Notes on Differential Geometry and Lie Groups

Jean Gallier
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@cis.upenn.edu

February 28, 2011

To my daughter Mia, my wife Anne,

my son Philippe, and my daughter Sylvie.

Preface

The motivations for writing these notes arose while I was coteaching a seminar on Special Topics in Machine Perception with Kostas Daniilidis in the Spring of 2004. In the Spring of 2005, I gave a version of my course *Advanced Geometric Methods in Computer Science* (CIS610), with the main goal of discussing statistics on diffusion tensors and shape statistics in medical imaging. This is when I realized that it was necessary to cover some material on Riemannian geometry but I ran out of time after presenting Lie groups and never got around to doing it! Then, in the Fall of 2006 I went on a wonderful and very productive sabbatical year in Nicholas Ayache's group (ACSEPIOS) at INRIA Sophia Antipolis where I learned about the beautiful and exciting work of Vincent Arsigny, Olivier Clatz, Hervé Delingette, Pierre Fillard, Grégoire Malandin, Xavier Pennec, Maxime Sermesant, and, of course, Nicholas Ayache, on statistics on manifolds and Lie groups applied to medical imaging. This inspired me to write chapters on differential geometry and, after a few additions made during Fall 2007 and Spring 2008, notably on left-invariant metrics on Lie groups, my little set of notes from 2004 had grown into the manuscript found here.

Let me go back to the seminar on Special Topics in Machine Perception given in 2004. The main theme of the seminar was group-theoretical methods in visual perception. In particular, Kostas decided to present some exciting results from Christopher Geyer's Ph.D. thesis [62] on scene reconstruction using two parabolic catadioptric cameras (Chapters 4 and 5). Catadioptric cameras are devices which use both mirrors (catioptric elements) and lenses (dioptric elements) to form images. Catadioptric cameras have been used in computer vision and robotics to obtain a wide field of view, often greater than 180°, unobtainable from perspective cameras. Applications of such devices include navigation, surveillance and vizualization, among others. Technically, certain matrices called *catadioptric fundamental matrices* come up. Geyer was able to give several equivalent characterizations of these matrices (see Chapter 5, Theorem 5.2). To my surprise, the Lorentz group $\mathbf{O}(3,1)$ (of the theory of special relativity) comes up naturally! The set of fundamental matrices turns out to form a manifold, $\mathcal{F}$, and the question then arises: What is the dimension of this manifold? Knowing the answer to this question is not only theoretically important but it is also practically very significant because it tells us what are the "degrees of freedom" of the problem.

Chris Geyer found an elegant and beautiful answer using some rather sophisticated concepts from the theory of group actions and Lie groups (Theorem 5.10): The space $\mathcal{F}$ is

isomorphic to the quotient

$$\mathbf{O}(3,1) \times \mathbf{O}(3,1)/H_F,$$

where H_F is the stabilizer of any element, F, in $\mathcal{F}$. Now, it is easy to determine the dimension of H_F by determining the dimension of its Lie algebra, which is 3. As $\dim \mathbf{O}(3,1) = 6$, we find that $\dim \mathcal{F} = 2 \cdot 6 - 3 = 9$.

Of course, a certain amount of machinery is needed in order to understand how the above results are obtained: group actions, manifolds, Lie groups, homogenous spaces, Lorentz groups, etc. As most computer science students, even those specialized in computer vision or robotics, are not familiar with these concepts, we thought that it would be useful to give a fairly detailed exposition of these theories.

During the seminar, I also used some material from my book, Gallier [58], especially from Chapters 11, 12 and 14. Readers might find it useful to read some of this material beforehand or in parallel with these notes, especially Chapter 14, which gives a more elementary introduction to Lie groups and manifolds. For the reader's convenience, I have incorporated a slightly updated version of chapter 14 from [58] as Chapter 1 of this manuscript. In fact, during the seminar, I lectured on most of Chapter 2, but only on the "gentler" versions of Chapters 3, 5, as in [58] and not at all on Chapter 7, which was written after the course had ended.

One feature worth pointing out is that we give a complete proof of the surjectivity of the exponential map, $\exp\colon \mathfrak{so}(1,3) \to \mathbf{SO}_0(1,3)$, for the Lorentz group $\mathbf{SO}_0(3,1)$ (see Section 5.5, Theorem 5.22). Although we searched the literature quite thoroughly, we did not find a proof of this specific fact (the physics books we looked at, even the most reputable ones, seem to take this fact as obvious and there are also wrong proofs, see the Remark following Theorem 2.6). We are aware of two proofs of the surjectivity of $\exp\colon \mathfrak{so}(1,n) \to \mathbf{SO}_0(1,n)$ in the general case where where n is arbitrary: One due to Nishikawa [118] (1983) and an earlier one due to Marcel Riesz [126] (1957). In both cases, the proof is quite involved (40 pages or so). In the case of $\mathbf{SO}_0(1,3)$, a much simpler argument can be made using the fact that $\varphi\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{SO}_0(1,3)$, is surjective and that its kernel is $\{I, -I\}$ (see Proposition 5.21). Actually, a proof of this fact is not easy to find in the literature either (and, beware there are wrong proofs, again, see the Remark following Theorem 2.6). We have made sure to provide all the steps of the proof of the surjectivity of $\exp\colon \mathfrak{so}(1,3) \to \mathbf{SO}_0(1,3)$. For more on this subject, see the discussion in Section 5.5, after Corollary 5.18.

One of the "revelations" I had while on sabbatical in Nicholas' group was that many of the data that radiologists deal with (for instance, "diffusion tensors") do not live in Euclidean spaces, which are flat, but instead in more complicated curved spaces (Riemannian manifolds). As a consequence, even a notion as simple as the average of a set of data does not make sense in such spaces. Similarly, it is not clear how to define the covariance matrix of a random vector.

Pennec [120], among others, introduced a framework based on Riemannian Geometry for defining some basic statistical notions on curved spaces and gave some algorithmic methods

to compute these basic notions. Based on work in Vincent Arsigny's Ph.D. thesis, Arsigny, Fillard, Pennec and Ayache [5] introduced a new Lie group structure on the space of symmetric positive definite matrices, which allowed them to transfer strandard statistical concepts to this space (abusively called "tensors".) One of my goals in writing these notes is to provide a rather thorough background in differential geometry so that one will then be well prepared to read the above papers by Arsigny, Fillard, Pennec, Ayache and others, on statistics on manifolds.

At first, when I was writing these notes, I felt that it was important to supply most proofs. However, when I reached manifolds and differential geometry concepts, such as connections, geodesics and curvature, I realized that how formidable a task it was! Since there are lots of very good book on differential geometry, not without regrets, I decided that it was best to try to "demistify" concepts rather than fill many pages with proofs. However, when omitting a proof, I give precise pointers to the literature. In some cases where the proofs are really beautiful, as in the Theorem of Hopf and Rinow, Myers' Theorem or the Cartan-Hadamard Theorem, I could not resist to supply complete proofs!

Experienced differential geometers may be surprised and perhaps even irritated by my selection of topics. I beg their forgiveness! Primarily, I have included topics that I felt would be useful for my purposes and thus, I have omitted some topics found in all respectable differential geomety book (such as spaces of constant curvature). On the other hand, I have occasionally included topics because I found them particularly beautiful (such as characteristic classes) even though they do not seem to be of any use in medical imaging or computer vision. I also hope that readers with a more modest background will not be put off by the level of abstraction in some of the chapters and instead will be inspired to read more about these concepts, including fibre bundles!

I have also included chapters that present material having significant practical applications. These include

1. Chapter 4, on constructing manifolds from gluing data, has applications to surface reconstruction from 3D meshes,

2. Chapter 16, on spherical harmonics, has applications in computer graphics and computer vision

3. Chapter 19, on the "Log-Euclidean framework", has applications in medical imaging and

4. Chapter 21, on Clifford algebras and spinnors, has applications in robotics and computer graphics.

Of course, as anyone who attempts to write about differential geometry and Lie groups, I faced the dilemma of including or not including a chapter on differential forms. Given that our intented audience probably knows very little about them, I decided to provide a fairly

detailed treatment including a brief treatment of vector-valued differential forms. Of course, this made it necessary to review tensor products, exterior powers, *etc.*, and I have included a rather extensive chapter on this material.

I must aknowledge my debt to two of my main sources of inspiration: Berger's *Panoramic View of Riemannian Geometry* [16] and Milnor's *Morse Theory* [106]. In my opinion, Milnor's book is still one of the best references on basic differential geometry. His exposition is remarkably clear and insightful and his treatment of the variational approach to geodesics is unsurpassed. We borrowed heavily from Milnor [106]. Since Milnor's book is typeset in "ancient" typewritten format (1973!), readers might enjoy reading parts of it typeset in LaTeX. I hope that the readers of these notes will be well prepared to read standard differential geometry texts such as do Carmo [50], Gallot, Hulin, Lafontaine [60] and O'Neill [119] but also more advanced sources such as Sakai [130], Petersen [121], Jost [83], Knapp [89] and of course, Milnor [106].

Acknowledgement: I would like to thank Eugenio Calabi, Chris Croke, Ron Donagi, David Harbater, Herman Gluck, Alexander Kirillov, Steve Shatz and Wolfgand Ziller for their encouragement, advice, inspiration and for what they taught us.

Contents

Chapter 1

Introduction to Manifolds and Lie Groups

> Le rôle prépondérant de la théorie des groupes en mathématiques a été longtemps insoupçonné; il y a quatre-vingts ans, le nom même de groupe était ignoré. C'est Galois qui, le premier, en a eu une notion claire, mais c'est seulement depuis les travaux de Klein et surtout de Lie que l'on a commencé à voir qu'il n'y a presque aucune théorie mathématique où cette notion ne tienne une place importante.
>
> —**Henri Poincaré**

1.1 The Exponential Map

The purpose of this chapter is to give a "gentle" and fairly concrete introduction to manifolds, Lie groups and Lie algebras, our main objects of study.

Most texts on Lie groups and Lie algebras begin with prerequisites in differential geometry that are often formidable to average computer scientists (or average scientists, whatever that means!). We also struggled for a long time, trying to figure out what Lie groups and Lie algebras are all about, but this can be done! A good way to sneak into the wonderful world of Lie groups and Lie algebras is to play with explicit matrix groups such as the group of rotations in $\mathbb{R}^2$ (or $\mathbb{R}^3$) and with the exponential map. After actually computing the exponential $A = e^B$ of a 2×2 skew symmetric matrix B and observing that it is a rotation matrix, and similarly for a 3×3 skew symmetric matrix B, one begins to suspect that there is something deep going on. Similarly, after the discovery that every real invertible $n \times n$ matrix A can be written as $A = RP$, where R is an orthogonal matrix and P is a positive definite symmetric matrix, and that P can be written as $P = e^S$ for some symmetric matrix S, one begins to appreciate the exponential map.

Our goal in this chapter is to give an elementary and concrete introduction to Lie groups and Lie algebras by studying a number of the so-called *classical groups*, such as the general linear group $\mathbf{GL}(n, \mathbb{R})$, the special linear group $\mathbf{SL}(n, \mathbb{R})$, the orthogonal group $\mathbf{O}(n)$, the

special orthogonal group $\mathbf{SO}(n)$, and the group of affine rigid motions $\mathbf{SE}(n)$, and their Lie algebras $\mathfrak{gl}(n,\mathbb{R})$ (all matrices), $\mathfrak{sl}(n,\mathbb{R})$ (matrices with null trace), $\mathfrak{o}(n)$, and $\mathfrak{so}(n)$ (skew symmetric matrices). Now, Lie groups are at the same time, groups, topological spaces and manifolds, so we will also have to introduce the crucial notion of a *manifold*.

The inventors of Lie groups and Lie algebras (starting with Lie!) regarded Lie groups as groups of symmetries of various topological or geometric objects. Lie algebras were viewed as the "infinitesimal transformations" associated with the symmetries in the Lie group. For example, the group $\mathbf{SO}(n)$ of rotations is the group of orientation-preserving isometries of the Euclidean space $\mathbb{E}^n$. The Lie algebra $\mathfrak{so}(n,\mathbb{R})$ consisting of real skew symmetric $n \times n$ matrices is the corresponding set of infinitesimal rotations. The geometric link between a Lie group and its Lie algebra is the fact that the Lie algebra can be viewed as the tangent space to the Lie group at the identity. There is a map from the tangent space to the Lie group, called the *exponential map*. The Lie algebra can be considered as a linearization of the Lie group (near the identity element), and the exponential map provides the "delinearization," i.e., it takes us back to the Lie group. These concepts have a concrete realization in the case of groups of matrices and, for this reason, we begin by studying the behavior of the exponential maps on matrices.

We begin by defining the exponential map on matrices and proving some of its properties. The exponential map allows us to "linearize" certain algebraic properties of matrices. It also plays a crucial role in the theory of linear differential equations with constant coefficients. But most of all, as we mentioned earlier, it is a stepping stone to Lie groups and Lie algebras. On the way to Lie algebras, we derive the classical "Rodrigues-like" formulae for rotations and for rigid motions in $\mathbb{R}^2$ and $\mathbb{R}^3$. We give an elementary proof that the exponential map is surjective for both $\mathbf{SO}(n)$ and $\mathbf{SE}(n)$, not using any topology, just certain normal forms for matrices (see Gallier [58], Chapters 11 and 12).

