at p, or differential of f at p*, is defined as follows: For every $v \in T_pM_1$ and every smooth curve $\gamma\colon I \to M_1$ such that $\gamma(0) = p$ and $\gamma'(0) = v$,

$$f'_p(v) = (f \circ \gamma)'(0).$$

The map f'_p is also denoted by df_p or T_pf.

Doing a few calculations involving the facts that

$$f \circ \gamma = (f \circ \varphi) \circ (\varphi^{-1} \circ \gamma) \quad \text{and} \quad \gamma = \varphi \circ (\varphi^{-1} \circ \gamma)$$

and using Lemma 2.9.3, it is not hard to show that $f'_p(v)$ does not depend on the choice of the curve γ. It is easily shown that f'_p is a linear map.

Finally, we define homomorphisms of Lie groups and Lie algebras and see how they are related.

Definition 2.9.18 Given two Lie groups G_1 and G_2, a *homomorphism (or map) of Lie groups* is a function $f: G_1 \to G_2$ that is a homomorphism of groups and a smooth map (between the manifolds G_1 and G_2).

Given two Lie algebras $\mathcal{A}_1$ and $\mathcal{A}_2$, a *homomorphism (or map) of Lie algebras* is a function $f: \mathcal{A}_1 \to \mathcal{A}_2$ that is a linear map between the vector spaces $\mathcal{A}_1$ and $\mathcal{A}_2$ and that preserves Lie brackets, i.e.,

$$f([A, B]) = [f(A), f(B)]$$

for all $A, B \in \mathcal{A}_1$.

An *isomorphism of Lie groups* is a bijective function f such that both f and f^{-1} are maps of Lie groups, and an *isomorphism of Lie algebras* is a bijective function f such that both f and f^{-1} are maps of Lie algebras.

It is immediately verified that if $f\colon G_1 \to G_2$ is a homomorphism of Lie groups, then $f'_I\colon \mathfrak{g}_1 \to \mathfrak{g}_2$ is a homomorphism of Lie algebras.

If some additional assumptions are made about G_1 and G_2 (for example, connected, simply connected), it can be shown that f is pretty much determined by f'_I.

Alert readers must have noticed that we only defined the Lie algebra of a linear group.

In the more general case, we can still define the Lie algebra $\mathfrak{g}$ of a Lie group G as the tangent space $T_I G$ at the identity I.

The tangent space $\mathfrak{g} = T_I G$ is a vector space, but we need to define the Lie bracket.

This can be done in several ways. We explain briefly how this can be done in terms of so-called *adjoint representations*.

This has the advantage of not requiring the definition of left-invariant vector fields, but it is still a little bizarre!

Given a Lie group G, for every $a \in G$ we define *left translation* as the map $L_a \colon G \to G$ such that $L_a(b) = ab$ for all $b \in G$, and *right translation* as the map $R_a \colon G \to G$ such that $R_a(b) = ba$ for all $b \in G$.

The maps L_a and R_a are diffeomorphisms, and their derivatives play an important role.

The inner automorphisms $R_{a^{-1}} \circ L_a$ (also written as $R_{a^{-1}} L_a$) also play an important role.

Note that

$$R_{a^{-1}} L_a(b) = aba^{-1}.$$

The derivative

$$(R_{a^{-1}}L_a)'_I \colon T_IG \to T_IG$$

of $R_{a^{-1}}L_a \colon G \to G$ at I is an isomorphism of Lie algebras, and since $T_IG = \mathfrak{g}$, we get a map denoted by $\mathrm{Ad}_a \colon \mathfrak{g} \to \mathfrak{g}$.

The map $a \mapsto \mathrm{Ad}_a$ is a map of Lie groups

$$\mathrm{Ad} \colon G \to \mathbf{GL}(\mathfrak{g}),$$

called the *adjoint representation of G* (where $\mathbf{GL}(\mathfrak{g})$ denotes the Lie group of all bijective linear maps on $\mathfrak{g}$).

In the case of a linear group, one can verify that

$$\mathrm{Ad}(a)(X) = \mathrm{Ad}_a(X) = aXa^{-1}$$

for all $a \in G$ and all $X \in \mathfrak{g}$.

The derivative

$$\mathrm{Ad}'_I \colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$$

of $\mathrm{Ad} \colon G \to \mathbf{GL}(\mathfrak{g})$ at I is map of Lie algebras, denoted by $\mathrm{ad} \colon \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g})$, called the *adjoint representation of $\mathfrak{g}$*.

Recall that Theorem 2.9.14 immediately implies that the Lie algebra, $\mathfrak{gl}(\mathfrak{g})$, of $\mathbf{GL}(\mathfrak{g})$ is the vector space of all linear maps on $\mathfrak{g}$.

In the case of a linear group, it can be verified that

$$\mathrm{ad}(A)(B) = [A,\, B]$$

for all $A, B \in \mathfrak{g}$.

One can also check that the Jacobi identity on $\mathfrak{g}$ is equivalent to the fact that ad preserves Lie brackets, i.e., ad is a map of Lie algebras:

$$\mathrm{ad}([A,\, B]) = [\mathrm{ad}(A),\, \mathrm{ad}(B)]$$

for all $A, B \in \mathfrak{g}$ (where on the right, the Lie bracket is the commutator of linear maps on $\mathfrak{g}$).

Thus, we recover the Lie bracket from ad.

This is the key to the definition of the Lie bracket in the case of a general Lie group (not just a linear Lie group).

We define the Lie bracket on $\mathfrak{g}$ as

$$[A,\, B] = \mathrm{ad}(A)(B).$$

To be complete, we still have to define the exponential map $\exp\colon \mathfrak{g} \to G$ for a general Lie group.

For this we need to introduce some left-invariant vector fields induced by the derivatives of the left translations and integral curves associated with such vector fields.