The last section gives a quick introduction to manifolds, Lie groups and Lie algebras. Rather than defining abstract manifolds in terms of charts, atlases, *etc.*, we consider the special case of embedded submanifolds of $\mathbb{R}^N$. This approach has the pedagogical advantage of being more concrete since it uses parametrizations of subsets of $\mathbb{R}^N$, which should be familiar to the reader in the case of curves and surfaces. The general definition of a manifold will be given in Chapter 3.

Also, rather than defining Lie groups in full generality, we define linear Lie groups using the famous result of Cartan (apparently actually due to Von Neumann) that a closed subgroup of $\mathbf{GL}(n,\mathbb{R})$ is a manifold, and thus a Lie group. This way, Lie algebras can be "computed" using tangent vectors to curves of the form $t \mapsto A(t)$, where $A(t)$ is a matrix. This section is inspired from Artin [7], Chevalley [34], Marsden and Ratiu [102], Curtis [38], Howe [80], and Sattinger and Weaver [134].

Given an $n \times n$ (real or complex) matrix $A = (a_{i,\,j})$, we would like to define the exponential

e^A of A as the sum of the series

$$e^A = I_n + \sum_{p\geq 1} \frac{A^p}{p!} = \sum_{p\geq 0} \frac{A^p}{p!},$$

letting $A^0 = I_n$. The problem is, Why is it well-defined? The following lemma shows that the above series is indeed absolutely convergent.

Lemma 1.1 *Let $A = (a_{i\,j})$ be a (real or complex) $n \times n$ matrix, and let*

$$\mu = \max\{|a_{i\,j}| \mid 1 \leq i, j \leq n\}.$$

If $A^p = (a_{i\,j}^{(p)})$, then

$$\left|a_{i\,j}^{(p)}\right| \leq (n\mu)^p$$

for all i, j, $1 \leq i, j \leq n$. As a consequence, the n^2 series

$$\sum_{p\geq 0} \frac{a_{i\,j}^{(p)}}{p!}$$

converge absolutely, and the matrix

$$e^A = \sum_{p\geq 0} \frac{A^p}{p!}$$

is a well-defined matrix.

Proof. The proof is by induction on p. For $p = 0$, we have $A^0 = I_n$, $(n\mu)^0 = 1$, and the lemma is obvious. Assume that

$$|a_{i\,j}^{(p)}| \leq (n\mu)^p$$

for all i, j, $1 \leq i, j \leq n$. Then we have

$$\left|a_{i\,j}^{(p+1)}\right| = \left|\sum_{k=1}^{n} a_{i\,k}^{(p)} a_{k\,j}\right| \leq \sum_{k=1}^{n} \left|a_{i\,k}^{(p)}\right| |a_{k\,j}| \leq \mu \sum_{k=1}^{n} \left|a_{i\,k}^{(p)}\right| \leq n\mu(n\mu)^p = (n\mu)^{p+1},$$

for all i, j, $1 \leq i, j \leq n$. For every pair (i, j) such that $1 \leq i, j \leq n$, since

$$\left|a_{i\,j}^{(p)}\right| \leq (n\mu)^p,$$

the series

$$\sum_{p\geq 0} \frac{\left|a_{i\,j}^{(p)}\right|}{p!}$$

is bounded by the convergent series

$$e^{n\mu} = \sum_{p\geq 0} \frac{(n\mu)^p}{p!},$$

and thus it is absolutely convergent. This shows that

$$e^A = \sum_{k\geq 0} \frac{A^k}{k!}$$

is well defined. □

It is instructive to compute explicitly the exponential of some simple matrices. As an example, let us compute the exponential of the real skew symmetric matrix

$$A = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}.$$

We need to find an inductive formula expressing the powers A^n. Let us observe that

$$\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \theta \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}^2 = -\theta^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Then, letting

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

we have

$$\begin{aligned} A^{4n} &= \theta^{4n} I_2, \\ A^{4n+1} &= \theta^{4n+1} J, \\ A^{4n+2} &= -\theta^{4n+2} I_2, \\ A^{4n+3} &= -\theta^{4n+3} J, \end{aligned}$$

and so

$$e^A = I_2 + \frac{\theta}{1!}J - \frac{\theta^2}{2!}I_2 - \frac{\theta^3}{3!}J + \frac{\theta^4}{4!}I_2 + \frac{\theta^5}{5!}J - \frac{\theta^6}{6!}I_2 - \frac{\theta^7}{7!}J + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \frac{\theta^6}{6!} + \cdots\right) I_2 + \left(\frac{\theta}{1!} - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \cdots\right) J.$$

We recognize the power series for $\cos\theta$ and $\sin\theta$, and thus

$$e^A = \cos\theta I_2 + \sin\theta J,$$

that is

$$e^A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Thus, e^A is a rotation matrix! This is a general fact. If A is a skew symmetric matrix, then e^A is an orthogonal matrix of determinant $+1$, i.e., a rotation matrix. Furthermore, every rotation matrix is of this form; i.e., the exponential map from the set of skew symmetric matrices to the set of rotation matrices is surjective. In order to prove these facts, we need to establish some properties of the exponential map. But before that, let us work out another example showing that the exponential map is not always surjective. Let us compute the exponential of a real 2×2 matrix with null trace of the form

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

We need to find an inductive formula expressing the powers A^n. Observe that

$$A^2 = (a^2 + bc)I_2 = -\det(A)I_2.$$

If $a^2 + bc = 0$, we have

$$e^A = I_2 + A.$$

If $a^2 + bc < 0$, let $\omega > 0$ be such that $\omega^2 = -(a^2 + bc)$. Then, $A^2 = -\omega^2 I_2$. We get

$$e^A = I_2 + \frac{A}{1!} - \frac{\omega^2}{2!}I_2 - \frac{\omega^2}{3!}A + \frac{\omega^4}{4!}I_2 + \frac{\omega^4}{5!}A - \frac{\omega^6}{6!}I_2 - \frac{\omega^6}{7!}A + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 - \frac{\omega^2}{2!} + \frac{\omega^4}{4!} - \frac{\omega^6}{6!} + \cdots\right) I_2 + \frac{1}{\omega}\left(\omega - \frac{\omega^3}{3!} + \frac{\omega^5}{5!} - \frac{\omega^7}{7!} + \cdots\right) A.$$

We recognize the power series for $\cos\omega$ and $\sin\omega$, and thus

$$e^A = \cos\omega\, I_2 + \frac{\sin\omega}{\omega}A.$$

If $a^2 + bc > 0$, let $\omega > 0$ be such that $\omega^2 = (a^2 + bc)$. Then $A^2 = \omega^2 I_2$. We get

$$e^A = I_2 + \frac{A}{1!} + \frac{\omega^2}{2!}I_2 + \frac{\omega^2}{3!}A + \frac{\omega^4}{4!}I_2 + \frac{\omega^4}{5!}A + \frac{\omega^6}{6!}I_2 + \frac{\omega^6}{7!}A + \cdots.$$

Rearranging the order of the terms, we have

$$e^A = \left(1 + \frac{\omega^2}{2!} + \frac{\omega^4}{4!} + \frac{\omega^6}{6!} + \cdots\right) I_2 + \frac{1}{\omega}\left(\omega + \frac{\omega^3}{3!} + \frac{\omega^5}{5!} + \frac{\omega^7}{7!} + \cdots\right) A.$$

If we recall that $\cosh\omega = \left(e^{\omega} + e^{-\omega}\right)/2$ and $\sinh\omega = \left(e^{\omega} - e^{-\omega}\right)/2$, we recognize the power series for $\cosh\omega$ and $\sinh\omega$, and thus

$$e^A = \cosh\omega\, I_2 + \frac{\sinh\omega}{\omega} A.$$

It immediately verified that in all cases,

$$\det\left(e^A\right) = 1.$$

This shows that the exponential map is a function from the set of 2×2 matrices with null trace to the set of 2×2 matrices with determinant 1. This function is not surjective. Indeed, $tr(e^A) = 2\cos\omega$ when $a^2 + bc < 0$, $tr(e^A) = 2\cosh\omega$ when $a^2 + bc > 0$, and $tr(e^A) = 2$ when $a^2 + bc = 0$. As a consequence, for any matrix A with null trace,

$$\operatorname{tr}\left(e^A\right) \geq -2,$$

and any matrix B with determinant 1 and whose trace is less than -2 is not the exponential e^A of any matrix A with null trace. For example,

$$B = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix},$$

where $a < 0$ and $a \neq -1$, is not the exponential of any matrix A with null trace.

A fundamental property of the exponential map is that if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then the eigenvalues of e^A are $e^{\lambda_1}, \ldots, e^{\lambda_n}$. For this we need two lemmas.

Lemma 1.2 *Let A and U be (real or complex) matrices, and assume that U is invertible. Then*

$$e^{UAU^{-1}} = Ue^AU^{-1}.$$

Proof. A trivial induction shows that

$$UA^pU^{-1} = (UAU^{-1})^p,$$

and thus

$$\begin{aligned} e^{UAU^{-1}} &= \sum_{p\geq 0} \frac{(UAU^{-1})^p}{p!} = \sum_{p\geq 0} \frac{UA^pU^{-1}}{p!} \\ &= U\left(\sum_{p\geq 0} \frac{A^p}{p!}\right) U^{-1} = Ue^AU^{-1}. \end{aligned}$$

□

Say that a square matrix A is an *upper triangular matrix* if it has the following shape,

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & a_{1\,3} & \dots & a_{1\,n-1} & a_{1\,n} \\ 0 & a_{2\,2} & a_{2\,3} & \dots & a_{2\,n-1} & a_{2\,n} \\ 0 & 0 & a_{3\,3} & \dots & a_{3\,n-1} & a_{3\,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1\,n-1} & a_{n-1\,n} \\ 0 & 0 & 0 & \dots & 0 & a_{n\,n} \end{pmatrix},$$

i.e., $a_{i\,j} = 0$ whenever $j < i$, $1 \leq i, j \leq n$.

Lemma 1.3 *Given any complex $n \times n$ matrix A, there is an invertible matrix P and an upper triangular matrix T such that*

$$A = PTP^{-1}.$$

Proof. We prove by induction on n that if $f\colon \mathbb{C}^n \to \mathbb{C}^n$ is a linear map, then there is a basis $(u_1, \dots, u_n)$ with respect to which f is represented by an upper triangular matrix. For $n = 1$ the result is obvious. If $n > 1$, since $\mathbb{C}$ is algebraically closed, f has some eigenvalue $\lambda_1 \in \mathbb{C}$, and let u_1 be an eigenvector for λ_1. We can find $n-1$ vectors $(v_2, \dots, v_n)$ such that $(u_1, v_2, \dots, v_n)$ is a basis of $\mathbb{C}^n$, and let W be the subspace of dimension $n-1$ spanned by $(v_2, \dots, v_n)$. In the basis $(u_1, v_2 \dots, v_n)$, the matrix of f is of the form

$$\begin{pmatrix} a_{1\,1} & a_{1\,2} & \dots & a_{1\,n} \\ 0 & a_{2\,2} & \dots & a_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n\,2} & \dots & a_{n\,n} \end{pmatrix},$$

since its first column contains the coordinates of $\lambda_1 u_1$ over the basis $(u_1, v_2, \dots, v_n)$. Letting $p\colon \mathbb{C}^n \to W$ be the projection defined such that $p(u_1) = 0$ and $p(v_i) = v_i$ when $2 \leq i \leq n$, the linear map $g\colon W \to W$ defined as the restriction of $p \circ f$ to W is represented by the $(n-1) \times (n-1)$ matrix $(a_{i\,j})_{2 \leq i,j \leq n}$ over the basis $(v_2, \dots, v_n)$. By the induction hypothesis, there is a basis $(u_2, \dots, u_n)$ of W such that g is represented by an upper triangular matrix $(b_{i\,j})_{1 \leq i,j \leq n-1}$.

However,

$$\mathbb{C}^n = \mathbb{C}u_1 \oplus W,$$

and thus $(u_1, \dots, u_n)$ is a basis for $\mathbb{C}^n$. Since p is the projection from $\mathbb{C}^n = \mathbb{C}u_1 \oplus W$ onto W and $g\colon W \to W$ is the restriction of $p \circ f$ to W, we have

$$f(u_1) = \lambda_1 u_1$$

and

$$f(u_{i+1}) = a_{1\,i}u_1 + \sum_{j=1}^{n-1} b_{i\,j}u_{j+1}$$

for some $a_{1\,i} \in \mathbb{C}$, when $1 \leq i \leq n-1$. But then the matrix of f with respect to $(u_1, \ldots, u_n)$ is upper triangular. Thus, there is a change of basis matrix P such that $A = PTP^{-1}$ where T is upper triangular. $\square$

Remark: If E is a Hermitian space, the proof of Lemma 1.3 can be easily adapted to prove that there is an *orthonormal* basis $(u_1, \ldots, u_n)$ with respect to which the matrix of f is upper triangular. In terms of matrices, this means that there is a unitary matrix U and an upper triangular matrix T such that $A = UTU^*$. This is usually known as *Schur's lemma*. Using this result, we can immediately rederive the fact that if A is a Hermitian matrix, then there is a unitary matrix U and a real diagonal matrix D such that $A = UDU^*$.

If $A = PTP^{-1}$ where T is upper triangular, note that the diagonal entries on T are the eigenvalues $\lambda_1, \ldots, \lambda_n$ of A. Indeed, A and T have the same characteristic polynomial. This is because if A and B are any two matrices such that $A = PBP^{-1}$, then

$$\begin{aligned}
\det(A - \lambda\, I) &= \det(PBP^{-1} - \lambda\, P\, IP^{-1}),\\
&= \det(P(B - \lambda\, I)P^{-1}),\\
&= \det(P)\det(B - \lambda\, I)\det(P^{-1}),\\
&= \det(P)\det(B - \lambda\, I)\det(P)^{-1},\\
&= \det(B - \lambda\, I).
\end{aligned}$$

Furthermore, it is well known that the determinant of a matrix of the form

$$\begin{pmatrix}
\lambda_1 - \lambda & a_{1\,2} & a_{1\,3} & \ldots & a_{1\,n-1} & a_{1\,n}\\
0 & \lambda_2 - \lambda & a_{2\,3} & \ldots & a_{2\,n-1} & a_{2\,n}\\
0 & 0 & \lambda_3 - \lambda & \ldots & a_{3\,n-1} & a_{3\,n}\\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\
0 & 0 & 0 & \ldots & \lambda_{n-1} - \lambda & a_{n-1\,n}\\
0 & 0 & 0 & \ldots & 0 & \lambda_n - \lambda
\end{pmatrix}$$

is $(\lambda_1 - \lambda)\cdots(\lambda_n - \lambda)$, and thus the eigenvalues of $A = PTP^{-1}$ are the diagonal entries of T. We use this property to prove the following lemma.

Lemma 1.4 *Given any complex $n \times n$ matrix A, if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A. Furthermore, if u is an eigenvector of A for λ_i, then u is an eigenvector of e^A for e^{λ_i}.*

Proof. By Lemma 1.3 there is an invertible matrix P and an upper triangular matrix T such that

$$A = PTP^{-1}.$$

By Lemma 1.2,

$$e^{PTP^{-1}} = Pe^{T}P^{-1}.$$

However, we showed that A and T have the same eigenvalues, which are the diagonal entries $\lambda_1, \ldots, \lambda_n$ of T, and $e^A = e^{PTP^{-1}} = Pe^TP^{-1}$ and e^T have the same eigenvalues, which are the diagonal entries of e^T. Clearly, the diagonal entries of e^T are $e^{\lambda_1}, \ldots, e^{\lambda_n}$. Now, if u is an eigenvector of A for the eigenvalue λ, a simple induction shows that u is an eigenvector of A^n for the eigenvalue λ^n, from which is follows that u is an eigenvector of e^A for e^λ. □

As a consequence, we can show that

$$\det(e^A) = e^{\mathrm{tr}(A)},$$

where $\mathrm{tr}(A)$ is the *trace of* A, i.e., the sum $a_{1\,1} + \cdots + a_{n\,n}$ of its diagonal entries, which is also equal to the sum of the eigenvalues of A. This is because the determinant of a matrix is equal to the product of its eigenvalues, and if $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A, then by Lemma 1.4, $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^A, and thus

$$\det\left(e^A\right) = e^{\lambda_1} \cdots e^{\lambda_n} = e^{\lambda_1 + \cdots + \lambda_n} = e^{\mathrm{tr}(A)}.$$

This shows that e^A is always an invertible matrix, since e^z is never null for every $z \in \mathbb{C}$. In fact, the inverse of e^A is e^{-A}, but we need to prove another lemma. This is because it is generally not true that

$$e^{A+B} = e^A e^B,$$

unless A and B commute, i.e., $AB = BA$. We need to prove this last fact.

Lemma 1.5 *Given any two complex $n \times n$ matrices A, B, if $AB = BA$, then*

$$e^{A+B} = e^A e^B.$$

Proof. Since $AB = BA$, we can expand $(A + B)^p$ using the binomial formula:

$$(A+B)^p = \sum_{k=0}^{p} \binom{p}{k} A^k B^{p-k},$$

and thus

$$\frac{1}{p!}(A+B)^p = \sum_{k=0}^{p} \frac{A^k B^{p-k}}{k!(p-k)!}.$$

Note that for any integer $N \geq 0$, we can write

$$\begin{aligned}\sum_{p=0}^{2N} \frac{1}{p!}(A+B)^p &= \sum_{p=0}^{2N}\sum_{k=0}^{p} \frac{A^k B^{p-k}}{k!(p-k)!} \\ &= \left(\sum_{p=0}^{N} \frac{A^p}{p!}\right)\left(\sum_{p=0}^{N} \frac{B^p}{p!}\right) + \sum_{\substack{\max(k,l) > N \\ k+l \leq 2N}} \frac{A^k}{k!}\frac{B^l}{l!},\end{aligned}$$

where there are $N(N+1)$ pairs (k, l) in the second term. Letting

$$\|A\| = \max\{|a_{i\,j}| \mid 1 \leq i, j \leq n\}, \quad \|B\| = \max\{|b_{i\,j}| \mid 1 \leq i, j \leq n\},$$

and $\mu = \max(\|A\|, \|B\|)$, note that for every entry $c_{i\,j}$ in $\left(A^k/k!\right)\left(B^l/l!\right)$ we have

$$|c_{i\,j}| \leq n\frac{(n\mu)^k}{k!}\frac{(n\mu)^l}{l!} \leq \frac{(n^2\mu)^{2N}}{N!}.$$

As a consequence, the absolute value of every entry in

$$\sum_{\substack{\max(k,l) > N \\ k+l \leq 2N}} \frac{A^k}{k!}\frac{B^l}{l!}$$

is bounded by

$$N(N+1)\frac{(n^2\mu)^{2N}}{N!},$$

which goes to 0 as $N \mapsto \infty$. From this, it immediately follows that

$$e^{A+B} = e^A e^B.$$

□

Now, using Lemma 1.5, since A and $-A$ commute, we have

$$e^A e^{-A} = e^{A+-A} = e^{0_n} = I_n,$$

which shows that the inverse of e^A is e^{-A}.

We will now use the properties of the exponential that we have just established to show how various matrices can be represented as exponentials of other matrices.

1.2 The Lie Groups $\mathbf{GL}(n,\mathbb{R})$, $\mathbf{SL}(n,\mathbb{R})$, $\mathbf{O}(n)$, $\mathbf{SO}(n)$, the Lie Algebras $\mathfrak{gl}(n,\mathbb{R})$, $\mathfrak{sl}(n,\mathbb{R})$, $\mathfrak{o}(n)$, $\mathfrak{so}(n)$, and the Exponential Map

First, we recall some basic facts and definitions. The set of real invertible $n \times n$ matrices forms a group under multiplication, denoted by $\mathbf{GL}(n,\mathbb{R})$. The subset of $\mathbf{GL}(n,\mathbb{R})$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{GL}(n,\mathbb{R})$, denoted by $\mathbf{SL}(n,\mathbb{R})$. It is also easy to check that the set of real $n \times n$ orthogonal matrices forms a group under multiplication, denoted by $\mathbf{O}(n)$. The subset of $\mathbf{O}(n)$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{O}(n)$, denoted by $\mathbf{SO}(n)$. We will also call matrices in $\mathbf{SO}(n)$ *rotation matrices*. Staying with easy things, we can check that the set of real $n \times n$ matrices with null trace forms a vector space under addition, and similarly for the set of skew symmetric matrices.

Definition 1.1 The group $\mathbf{GL}(n,\mathbb{R})$ is called the *general linear group*, and its subgroup $\mathbf{SL}(n,\mathbb{R})$ is called the *special linear group*. The group $\mathbf{O}(n)$ of orthogonal matrices is called the *orthogonal group*, and its subgroup $\mathbf{SO}(n)$ is called the *special orthogonal group* (or *group of rotations*). The vector space of real $n \times n$ matrices with null trace is denoted by $\mathfrak{sl}(n,\mathbb{R})$, and the vector space of real $n \times n$ skew symmetric matrices is denoted by $\mathfrak{so}(n)$.

Remark: The notation $\mathfrak{sl}(n,\mathbb{R})$ and $\mathfrak{so}(n)$ is rather strange and deserves some explanation. The groups $\mathbf{GL}(n,\mathbb{R})$, $\mathbf{SL}(n,\mathbb{R})$, $\mathbf{O}(n)$, and $\mathbf{SO}(n)$ are more than just groups. They are also topological groups, which means that they are topological spaces (viewed as subspaces of $\mathbb{R}^{n^2}$) and that the multiplication and the inverse operations are continuous (in fact, smooth). Furthermore, they are smooth real manifolds.[1] Such objects are called *Lie groups*. The real vector spaces $\mathfrak{sl}(n)$ and $\mathfrak{so}(n)$ are what is called *Lie algebras*. However, we have not defined the algebra structure on $\mathfrak{sl}(n,\mathbb{R})$ and $\mathfrak{so}(n)$ yet. The algebra structure is given by what is called the *Lie bracket*, which is defined as

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

Lie algebras are associated with Lie groups. What is going on is that the Lie algebra of a Lie group is its tangent space at the identity, i.e., the space of all tangent vectors at the identity (in this case, I_n). In some sense, the Lie algebra achieves a "linearization" of the Lie group. The exponential map is a map from the Lie algebra to the Lie group, for example,

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

and

$$\exp\colon \mathfrak{sl}(n,\mathbb{R}) \to \mathbf{SL}(n,\mathbb{R}).$$

[1] We refrain from defining manifolds right now, not to interupt the flow of intuitive ideas.

The exponential map often allows a parametrization of the Lie group elements by simpler objects, the Lie algebra elements.

One might ask, What happened to the Lie algebras $\mathfrak{gl}(n,\mathbb{R})$ and $\mathfrak{o}(n)$ associated with the Lie groups $\mathbf{GL}(n,\mathbb{R})$ and $\mathbf{O}(n)$? We will see later that $\mathfrak{gl}(n,\mathbb{R})$ is the set of *all* real $n \times n$ matrices, and that $\mathfrak{o}(n) = \mathfrak{so}(n)$.

The properties of the exponential map play an important role in studying a Lie group. For example, it is clear that the map

$$\exp\colon \mathfrak{gl}(n,\mathbb{R}) \to \mathbf{GL}(n,\mathbb{R})$$

is well-defined, but since every matrix of the form e^A has a positive determinant, exp is not surjective. Similarly, since

$$\det(e^A) = e^{\mathrm{tr}(A)},$$

the map

$$\exp\colon \mathfrak{sl}(n,\mathbb{R}) \to \mathbf{SL}(n,\mathbb{R})$$

is well-defined. However, we showed in Section 1.1 that it is not surjective either. As we will see in the next theorem, the map

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is well-defined and surjective. The map

$$\exp\colon \mathfrak{o}(n) \to \mathbf{O}(n)$$

is well-defined, but it is not surjective, since there are matrices in $\mathbf{O}(n)$ with determinant -1.

Remark: The situation for matrices over the field $\mathbb{C}$ of complex numbers is quite different, as we will see later.

We now show the fundamental relationship between $\mathbf{SO}(n)$ and $\mathfrak{so}(n)$.

Theorem 1.6 *The exponential map*

$$\exp\colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is well-defined and surjective.

Proof. First, we need to prove that if A is a skew symmetric matrix, then e^A is a rotation matrix. For this, first check that

$$\left(e^A\right)^\top = e^{A^\top}.$$

Then, since $A^\top = -A$, we get

$$\left(e^A\right)^\top = e^{A^\top} = e^{-A},$$

and so

$$\left(e^A\right)^\top e^A = e^{-A}e^A = e^{-A+A} = e^{0_n} = I_n,$$

and similarly,

$$e^A \left(e^A\right)^\top = I_n,$$

showing that e^A is orthogonal. Also,

$$\det\left(e^A\right) = e^{\mathrm{tr}(A)},$$

and since A is real skew symmetric, its diagonal entries are 0, i.e., $\mathrm{tr}(A) = 0$, and so $\det(e^A) = +1$.

For the surjectivity, we will use Theorem 11.4.4 and Theorem 11.4.5, from Chapter 11 of Gallier [58]. Theorem 11.4.4 says that for every skew symmetric matrix A there is an orthogonal matrix P such that $A = PD\,P^\top$, where D is a block diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & & \cdots & \\ & D_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & D_p \end{pmatrix}$$

such that each block D_i is either 0 or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

where $\theta_i \in \mathbb{R}$, with $\theta_i > 0$. Theorem 11.4.5 says that for every orthogonal matrix R there is an orthogonal matrix P such that $R = PE\,P^\top$, where E is a block diagonal matrix of the form

$$E = \begin{pmatrix} E_1 & & \cdots & \\ & E_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & E_p \end{pmatrix}$$

such that each block E_i is either 1, -1, or a two-dimensional matrix of the form

$$E_i = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}.$$

If R is a rotation matrix, there is an even number of -1's and they can be grouped into

blocks of size 2 associated with $\theta = \pi$. Let D be the block matrix associated with E in the obvious way (where an entry 1 in E is associated with a 0 in D). Since by Lemma 1.2

$$e^A = e^{PDP^{-1}} = Pe^D P^{-1},$$

and since D is a block diagonal matrix, we can compute e^D by computing the exponentials of its blocks. If $D_i = 0$, we get $E_i = e^0 = +1$, and if

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix},$$

we showed earlier that

$$e^{D_i} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix},$$

exactly the block E_i. Thus, $E = e^D$, and as a consequence,

$$e^A = e^{PDP^{-1}} = Pe^D P^{-1} = PEP^{-1} = PE\,P^\top = R.$$

This shows the surjectivity of the exponential. □

When $n = 3$ (and A is skew symmetric), it is possible to work out an explicit formula for e^A. For any 3×3 real skew symmetric matrix

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

letting $\theta = \sqrt{a^2 + b^2 + c^2}$ and

$$B = \begin{pmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{pmatrix},$$

we have the following result known as *Rodrigues's formula* (1840).

Lemma 1.7 *The exponential map* $\exp\colon \mathfrak{so}(3) \to \mathbf{SO}(3)$ *is given by*

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

or, equivalently, by

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

if $\theta \neq 0$, *with* $e^{0_3} = I_3$.

Proof sketch. First, prove that

$$\begin{aligned} A^2 &= -\theta^2 I + B, \\ AB &= BA = 0. \end{aligned}$$

From the above, deduce that

$$A^3 = -\theta^2 A,$$

and for any $k \geq 0$,

$$\begin{aligned} A^{4k+1} &= \theta^{4k} A, \\ A^{4k+2} &= \theta^{4k} A^2, \\ A^{4k+3} &= -\theta^{4k+2} A, \\ A^{4k+4} &= -\theta^{4k+2} A^2. \end{aligned}$$

Then prove the desired result by writing the power series for e^A and regrouping terms so that the power series for cos and sin show up. □

The above formulae are the well-known formulae expressing a rotation of axis specified by the vector (a, b, c) and angle θ. Since the exponential is surjective, it is possible to write down an explicit formula for its inverse (but it is a multivalued function!). This has applications in kinematics, robotics, and motion interpolation.

1.3 Symmetric Matrices, Symmetric Positive Definite Matrices, and the Exponential Map

Recall that a real symmetric matrix is called *positive* (or *positive semidefinite*) if its eigenvalues are all positive or null, and *positive definite* if its eigenvalues are all strictly positive. We denote the vector space of real symmetric $n \times n$ matrices by $\mathbf{S}(n)$, the set of symmetric positive matrices by $\mathbf{SP}(n)$, and the set of symmetric positive definite matrices by $\mathbf{SPD}(n)$.

The next lemma shows that every symmetric positive definite matrix A is of the form e^B for some unique symmetric matrix B. The set of symmetric matrices is a vector space, but it is not a Lie algebra because the Lie bracket $[A, B]$ is not symmetric unless A and B commute, and the set of symmetric (positive) definite matrices is not a multiplicative group, so this result is of a different flavor as Theorem 1.6.

Lemma 1.8 *For every symmetric matrix B, the matrix e^B is symmetric positive definite. For every symmetric positive definite matrix A, there is a unique symmetric matrix B such that $A = e^B$.*

Proof. We showed earlier that

$$\left(e^B\right)^\top = e^{B^\top}.$$

If B is a symmetric matrix, then since $B^\top = B$, we get

$$\left(e^B\right)^\top = e^{B^\top} = e^B,$$

and e^B is also symmetric. Since the eigenvalues $\lambda_1, \ldots, \lambda_n$ of the symmetric matrix B are real and the eigenvalues of e^B are $e^{\lambda_1}, \ldots, e^{\lambda_n}$, and since $e^\lambda > 0$ if $\lambda \in \mathbb{R}$, e^B is positive definite.

If A is symmetric positive definite, by Theorem 11.4.3 from Chapter 11 of Gallier [58], there is an orthogonal matrix P such that $A = PD\,P^\top$, where D is a diagonal matrix

$$D = \begin{pmatrix} \lambda_1 & & \cdots & \\ & \lambda_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_i > 0$, since A is positive definite. Letting

$$L = \begin{pmatrix} \log \lambda_1 & & \cdots & \\ & \log \lambda_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & \log \lambda_n \end{pmatrix},$$

it is obvious that $e^L = D$, with $\log \lambda_i \in \mathbb{R}$, since $\lambda_i > 0$.

Let

$$B = PL\,P^\top.$$

By Lemma 1.2, we have

$$e^B = e^{PL\,P^\top} = e^{PLP^{-1}} = Pe^L P^{-1} = Pe^L\,P^\top = PD\,P^\top = A.$$

Finally, we prove that if B_1 and B_2 are symmetric and $A = e^{B_1} = e^{B_2}$, then $B_1 = B_2$. Since B_1 is symmetric, there is an orthonormal basis $(u_1, \ldots, u_n)$ of eigenvectors of B_1. Let $\mu_1, \ldots, \mu_n$ be the corresponding eigenvalues. Similarly, there is an orthonormal basis $(v_1, \ldots, v_n)$ of eigenvectors of B_2. We are going to prove that B_1 and B_2 agree on the basis $(v_1, \ldots, v_n)$, thus proving that $B_1 = B_2$.

Let μ be some eigenvalue of B_2, and let $v = v_i$ be some eigenvector of B_2 associated with μ. We can write

$$v = \alpha_1 u_1 + \cdots + \alpha_n u_n.$$

Since v is an eigenvector of B_2 for μ and $A = e^{B_2}$, by Lemma 1.4

$$A(v) = e^\mu v = e^\mu \alpha_1 u_1 + \cdots + e^\mu \alpha_n u_n.$$

On the other hand,

$$A(v) = A(\alpha_1 u_1 + \cdots + \alpha_n u_n) = \alpha_1 A(u_1) + \cdots + \alpha_n A(u_n),$$

and since $A = e^{B_1}$ and $B_1(u_i) = \mu_i u_i$, by Lemma 1.4 we get

$$A(v) = e^{\mu_1}\alpha_1 u_1 + \cdots + e^{\mu_n}\alpha_n u_n.$$

Therefore, $\alpha_i = 0$ if $\mu_i \neq \mu$. Letting

$$I = \{i \mid \mu_i = \mu,\ i \in \{1, \ldots, n\}\},$$

we have

$$v = \sum_{i \in I} \alpha_i u_i.$$

Now,

$$\begin{aligned} B_1(v) &= B_1\Big(\sum_{i \in I} \alpha_i u_i\Big) = \sum_{i \in I} \alpha_i B_1(u_i) = \sum_{i \in I} \alpha_i \mu_i u_i \\ &= \sum_{i \in I} \alpha_i \mu u_i = \mu\Big(\sum_{i \in I} \alpha_i u_i\Big) = \mu v, \end{aligned}$$

since $\mu_i = \mu$ when $i \in I$. Since v is an eigenvector of B_2 for μ,

$$B_2(v) = \mu v,$$

which shows that

$$B_1(v) = B_2(v).$$

Since the above holds for every eigenvector v_i, we have $B_1 = B_2$. □

Lemma 1.8 can be reformulated as stating that the map $\exp\colon \mathbf{S}(n) \to \mathbf{SPD}(n)$ is a bijection. It can be shown that it is a homeomorphism. In the case of invertible matrices, the polar form theorem can be reformulated as stating that there is a bijection between the topological space $\mathbf{GL}(n, \mathbb{R})$ of real $n \times n$ invertible matrices (also a group) and $\mathbf{O}(n) \times \mathbf{SPD}(n)$.

As a corollary of the polar form theorem (Theorem 12.1.3 in Chapter 12 of Gallier [58]) and Lemma 1.8, we have the following result: For every invertible matrix A there is a unique orthogonal matrix R and a unique symmetric matrix S such that

$$A = R\, e^S.$$

Thus, we have a bijection between $\mathbf{GL}(n, \mathbb{R})$ and $\mathbf{O}(n) \times \mathbf{S}(n)$. But $\mathbf{S}(n)$ itself is isomorphic to $\mathbb{R}^{n(n+1)/2}$. Thus, there is a bijection between $\mathbf{GL}(n, \mathbb{R})$ and $\mathbf{O}(n) \times \mathbb{R}^{n(n+1)/2}$. It can also be shown that this bijection is a homeomorphism. This is an interesting fact. Indeed, this

homeomorphism essentially reduces the study of the topology of $\mathbf{GL}(n, \mathbb{R})$ to the study of the topology of $\mathbf{O}(n)$. This is nice, since it can be shown that $\mathbf{O}(n)$ is compact.

In $A = R\, e^S$, if $\det(A) > 0$, then R must be a rotation matrix (i.e., $\det(R) = +1$), since $\det\left(e^S\right) > 0$. In particular, if $A \in \mathbf{SL}(n, \mathbb{R})$, since $\det(A) = \det(R) = +1$, the symmetric matrix S must have a null trace, i.e., $S \in \mathbf{S}(n) \cap \mathfrak{sl}(n, \mathbb{R})$. Thus, we have a bijection between $\mathbf{SL}(n, \mathbb{R})$ and $\mathbf{SO}(n) \times (\mathbf{S}(n) \cap \mathfrak{sl}(n, \mathbb{R}))$.

We can also show that the exponential map is a surjective map from the skew Hermitian matrices to the unitary matrices (use Theorem 11.4.7 from Chapter 11 in Gallier [58]).

1.4 The Lie Groups $\mathbf{GL}(n, \mathbb{C})$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$, $\mathbf{SU}(n)$, the Lie Algebras $\mathfrak{gl}(n, \mathbb{C})$, $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, $\mathfrak{su}(n)$, and the Exponential Map

The set of complex invertible $n \times n$ matrices forms a group under multiplication, denoted by $\mathbf{GL}(n, \mathbb{C})$. The subset of $\mathbf{GL}(n, \mathbb{C})$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{GL}(n, \mathbb{C})$, denoted by $\mathbf{SL}(n, \mathbb{C})$. It is also easy to check that the set of complex $n \times n$ unitary matrices forms a group under multiplication, denoted by $\mathbf{U}(n)$. The subset of $\mathbf{U}(n)$ consisting of those matrices having determinant $+1$ is a subgroup of $\mathbf{U}(n)$, denoted by $\mathbf{SU}(n)$. We can also check that the set of complex $n \times n$ matrices with null trace forms a real vector space under addition, and similarly for the set of skew Hermitian matrices and the set of skew Hermitian matrices with null trace.

Definition 1.2 The group $\mathbf{GL}(n, \mathbb{C})$ is called the *general linear group*, and its subgroup $\mathbf{SL}(n, \mathbb{C})$ is called the *special linear group*. The group $\mathbf{U}(n)$ of unitary matrices is called the *unitary group*, and its subgroup $\mathbf{SU}(n)$ is called the *special unitary group*. The real vector space of complex $n \times n$ matrices with null trace is denoted by $\mathfrak{sl}(n, \mathbb{C})$, the real vector space of skew Hermitian matrices is denoted by $\mathfrak{u}(n)$, and the real vector space $\mathfrak{u}(n) \cap \mathfrak{sl}(n, \mathbb{C})$ is denoted by $\mathfrak{su}(n)$.

Remarks:

(1) As in the real case, the groups $\mathbf{GL}(n, \mathbb{C})$, $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$, and $\mathbf{SU}(n)$ are also topological groups (viewed as subspaces of $\mathbb{R}^{2n^2}$), and in fact, smooth real manifolds. Such objects are called *(real) Lie groups*. The real vector spaces $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$ are *Lie algebras* associated with $\mathbf{SL}(n, \mathbb{C})$, $\mathbf{U}(n)$, and $\mathbf{SU}(n)$. The algebra structure is given by the *Lie bracket*, which is defined as

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

(2) It is also possible to define complex Lie groups, which means that they are topological groups and smooth *complex* manifolds. It turns out that $\mathbf{GL}(n, \mathbb{C})$ and $\mathbf{SL}(n, \mathbb{C})$ are complex manifolds, but not $\mathbf{U}(n)$ and $\mathbf{SU}(n)$.

One should be very careful to observe that even though the Lie algebras $\mathfrak{sl}(n, \mathbb{C})$, $\mathfrak{u}(n)$, and $\mathfrak{su}(n)$ consist of matrices with complex coefficients, we view them as *real* vector spaces. The Lie algebra $\mathfrak{sl}(n, \mathbb{C})$ is also a complex vector space, but $\mathfrak{u}(n)$ and $\mathfrak{su}(n)$ are not! Indeed, if A is a skew Hermitian matrix, iA is *not* skew Hermitian, but Hermitian!

Again the Lie algebra achieves a "linearization" of the Lie group. In the complex case, the Lie algebras $\mathfrak{gl}(n, \mathbb{C})$ is the set of *all* complex $n \times n$ matrices, but $\mathfrak{u}(n) \neq \mathfrak{su}(n)$, because a skew Hermitian matrix does not necessarily have a null trace.

The properties of the exponential map also play an important role in studying complex Lie groups. For example, it is clear that the map

$$\exp\colon \mathfrak{gl}(n, \mathbb{C}) \to \mathbf{GL}(n, \mathbb{C})$$

is well-defined, but this time, it is surjective! One way to prove this is to use the Jordan normal form. Similarly, since

$$\det\left(e^A\right) = e^{\mathrm{tr}(A)},$$

the map

$$\exp\colon \mathfrak{sl}(n, \mathbb{C}) \to \mathbf{SL}(n, \mathbb{C})$$

is well-defined, but it is not surjective! As we will see in the next theorem, the maps

$$\exp\colon \mathfrak{u}(n) \to \mathbf{U}(n)$$

and

$$\exp\colon \mathfrak{su}(n) \to \mathbf{SU}(n)$$

are well-defined and surjective.

Theorem 1.9 *The exponential maps*

$$\exp\colon \mathfrak{u}(n) \to \mathbf{U}(n) \quad \textit{and} \quad \exp\colon \mathfrak{su}(n) \to \mathbf{SU}(n)$$

are well-defined and surjective.

Proof. First, we need to prove that if A is a skew Hermitian matrix, then e^A is a unitary matrix. For this, first check that

$$\left(e^A\right)^* = e^{A^*}.$$

Then, since $A^* = -A$, we get

$$\left(e^A\right)^* = e^{A^*} = e^{-A},$$

and so

$$\left(e^A\right)^* e^A = e^{-A}e^A = e^{-A+A} = e^{0_n} = I_n,$$

and similarly, $e^A \left(e^A\right)^* = I_n$, showing that e^A is unitary. Since

$$\det\left(e^A\right) = e^{\mathrm{tr}(A)},$$

if A is skew Hermitian and has null trace, then $\det(e^A) = +1$.

For the surjectivity we will use Theorem 11.4.7 in Chapter 11 of Gallier [58]. First, assume that A is a unitary matrix. By Theorem 11.4.7, there is a unitary matrix U and a diagonal matrix D such that $A = UDU^*$. Furthermore, since A is unitary, the entries $\lambda_1, \ldots, \lambda_n$ in D (the eigenvalues of A) have absolute value $+1$. Thus, the entries in D are of the form $\cos\theta + i\sin\theta = e^{i\theta}$. Thus, we can assume that D is a diagonal matrix of the form

$$D = \begin{pmatrix} e^{i\theta_1} & & \cdots & \\ & e^{i\theta_2} & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & e^{i\theta_p} \end{pmatrix}.$$

If we let E be the diagonal matrix

$$E = \begin{pmatrix} i\theta_1 & & \cdots & \\ & i\theta_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & i\theta_p \end{pmatrix}$$

it is obvious that E is skew Hermitian and that

$$e^E = D.$$

Then, letting $B = UEU^*$, we have

$$e^B = A,$$

and it is immediately verified that B is skew Hermitian, since E is.

If A is a unitary matrix with determinant $+1$, since the eigenvalues of A are $e^{i\theta_1}, \ldots, e^{i\theta_p}$ and the determinant of A is the product

$$e^{i\theta_1} \cdots e^{i\theta_p} = e^{i(\theta_1 + \cdots + \theta_p)}$$

of these eigenvalues, we must have

$$\theta_1 + \cdots + \theta_p = 0,$$

and so, E is skew Hermitian and has zero trace. As above, letting

$$B = UEU^*,$$

we have

$$e^B = A,$$

where B is skew Hermitian and has null trace. □

We now extend the result of Section 1.3 to Hermitian matrices.

1.5 Hermitian Matrices, Hermitian Positive Definite Matrices, and the Exponential Map

Recall that a Hermitian matrix is called *positive* (or *positive semidefinite*) if its eigenvalues are all positive or null, and *positive definite* if its eigenvalues are all strictly positive. We denote the real vector space of Hermitian $n \times n$ matrices by $\mathbf{H}(n)$, the set of Hermitian positive matrices by $\mathbf{HP}(n)$, and the set of Hermitian positive definite matrices by $\mathbf{HPD}(n)$.

The next lemma shows that every Hermitian positive definite matrix A is of the form e^B for some unique Hermitian matrix B. As in the real case, the set of Hermitian matrices is a real vector space, but it is not a Lie algebra because the Lie bracket $[A, B]$ is not Hermitian unless A and B commute, and the set of Hermitian (positive) definite matrices is not a multiplicative group.

Lemma 1.10 *For every Hermitian matrix B, the matrix e^B is Hermitian positive definite. For every Hermitian positive definite matrix A, there is a unique Hermitian matrix B such that $A = e^B$.*

Proof. It is basically the same as the proof of Theorem 1.10, except that a Hermitian matrix can be written as $A = UDU^*$, where D is a real diagonal matrix and U is unitary instead of orthogonal. □

Lemma 1.10 can be reformulated as stating that the map $\exp\colon \mathbf{H}(n) \to \mathbf{HPD}(n)$ is a bijection. In fact, it can be shown that it is a homeomorphism. In the case of complex invertible matrices, the polar form theorem can be reformulated as stating that there is a bijection between the topological space $\mathbf{GL}(n, \mathbb{C})$ of complex $n \times n$ invertible matrices (also a group) and $\mathbf{U}(n) \times \mathbf{HPD}(n)$. As a corollary of the polar form theorem and Lemma 1.10, we have the following result: For every complex invertible matrix A, there is a unique unitary matrix U and a unique Hermitian matrix S such that

$$A = U\, e^S.$$

Thus, we have a bijection between $\mathbf{GL}(n,\mathbb{C})$ and $\mathbf{U}(n)\times\mathbf{H}(n)$. But $\mathbf{H}(n)$ itself is isomorphic to $\mathbb{R}^{n^2}$, and so there is a bijection between $\mathbf{GL}(n,\mathbb{C})$ and $\mathbf{U}(n)\times\mathbb{R}^{n^2}$. It can also be shown that this bijection is a homeomorphism. This is an interesting fact. Indeed, this homeomorphism essentially reduces the study of the topology of $\mathbf{GL}(n,\mathbb{C})$ to the study of the topology of $\mathbf{U}(n)$. This is nice, since it can be shown that $\mathbf{U}(n)$ is compact (as a real manifold).

In the polar decomposition $A = Ue^S$, we have $|\det(U)| = 1$, since U is unitary, and $\mathrm{tr}(S)$ is real, since S is Hermitian (since it is the sum of the eigenvalues of S, which are real), so that $\det\left(e^S\right) > 0$. Thus, if $\det(A) = 1$, we must have $\det\left(e^S\right) = 1$, which implies that $S \in \mathbf{H}(n)\cap\mathfrak{sl}(n,\mathbb{C})$. Thus, we have a bijection between $\mathbf{SL}(n,\mathbb{C})$ and $\mathbf{SU}(n)\times(\mathbf{H}(n)\cap\mathfrak{sl}(n,\mathbb{C}))$.

In the next section we study the group $\mathbf{SE}(n)$ of affine maps induced by orthogonal transformations, also called rigid motions, and its Lie algebra. We will show that the exponential map is surjective. The groups $\mathbf{SE}(2)$ and $\mathbf{SE}(3)$ play play a fundamental role in robotics, dynamics, and motion planning.

1.6 The Lie Group $\mathbf{SE}(n)$ and the Lie Algebra $\mathfrak{se}(n)$

First, we review the usual way of representing affine maps of $\mathbb{R}^n$ in terms of $(n+1)\times(n+1)$ matrices.

Definition 1.3 The set of affine maps ρ of $\mathbb{R}^n$, defined such that

$$\rho(X) = RX + U,$$

where R is a rotation matrix ($R \in \mathbf{SO}(n)$) and U is some vector in $\mathbb{R}^n$, is a group under composition called the group of *direct affine isometries, or rigid motions*, denoted by $\mathbf{SE}(n)$.

Every rigid motion can be represented by the $(n+1)\times(n+1)$ matrix

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

in the sense that

$$\begin{pmatrix} \rho(X) \\ 1 \end{pmatrix} = \begin{pmatrix} R & U \\ 0 & 1 \end{pmatrix}\begin{pmatrix} X \\ 1 \end{pmatrix}$$

iff

$$\rho(X) = RX + U.$$

Definition 1.4 The vector space of real $(n+1)\times(n+1)$ matrices of the form

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix},$$

where Ω is a skew symmetric matrix and U is a vector in $\mathbb{R}^n$, is denoted by $\mathfrak{se}(n)$.

Remark: The group $\mathbf{SE}(n)$ is a Lie group, and its Lie algebra turns out to be $\mathfrak{se}(n)$.

We will show that the exponential map $\exp\colon \mathfrak{se}(n) \to \mathbf{SE}(n)$ is surjective. First, we prove the following key lemma.

Lemma 1.11 *Given any $(n+1) \times (n+1)$ matrix of the form*

$$A = \begin{pmatrix} \Omega & U \\ 0 & 0 \end{pmatrix}$$

where Ω is any matrix and $U \in \mathbb{R}^n$,

$$A^k = \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix},$$

where $\Omega^0 = I_n$. As a consequence,

$$e^A = \begin{pmatrix} e^\Omega & VU \\ 0 & 1 \end{pmatrix},$$

where

$$V = I_n + \sum_{k \geq 1} \frac{\Omega^k}{(k+1)!}.$$

Proof. A trivial induction on k shows that

$$A^k = \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix}.$$

Then we have

$$\begin{aligned} e^A &= \sum_{k \geq 0} \frac{A^k}{k!}, \\ &= I_{n+1} + \sum_{k \geq 1} \frac{1}{k!} \begin{pmatrix} \Omega^k & \Omega^{k-1}U \\ 0 & 0 \end{pmatrix}, \\ &= \begin{pmatrix} I_n + \sum_{k \geq 0} \frac{\Omega^k}{k!} & \sum_{k \geq 1} \frac{\Omega^{k-1}}{k!} U \\ 0 & 1 \end{pmatrix}, \\ &= \begin{pmatrix} e^\Omega & VU \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

□

We can now prove our main theorem. We will need to prove that V is invertible when Ω is a skew symmetric matrix. It would be tempting to write V as

$$V = \Omega^{-1}(e^{\Omega} - I).$$

Unfortunately, for odd n, a skew symmetric matrix of order n is not invertible! Thus, we have to find another way of proving that V is invertible. However, observe that we have the following useful fact:

$$V = I_n + \sum_{k \geq 1} \frac{\Omega^k}{(k+1)!} = \int_0^1 e^{\Omega t} dt.$$

This is what we will use in Theorem 1.12 to prove surjectivity.

Theorem 1.12 *The exponential map*

$$\exp \colon \mathfrak{se}(n) \to \mathbf{SE}(n)$$

is well-defined and surjective.

Proof. Since Ω is skew symmetric, e^{Ω} is a rotation matrix, and by Theorem 1.6, the exponential map

$$\exp \colon \mathfrak{so}(n) \to \mathbf{SO}(n)$$

is surjective. Thus, it remains to prove that for every rotation matrix R, there is some skew symmetric matrix Ω such that $R = e^{\Omega}$ and

$$V = I_n + \sum_{k \geq 1} \frac{\Omega^k}{(k+1)!}$$

is invertible. By Theorem 11.4.4 in Chapter 11 of Gallier [58], for every skew symmetric matrix Ω there is an orthogonal matrix P such that $\Omega = PDP^{\top}$, where D is a block diagonal matrix of the form

$$D = \begin{pmatrix} D_1 & & \cdots & \\ & D_2 & \cdots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \cdots & D_p \end{pmatrix}$$

such that each block D_i is either 0 or a two-dimensional matrix of the form

$$D_i = \begin{pmatrix} 0 & -\theta_i \\ \theta_i & 0 \end{pmatrix}$$

where $\theta_i \in \mathbb{R}$, with $\theta_i > 0$. Actually, we can assume that $\theta_i \neq k2\pi$ for all $k \in \mathbb{Z}$, since when $\theta_i = k2\pi$ we have $e^{D_i} = I_2$, and D_i can be replaced by two one-dimensional blocks each consisting of a single zero. To compute V, since $\Omega = PD\,P^\top = PDP^{-1}$, observe that

$$\begin{aligned} V &= I_n + \sum_{k\geq 1} \frac{\Omega^k}{(k+1)!} \\ &= I_n + \sum_{k\geq 1} \frac{PD^kP^{-1}}{(k+1)!} \\ &= P\left(I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!}\right)P^{-1} \\ &= PWP^{-1}, \end{aligned}$$

where

$$W = I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!}.$$

We can compute

$$W = I_n + \sum_{k\geq 1} \frac{D^k}{(k+1)!} = \int_0^1 e^{Dt}dt,$$

by computing

$$W = \begin{pmatrix} W_1 & & \dots & \\ & W_2 & \dots & \\ \vdots & \vdots & \ddots & \vdots \\ & & \dots & W_p \end{pmatrix}$$

by blocks. Since

$$e^{D_i} = \begin{pmatrix} \cos\theta_i & -\sin\theta_i \\ \sin\theta_i & \cos\theta_i \end{pmatrix}$$

when D_i is a 2×2 skew symmetric matrix and $W_i = \int_0^1 e^{D_i t}dt$, we get

$$W_i = \begin{pmatrix} \int_0^1 \cos(\theta_i t)dt & \int_0^1 -\sin(\theta_i t)dt \\ \int_0^1 \sin(\theta_i t)dt & \int_0^1 \cos(\theta_i t)dt \end{pmatrix} = \frac{1}{\theta_i}\begin{pmatrix} \sin(\theta_i t)\,|_0^1 & \cos(\theta_i t)\,|_0^1 \\ -\cos(\theta_i t)\,|_0^1 & \sin(\theta_i t)\,|_0^1 \end{pmatrix},$$

that is,

$$W_i = \frac{1}{\theta_i}\begin{pmatrix} \sin\theta_i & -(1-\cos\theta_i) \\ 1-\cos\theta_i & \sin\theta_i \end{pmatrix},$$

and $W_i = 1$ when $D_i = 0$. Now, in the first case, the determinant is

$$\frac{1}{\theta_i^2}\left((\sin\theta_i)^2 + (1-\cos\theta_i)^2\right) = \frac{2}{\theta_i^2}(1-\cos\theta_i),$$

which is nonzero, since $\theta_i \neq k2\pi$ for all $k \in \mathbb{Z}$. Thus, each W_i is invertible, and so is W, and thus, $V = PWP^{-1}$ is invertible. $\square$

In the case $n = 3$, given a skew symmetric matrix

$$\Omega = \begin{pmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{pmatrix},$$

letting $\theta = \sqrt{a^2 + b^2 + c^2}$, it it easy to prove that if $\theta = 0$, then

$$e^A = \begin{pmatrix} I_3 & U \\ 0 & 1 \end{pmatrix},$$

and that if $\theta \neq 0$ (using the fact that $\Omega^3 = -\theta^2\Omega$), then

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

and

$$V = I_3 + \frac{(1 - \cos\theta)}{\theta^2}\Omega + \frac{(\theta - \sin\theta)}{\theta^3}\Omega^2.$$

Our next goal is to define *embedded submanifolds* and (linear) Lie groups. Before doing this, we believe that some readers might appreciate a review of the notion of the *derivative* of a function between two normed vector spaces.

1.7 The Derivative of a Function Between Normed Vector Spaces, a Review

In this brief section, we review some basic notions of differential calculus, in particular, the *derivative* of a function, $f\colon E \to F$, where E and F are normed vector spaces. In most cases, $E = \mathbb{R}^n$ and $F = \mathbb{R}^m$. However, if we need to deal with infinite dimensional manifolds, then it is necessary to allow E and F to be infinite dimensional. This section can be omitted by readers already familiar with this standard material. We omit all proofs and refer the reader to standard analysis textbooks such as Lang [94, 93], Munkres [116], Choquet-Bruhat [37] or Schwartz [135], for a complete exposition.

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 $[r, s] \subseteq A$ containing a, form a line segment 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 1.5 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 1.5 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 $E = \mathbb{R}$, $F = \mathbb{R}$ 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 1.6 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 *total derivative*, or *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.

The following proposition shows that our new definition is consistent with the definition of the directional derivative and that *the continuous linear map L is unique*, if it exists.

Proposition 1.13 *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,*

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

and thus, $\mathrm{D}f(a)$ is uniquely defined.

Proof. If $L = \mathrm{D}f(a)$ exists, then for any nonzero vector $u \in E$, because A is open, for any $t \in \mathbb{R} - \{0\}$ (or $t \in \mathbb{C} - \{0\}$) small enough, $a + tu \in A$, so

$$\begin{aligned} f(a+tu) &= f(a) + L(tu) + \epsilon(tu)\|tu\| \\ &= f(a) + tL(u) + |t|\epsilon(tu)\|u\| \end{aligned}$$

which implies that

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

and since $\lim_{t \mapsto 0} \epsilon(tu) = 0$, we deduce that

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

Because

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

for all h such that $\|h\|$ is small enough, L is continuous, and $\lim_{h \mapsto 0} \epsilon(h)\|h\| = 0$, we have $\lim_{h \mapsto 0} f(a+h) = f(a)$, that is, f is continuous at a. □

Observe that the uniqueness of $\mathrm{D}f(a)$ follows from Proposition 1.13. 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 1.7 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 Leibnitz, 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 1.14 *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 $\mathrm{D}f(a) = f$, for every $a \in E$, where f denotes the linear map associated with f.*

Proposition 1.15 *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*

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

Proposition 1.16 *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 1.17 *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 1.17 has many interesting consequences. We mention two corollaries.

Proposition 1.18 *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 1.18 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 $J(f)(a) = \left(\dfrac{\partial f_i}{\partial x_j}(a)\right)$:

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

or

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

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 1.1 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 1.8 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 1.2

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
$$J(f)(a) = \begin{pmatrix} \dfrac{\partial f}{\partial x}(a) & \dfrac{\partial f}{\partial y}(a) & \dfrac{\partial f}{\partial z}(a) \end{pmatrix}.$$

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

$$J(f)(a) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(a) & \cdots & \dfrac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$

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 1.17, 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 $(v_1, \ldots, v_n)$:

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

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 1.19 *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* $\dfrac{\partial f}{\partial x_j}$*) 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* $\partial_j f_i$ $\left(\text{or } \dfrac{\partial f_i}{\partial x_j}\right)$ *is defined and continuous on* A*, for all* i, j*,* $1 \leq i \leq m$*,* $1 \leq j \leq n$*.*

Definition 1.9 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 1.10 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 1.18, if f is a diffeomorphism on A, then $\mathrm{D}f(a)$ is a bijection for every $a \in A$.

Theorem 1.20 *(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 1.20 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.

1.8 Manifolds, Lie Groups and Lie Algebras

In this section we define precisely manifolds, Lie groups and 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, and that the loss of information is not too severe. The challenge that we are facing is that unless our readers are already familiar with manifolds, the amount of basic differential geometry required to define Lie groups and Lie algebras in full generality is overwhelming.

Fortunately, most of the Lie groups that we will consider are subspaces of $\mathbb{R}^N$ for some sufficiently large N. In fact, most of them are 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 these groups are subspaces of $\mathbb{R}^N$, in a first stage, we do not need the definition of an abstract manifold. We just have to define embedded submanifolds (also called submanifolds) of $\mathbb{R}^N$ (in the case of $\mathbf{GL}(n, \mathbb{R})$, $N = n^2$). This is the path that we will follow. The general definition of manifold will be given in Chapter 3.

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 a characterization of the linear groups that obviates much of the work. 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 (see Warner [147], Chapter 3, Theorem 3.21, page 97). 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 (see Warner [147], Chapter 3, Theorem 3.42, page 110). 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. This approach to define the Lie algebra of a matrix group is followed by a number of authors, such as Curtis [38]. However, Curtis is rather cavalier, since he does not explain why the required curves actually exist, and thus, according to his definition, Lie algebras could be the trivial vector space! Although we will not prove the theorem of Von Neumann and Cartan, we feel that it is important to make clear why the definitions make sense, i.e., why we are not dealing with trivial objects.

A small annoying technical problem will arise 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 (which means that its complement is open). Thus, discrete subgroups of $\mathbf{GL}(n, \mathbb{R})$ are Lie groups! But these are not very interesting Lie groups, and 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 detailed presentations, see DoCarmo [49, 50], Milnor [108], Marsden and Ratiu [102], Berger and Gostiaux [17], or Warner [147]. 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 set 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 that allow the definition of smooth functions on M and 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 1.11 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 point $p \in M$ 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 1.11. 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.

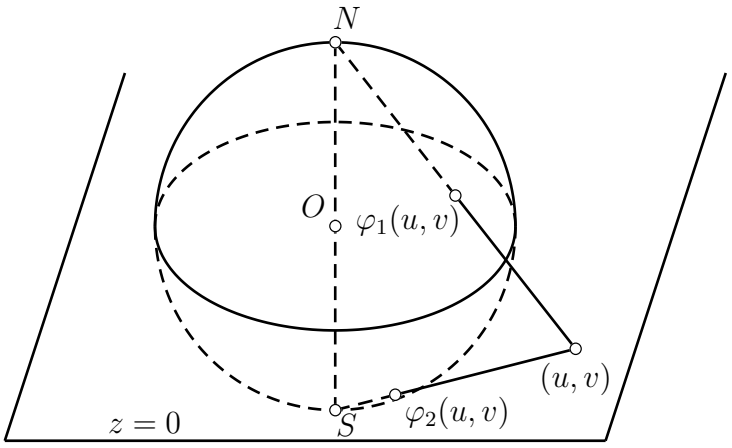

Figure 1.1: Inverse stereographic projections

Example 1.3 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 1.1. We leave as an exercise to 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. 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 1.21 *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 1.21 (see Berger and Gostiaux [17], Theorem 2.1.9 or DoCarmo [50], Chapter 0, Section 4). It is a key technical result used to show that interesting properties of maps between manifolds do not depend on parametrizations.

Lemma 1.22 *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 1.22 is illustrated in Figure 1.2.

Using Definition 1.11, 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:

Proposition 1.23 *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

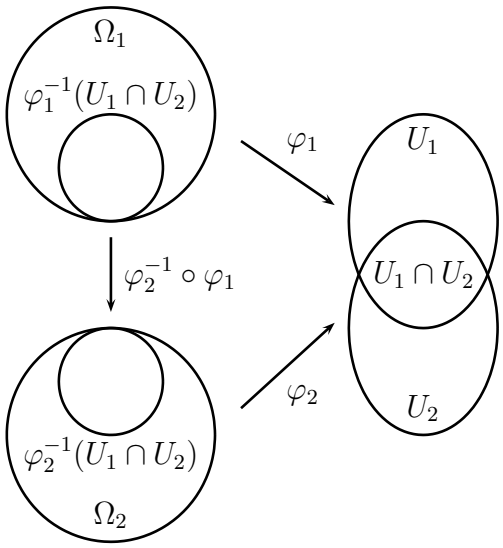

Figure 1.2: Parametrizations and transition functions

(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 $f_1^{-1}(0) = W_1 \cap f^{-1}(0) = W_1 \cap W \cap M = W_1 \cap M$.

A proof of Proposition 1.23 can be found in Lafontaine [92] or Berger and Gostiaux [17]. Lemma 1.21 and Proposition 1.23 are actually *equivalent* to Definition 1.11. This equivalence is also proved in Lafontaine [92] and Berger and Gostiaux [17].

The proof, which is somewhat illuminating, is based on two technical lemmas that are proved using the inverse function theorem (for example, see Guillemin and Pollack [69], Chapter 1, Sections 3 and 4).

Lemma 1.24 *Let $U \subseteq \mathbb{R}^m$ be an open subset of $\mathbb{R}^m$ and pick some $a \in U$. If $f\colon U \to \mathbb{R}^n$ 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, $\theta\colon V \to U' \times O$, so that*

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

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

Proof. Since f is an immersion, its Jacobian matrix, $J(f)$, (an $n \times m$ matrix) has rank m and by permuting coordinates if needed, we may assume that the first m rows of $J(f)$ are linearly independent and we let

$$A = \left(\frac{\partial f_i}{\partial x_j}(a)\right)$$

be this invertible $m \times m$ matrix. Define the map, $g\colon U \times \mathbb{R}^{n-m} \to \mathbb{R}^n$, by

$$g(x, y) = (f_1(x), \ldots, f_m(x), y_1 + f_{m+1}(x), \ldots, y_{n-m} + f_n(x)),$$

for all $x \in U$ and all $y \in \mathbb{R}^{n-m}$. The Jacobian matrix of g at $(a, 0)$ is of the form

$$J = \begin{pmatrix} A & 0 \\ B & I \end{pmatrix}$$

so $\det(J) = \det(A)\det(I) = \det(A) \neq 0$, since A is invertible. By the inverse function theorem, there are some open subsets $W \subseteq U \times \mathbb{R}^{n-m}$ with $(a, 0) \in W$ and $V \subseteq \mathbb{R}^n$ such that the restriction of g to W is a diffeomorphism between W and V. Since $W \subseteq U \times \mathbb{R}^{n-m}$ is an open set, we can find some open subsets $U' \subseteq U$ and $O \subseteq \mathbb{R}^{n-m}$ so that $U' \times O \subseteq W$, $a \in U'$, and we can replace W by $U' \times O$ and restrict further g to this open set so that we obtain a diffeomorphism from $U' \times O$ to (a smaller) V. If $\theta\colon V \to U' \times O$ is the inverse of this diffeomorphism, then $f(U') \subseteq V$ and since $g(x, 0) = f(x)$,

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

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

Lemma 1.25 *Let $W \subseteq \mathbb{R}^m$ be an open subset of $\mathbb{R}^m$ and pick some $a \in W$. If $f\colon W \to \mathbb{R}^n$ 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 $(x_1, \ldots, x_m) \in O$.

Proof. Since f is a submersion, its Jacobian matrix, $J(f)$, (an $n \times m$ matrix) has rank n and by permuting coordinates if needed, we may assume that the first n columns of $J(f)$ are linearly independent and we let

$$A = \left(\frac{\partial f_i}{\partial x_j}(a)\right)$$

be this invertible $n \times n$ matrix. Define the map, $g\colon W \to \mathbb{R}^m$, by

$$g(x) = (f(x), x_{n+1}, \ldots, x_m),$$

for all $x \in W$. The Jacobian matrix of g at a is of the form

$$J = \begin{pmatrix} A & B \\ 0 & I \end{pmatrix}$$

so $\det(J) = \det(A)\det(I) = \det(A) \neq 0$, since A is invertible. By the inverse function theorem, there are some open subsets $V \subseteq W$ with $a \in V$ and $O \subseteq \mathbb{R}^m$ such that the restriction of g to V is a diffeomorphism between V and O. Let $\psi\colon O \to V$ be the inverse of this diffeomorphism. Because $g \circ \psi = \mathrm{id}$, we have

$$(x_1, \ldots, x_m) = g(\psi(x)) = (f(\psi(x)), \psi_{n+1}(x), \ldots, \psi_m(x)),$$

that is,

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

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

Using Lemmas 1.24 and 1.25, we can prove the following theorem which confirms that all our characterizations of a manifold are equivalent.

Theorem 1.26 *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

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

Proof. If (1) holds, then by Lemma 1.24, replacing Ω by a smaller open subset $\Omega' \subseteq \Omega$ if necessary, there is some open subset $V \subseteq \mathbb{R}^N$ with $p \in V$ and $\varphi(\Omega') \subseteq V$, an open subset, $O \subseteq \mathbb{R}^{N-m}$, and some diffeomorphism, $\theta\colon V \to \Omega' \times O$, so that

$$(\theta \circ \varphi)(x_1, \ldots, x_m) = (x_1, \ldots, x_m, 0, \ldots, 0),$$

for all $(x_1, \ldots, x_m) \in \Omega'$. Observe that the above condition implies that

$$(\theta \circ \varphi)(\Omega') = \theta(V) \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\}).$$

Since φ is a homeomorphism between Ω and its image in M and since $\Omega' \subseteq \Omega$ is an open subset, $\varphi(\Omega') = M \cap W'$ for some open subset $W' \subseteq \mathbb{R}^N$, so if we let $W = V \cap W'$, because $\varphi(\Omega') \subseteq V$ it follows that $\varphi(\Omega') = M \cap W$ and

$$\theta(W \cap M) = \theta(\varphi(\Omega')) = \theta(V) \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\}).$$

However, θ is injective and $\theta(W \cap M) \subseteq \theta(W)$ so

$$\begin{aligned} \theta(W \cap M) &= \theta(W) \cap \theta(V) \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\}) \\ &= \theta(W \cap V) \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\}) \\ &= \theta(W) \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\}). \end{aligned}$$

If we let $O = \theta(W)$, we get

$$\theta^{-1}(O \cap (\mathbb{R}^m \times \{(0, \ldots, 0)\})) = M \cap W,$$

which is (2).

If (2) holds, we can write $\varphi^{-1} = (f_1, \ldots, f_N)$ and because $\varphi^{-1} \colon W \to O$ is a diffeomorphism, $df_1(q), \ldots, df_N(q)$ are linearly independent for all $q \in W$, so the map

$$f = (f_{m+1}, \ldots, f_N)$$

is a submersion, $f \colon W \to \mathbb{R}^{N-m}$, and we have $f(x) = 0$ iff $f_{m+1}(x) = \cdots = f_N(x) = 0$ iff

$$\varphi^{-1}(x) = (f_1(x), \ldots, f_m(x), 0, \ldots, 0)$$

iff $\varphi^{-1}(x) \in O \cap (\mathbb{R}^m \times \{0_{N-m}\})$ iff $x \in \varphi(O \cap (\mathbb{R}^m \times \{0_{N-m}\}) = M \cap W$, because

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

Thus, $M \cap W = f^{-1}(0)$, which is (3).

The proof that (3) implies (2) uses Lemma 1.25 instead of Lemma 1.24. If $f \colon W \to \mathbb{R}^{N-m}$ is the submersion such that $M \cap W = f^{-1}(0)$ given by (3), then by Lemma 1.25, there are open subsets $V \subseteq W$, $O \subseteq \mathbb{R}^N$ and a diffeomorphism, $\psi \colon O \to V$ so that

$$f(\psi(x_1, \ldots, x_N)) = (x_1, \ldots, x_{N-m})$$

for all $(x_1, \ldots, x_N) \in O$. If σ is the permutation of variables given by

$$\sigma(x_1, \ldots, x_m, x_{m+1}, \ldots, x_N) = (x_{m+1}, \ldots, x_N, x_1, \ldots, x_m),$$

then $\varphi = \psi \circ \sigma$ is a diffeomorphism such that

$$f(\varphi(x_1, \ldots, x_N)) = (x_{m+1}, \ldots, x_N)$$

for all $(x_1, \ldots, x_N) \in O$. If we denote the restriction of f to V by g, it is clear that

$$M \cap V = g^{-1}(0)$$

and because $g(\varphi(x_1, \ldots, x_N)) = 0$ iff $(x_{m+1}, \ldots, x_N) = 0_{N-m}$ and φ is a bijection,

$$\begin{aligned} M \cap V &= \{(y_1, \ldots, y_N) \in V \mid g(y_1, \ldots, y_N) = 0\} \\ &= \{\varphi(x_1, \ldots, x_N) \mid (\exists (x_1, \ldots, x_N) \in O)(g(\varphi(x_1, \ldots, x_N)) = 0)\} \\ &= \varphi(O \cap (\mathbb{R}^m \times \{0_{N-m}\})), \end{aligned}$$

which is (2).

If (2) holds, then $\varphi\colon O \to W$ is a diffeomorphism,

$$O \cap (\mathbb{R}^m \times \{0_{N-m}\}) = \Omega \times \{0_{N-m}\}$$

for some open subset, $\Omega \subseteq \mathbb{R}^m$, and the map $\psi\colon \Omega \to \mathbb{R}^N$ given by

$$\psi(x) = \varphi(x, 0_{N-m})$$

is an immersion on Ω and a homeomorhism onto $U \cap M$, which implies (1).

If (3) holds, then if we write $f = (f_1, \ldots, f_{N-m})$, with $f_i\colon W \to \mathbb{R}$, then the fact that $df(p)$ is a submersion is equivalent to the fact that the linear forms $df_1(p), \ldots, df_{N-m}(p)$ are linearly independent and

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

Finally, if (4) holds, then if we define $f\colon W \to \mathbb{R}^{N-m}$ by

$$f = (f_1, \ldots, f_{N-m}),$$

because $df_1(p), \ldots, df_{N-m}(p)$ are linearly independent we get a smooth map which is a submersion at p such that

$$M \cap W = f^{-1}(0).$$

Now, f is a submersion at p iff $df(p)$ is surjective, which means that a certain determinant is nonzero and since the determinant function is continuous, this determinant is nonzero on some open subset, $W' \subseteq W$, containing p, so if we restrict f to W', we get an immersion on W' such that $M \cap W' = f^{-1}(0)$. $\square$

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 1.26, we can show again that 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 (for $x \neq 0$) since

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

We can also show that 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 $A \in \mathbf{GL}(n)$, where $f(A) = \det(A)$.

The third characterization of Theorem 1.26 suggests the following definition.

Definition 1.12 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 1.26 implies the following useful proposition:

Proposition 1.27 *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 1.12 and Proposition 1.27 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

$$f_a(x) = f(x) + a_1 x_1 + \cdots + a_{m+1} x_{m+1}$$

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 this book. A fairly elementary exposition of nondegenerate critical points and Morse functions can be found in Guillemin and Pollack [69] (Chapter 1, Section 7). Sard's theorem is proved in Appendix 1 of Guillemin and Pollack [69] and also in Chapter 2 of Milnor [108]. Morse theory (starting with Morse lemma) and much more, is discussed in Milnor [106], widely recognized as a mathematical masterpiece. An excellent

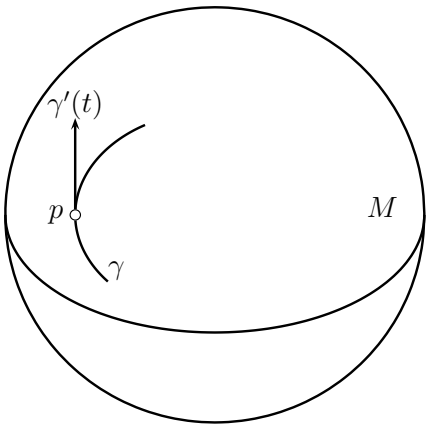

Figure 1.3: Tangent vector to a curve on a manifold

and more leisurely introduction to Morse theory is given in Matsumoto [105], where a proof of Morse lemma is also given.

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

Definition 1.13 Let M be an m-dimensional manifold in $\mathbb{R}^N$. A *smooth curve γ in M* is any function $\gamma\colon I \to M$ 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 $\varphi\colon \Omega \to U$ of M at p and some open interval $]t-\epsilon,\, t+\epsilon[\subseteq I$ such that the curve $\varphi^{-1} \circ \gamma\colon]t-\epsilon,\, t+\epsilon[\to \mathbb{R}^m$ is smooth.

Using Lemma 1.22, it is easily shown that Definition 1.13 does not depend on the choice of the parametrization $\varphi\colon \Omega \to U$ at p.

Lemma 1.22 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 1.3.

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 1.14 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 $\gamma \colon I \to M$ 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 (postponing defining smooth maps).

Definition 1.15 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 $^{-1} \colon G \to G$ are smooth.

(Smooth maps are defined in Definition 1.18). 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 1.16 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 1.28 *A closed subgroup G of $\mathbf{GL}(n, \mathbb{R})$ is a linear Lie group. Furthermore, the set $\mathfrak{g}$ defined such that*

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

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 1.28 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 nonnull 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 nonnull rational numbers is a multiplicative subgroup of $\mathbb{R}^*$, but it is not closed, since $\mathbb{Q}$ is dense in $\mathbb{R}$.

The proof of Theorem 1.28 involves proving that when G is not a discrete subgroup, there is an open subset $\Omega \subseteq \mathbf{M}(n, \mathbb{R})$ such that $0_{n,n} \in \Omega$, an open subset $W \subseteq \mathbf{M}(n, \mathbb{R})$ such that $I \in W$, and that $\exp\colon \Omega \to W$ is a diffeomorphism such that

$$\exp(\Omega \cap \mathfrak{g}) = W \cap G.$$

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

With the help of Theorem 1.28 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 and to figure out what are their Lie algebras. (Of course, $\mathbf{GL}(n, \mathbb{R})$ is a Lie group, as we already know.)

For example, if $G = \mathbf{GL}(n, \mathbb{R})$, as e^{tA} is invertible for *every* matrix, $A \in \mathbf{M}(n, \mathbb{R})$, we deduce that the Lie algebra, $\mathfrak{gl}(n, \mathbb{R})$, of $\mathbf{GL}(n, \mathbb{R})$ is equal to $\mathbf{M}(n, \mathbb{R})$. We also claim that the Lie algebra, $\mathfrak{sl}(n, \mathbb{R})$, of $\mathbf{SL}(n, \mathbb{R})$ is the set of all matrices with zero trace. Indeed, $\mathfrak{sl}(n, \mathbb{R})$ is the subalgebra of $\mathfrak{gl}(n, \mathbb{R})$ consisting of all matrices $X \in \mathfrak{gl}(n, \mathbb{R})$ such that

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

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

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 $\mathbb{R}^n$; see Gallier [58], Chapter 2, Problem 2.19).

Returning to Theorem 1.28, the vector space $\mathfrak{g}$ is called the *Lie algebra* of the Lie group G. Lie algebras are defined as follows.

Definition 1.17 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 1.28, 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 1.28 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. 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 1.7):

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

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 1.4. 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. Every such matrix is of the form

$$i(d\sigma_1 + c\sigma_2 + b\sigma_3) = \begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix},$$

where $b, c, d \in \mathbb{R}$, and $\sigma_1, \sigma_2, \sigma_3$ are the Pauli spin matrices

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

and thus the matrices $i\sigma_1, i\sigma_2, i\sigma_3$ form a basis of the Lie algebra $\mathfrak{su}(2)$. The Lie bracket $[A, B]$ in $\mathfrak{su}(2)$ is given by the usual commutator, $[A, B] = AB - BA$.

It is easily checked that the vector space $\mathbb{R}^3$ is a Lie algebra if we define the Lie bracket on $\mathbb{R}^3$ as the usual cross product $u \times v$ of vectors. Then we can define an isomorphism of Lie algebras $\varphi\colon (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$ by the formula

$$\varphi(b, c, d) = \frac{i}{2}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{2}\begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix}.$$

It is indeed easy to verify that

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

Returning to $\mathfrak{su}(2)$, letting $\theta = \sqrt{b^2 + c^2 + d^2}$, we can write

$$d\sigma_1 + c\sigma_2 + b\sigma_3 = \begin{pmatrix} b & -ic + d \\ ic + d & -b \end{pmatrix} = \theta A,$$

where

$$A = \frac{1}{\theta}(d\sigma_1 + c\sigma_2 + b\sigma_3) = \frac{1}{\theta}\begin{pmatrix} b & -ic + d \\ ic + d & -b \end{pmatrix},$$

so that $A^2 = I$, and it can be shown that the exponential map $\exp\colon \mathfrak{su}(2) \to \mathbf{SU}(2)$ is given by

$$\exp(i\theta A) = \cos\theta\, \mathbf{1} + i \sin\theta\, A.$$

In view of the isomorphism $\varphi\colon (\mathbb{R}^3, \times) \to \mathfrak{su}(2)$, where

$$\varphi(b, c, d) = \frac{1}{2}\begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix} = i\frac{\theta}{2}A,$$

the exponential map can be viewed as a map $\exp\colon (\mathbb{R}^3, \times) \to \mathbf{SU}(2)$ given by the formula

$$\exp(\theta v) = \left[\cos\frac{\theta}{2}, \sin\frac{\theta}{2}\, v\right],$$

for every vector θv, where v is a unit vector in $\mathbb{R}^3$ and $\theta \in \mathbb{R}$. In this form, $\exp(\theta v)$ is a quaternion corresponding to a rotation of axis v and angle θ.

As we showed, $\mathbf{SE}(n)$ is a Lie group, and its lie algebra $\mathfrak{se}(n)$ described in Section 1.6 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 1.18 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 1.22, it is easily shown that Definition 1.18 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 1.19 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 1.22, 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 1.20 Given two Lie groups G_1 and G_2, a *homomorphism (or map) of Lie groups* is a function $f\colon G_1 \to G_2$ 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 $f\colon \mathcal{A}_1 \to \mathcal{A}_2$ 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

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

of $R_{a^{-1}} L_a\colon G \to G$ at I is an isomorphism of Lie algebras, and since $T_I G = \mathfrak{g}$, 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 1.28 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 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. We will do this in Chapter 5 but for this we will need a deeper study of manifolds (see Chapter 3).

Readers who wish to learn more about Lie groups and Lie algebras should consult (more or less listed in order of difficulty) Curtis [38], Sattinger and Weaver [134], Hall [70] and Marsden and Ratiu [102]. The excellent lecture notes by Carter, Segal, and Macdonald [31] constitute a very efficient (although somewhat terse) introduction to Lie algebras and Lie groups. Classics such as Weyl [151] and Chevalley [34] are definitely worth consulting, although the presentation and the terminology may seem a bit old fashioned. For more advanced texts, one may consult Abraham and Marsden [1], Warner [147], Sternberg [143], Bröcker and tom Dieck [25], and Knapp [89]. For those who read French, Mneimné and Testard [111] is very clear and quite thorough, and uses very little differential geometry, although it is more advanced than Curtis. Chapter 1, by Bryant, in Freed and Uhlenbeck [26] is also worth reading, but the pace is fast.

Chapter 2

Review of Groups and Group Actions

2.1 Groups

Definition 2.1 A *group* is a set, G, equipped with an operation, $\cdot : G \times G \to G$, having the following properties: $\cdot$ is *associative*, has an *identity element*, $e \in G$, and every element in G is *invertible* (w.r.t. $\cdot$). More explicitly, this means that the following equations hold for all $a, b, c \in G$:

(G1) $a \cdot (b \cdot c) = (a \cdot b) \cdot c$. (associativity);

(G2) $a \cdot e = e \cdot a = a$. (identity);

(G3) For every $a \in G$, there is some $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e$ (inverse).

A group G is *abelian* (or *commutative*) if

$$a \cdot b = b \cdot a$$

for all $a, b \in G$.

A set M together with an operation $\cdot : M \times M \to M$ and an element e satisfying only conditions (G1) and (G2) is called a *monoid*. For example, the set $\mathbb{N} = \{0, 1, \ldots, n \ldots\}$ of natural numbers is a (commutative) monoid. However, it is not a group.

Observe that a group (or a monoid) is never empty, since $e \in G$.

Some examples of groups are given below:

Example 2.1

1. The set $\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n \ldots\}$ of integers is a group under addition, with identity element 0. However, $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ is not a group under multiplication.

2. The set $\mathbb{Q}$ of rational numbers is a group under addition, with identity element 0. The set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ is also a group under multiplication, with identity element 1.

3. Similarly, the sets $\mathbb{R}$ of real numbers and $\mathbb{C}$ of complex numbers are groups under addition (with identity element 0), and $\mathbb{R}^* = \mathbb{R} - \{0\}$ and $\mathbb{C}^* = \mathbb{C} - \{0\}$ are groups under multiplication (with identity element 1).

4. The sets $\mathbb{R}^n$ and $\mathbb{C}^n$ of n-tuples of real or complex numbers are groups under componentwise addition:

$$(x_1, \ldots, x_n) + (y_1, \cdots, y_n) = (x_1 + y_n, \ldots, x_n + y_n),$$

with identity element $(0, \ldots, 0)$. All these groups are abelian.

5. Given any nonempty set S, the set of bijections $f\colon S \to S$, also called *permutations of* S, is a group under function composition (i.e., the multiplication of f and g is the composition $g \circ f$), with identity element the identity function id_S. This group is not abelian as soon as S has more than two elements.

6. The set of $n \times n$ matrices with real (or complex) coefficients is a group under addition of matrices, with identity element the null matrix. It is denoted by $\mathrm{M}_n(\mathbb{R})$ (or $\mathrm{M}_n(\mathbb{C})$).

7. The set $\mathbb{R}[X]$ of polynomials in one variable with real coefficients is a group under addition of polynomials.

8. The set of $n \times n$ invertible matrices with real (or complex) coefficients is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *general linear group* and is usually denoted by $\mathbf{GL}(n, \mathbb{R})$ (or $\mathbf{GL}(n, \mathbb{C})$).

9. The set of $n \times n$ invertible matrices with real (or complex) coefficients and determinant $+1$ is a group under matrix multiplication, with identity element the identity matrix I_n. This group is called the *special linear group* and is usually denoted by $\mathbf{SL}(n, \mathbb{R})$ (or $\mathbf{SL}(n, \mathbb{C})$).

10. The set of $n \times n$ invertible matrices with real coefficients such that $RR^\top = I_n$ and of determinant $+1$ is a group called the *orthogonal group* and is usually denoted by $\mathbf{SO}(n)$ (where $R^\top$ is the *transpose* of the matrix R, i.e., the rows of $R^\top$ are the columns of R). It corresponds to the rotations in $\mathbb{R}^n$.

11. Given an open interval $]a, b[$, the set $C(]a, b[)$ of continuous functions $f\colon]a, b[\to \mathbb{R}$ is a group under the operation $f + g$ defined such that

$$(f + g)(x) = f(x) + g(x)$$

for all $x \in]a, b[$.

Given a group, G, for any two subsets $R, S \subseteq G$, we let

$$RS = \{r \cdot s \mid r \in R,\ s \in S\}.$$

In particular, for any $g \in G$, if $R = \{g\}$, we write

$$gS = \{g \cdot s \mid s \in S\}$$

and similarly, if $S = \{g\}$, we write

$$Rg = \{r \cdot g \mid r \in R\}.$$

From now on, we will drop the multiplication sign and write g_1g_2 for $g_1 \cdot g_2$.

Definition 2.2 Given a group, G, a subset, H, of G is a *subgroup of* G iff

(1) The identity element, e, of G also belongs to H ($e \in H$);

(2) For all $h_1, h_2 \in H$, we have $h_1h_2 \in H$;

(3) For all $h \in H$, we have $h^{-1} \in H$.

It is easily checked that a subset, $H \subseteq G$, is a subgroup of G iff H is nonempty and whenever $h_1, h_2 \in H$, then $h_1h_2^{-1} \in H$.

If H is a subgroup of G and $g \in G$ is any element, the sets of the form gH are called *left cosets of H in G* and the sets of the form Hg are called *right cosets of H in G*. The left cosets (resp. right cosets) of H induce an equivalence relation, $\sim$, defined as follows: For all $g_1, g_2 \in G$,

$$g_1 \sim g_2 \quad \text{iff} \quad g_1H = g_2H$$

(resp. $g_1 \sim g_2$ iff $Hg_1 = Hg_2$).

Obviously, $\sim$ is an equivalence relation. Now, it is easy to see that $g_1H = g_2H$ iff $g_2^{-1}g_1 \in H$, so the equivalence class of an element $g \in G$ is the coset gH (resp. Hg). The set of left cosets of H in G (which, in general, is **not** a group) is denoted G/H. The "points" of G/H are obtained by "collapsing" all the elements in a coset into a single element.

It is tempting to define a multiplication operation on left cosets (or right cosets) by setting

$$(g_1H)(g_2H) = (g_1g_2)H,$$

but this operation is not well defined in general, unless the subgroup H possesses a special property. This property is typical of the kernels of group homomorphisms, so we are led to

Definition 2.3 Given any two groups, G, G', a function $\varphi\colon G \to G'$ is a *homomorphism* iff

$$\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2), \quad \text{for all } g_1, g_2 \in G.$$

Taking $g_1 = g_2 = e$ (in G), we see that

$$\varphi(e) = e',$$

and taking $g_1 = g$ and $g_2 = g^{-1}$, we see that

$$\varphi(g^{-1}) = \varphi(g)^{-1}.$$

If $\varphi\colon G \to G'$ and $\psi\colon G' \to G''$ are group homomorphisms, then $\psi \circ \varphi\colon G \to G''$ is also a homomorphism. If $\varphi\colon G \to G'$ is a homomorphism of groups and $H \subseteq G$ and $H' \subseteq G'$ are two subgroups, then it is easily checked that

$$\text{Im } H = \varphi(H) = \{\varphi(g) \mid g \in H\} \quad \text{is a subgroup of } G'$$

(Im H is called the *image of H by φ*) and

$$\varphi^{-1}(H') = \{g \in G \mid \varphi(g) \in H'\} \quad \text{is a subgroup of } G.$$

In particular, when $H' = \{e'\}$, we obtain the *kernel*, Ker φ, of φ. Thus,

$$\text{Ker } \varphi = \{g \in G \mid \varphi(g) = e'\}.$$

It is immediately verified that $\varphi\colon G \to G'$ is injective iff Ker $\varphi = \{e\}$. (We also write Ker $\varphi = (0)$.) We say that φ is an *isomorphism* if there is a homomorphism, $\psi\colon G' \to G$, so that

$$\psi \circ \varphi = \text{id}_G \quad \text{and} \quad \varphi \circ \psi = \text{id}_{G'}.$$

In this case, ψ is unique and it is denoted φ^{-1}. When φ is an isomorphism we say the the groups G and G' are *isomorphic*. When $G' = G$, a group isomorphism is called an *automorphism*.

We claim that $H = \text{Ker } \varphi$ satisfies the following property:

$$gH = Hg, \quad \text{for all } g \in G. \tag{$*$}$$

First, note that $(*)$ is equivalent to

$$gHg^{-1} = H, \quad \text{for all } g \in G,$$

and the above is equivalent to

$$gHg^{-1} \subseteq H, \quad \text{for all } g \in G. \tag{$**$}$$

This is because $gHg^{-1} \subseteq H$ implies $H \subseteq g^{-1}Hg$, and this for all $g \in G$. But,

$$\varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g^{-1}) = \varphi(g)e'\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = e',$$

for all $h \in H = \text{Ker } \varphi$ and all $g \in G$. Thus, by definition of $H = \text{Ker } \varphi$, we have $gHg^{-1} \subseteq H$.

Definition 2.4 For any group, G, a subgroup, $N \subseteq G$, is a *normal subgroup* of G iff

$$gNg^{-1} = N, \quad \text{for all } g \in G.$$

This is denoted by $N \lhd G$.

If N is a normal subgroup of G, the equivalence relation induced by left cosets is the same as the equivalence induced by right cosets. Furthermore, this equivalence relation, $\sim$, is a *congruence*, which means that: For all $g_1, g_2, g_1', g_2' \in G$,

(1) If $g_1N = g_1'N$ and $g_2N = g_2'N$, then $g_1g_2N = g_1'g_2'N$, and

(2) If $g_1N = g_2N$, then $g_1^{-1}N = g_2^{-1}N$.

As a consequence, we can define a group structure on the set $G/\sim$ of equivalence classes modulo $\sim$, by setting

$$(g_1N)(g_2N) = (g_1g_2)N.$$

This group is denoted G/N. The equivalence class, gN, of an element $g \in G$ is also denoted $\overline{g}$. The map $\pi\colon G \to G/N$, given by

$$\pi(g) = \overline{g} = gN,$$

is clearly a group homomorphism called the *canonical projection*.

Given a homomorphism of groups, $\varphi\colon G \to G'$, we easily check that the groups $G/\mathrm{Ker}\ \varphi$ and $\mathrm{Im}\ \varphi = \varphi(G)$ are isomorphic.

2.2 Group Actions and Homogeneous Spaces, I

If X is a set (usually, some kind of geometric space, for example, the sphere in $\mathbb{R}^3$, the upper half-plane, etc.), the "symmetries" of X are often captured by the action of a group, G, on X. In fact, if G is a Lie group and the action satisfies some simple properties, the set X can be given a manifold structure which makes it a projection (quotient) of G, a so-called "homogeneous space".

Definition 2.5 Given a set, X, and a group, G, a *left action of G on X* (for short, an *action of G on X*) is a function, $\varphi\colon G \times X \to X$, such that

(1) For all $g, h \in G$ and all $x \in X$,

$$\varphi(g, \varphi(h, x)) = \varphi(gh, x),$$

(2) For all $x \in X$,

$$\varphi(1, x) = x,$$

where $1 \in G$ is the identity element of G.

To alleviate the notation, we usually write $g \cdot x$ or even gx for $\varphi(g, x)$, in which case, the above axioms read:

(1) For all $g, h \in G$ and all $x \in X$,

$$g \cdot (h \cdot x) = gh \cdot x,$$

(2) For all $x \in X$,

$$1 \cdot x = x.$$

The set X is called a *(left) G-set*. The action φ is *faithful* or *effective* iff for every g, if $g \cdot x = x$ for all $x \in X$, then $g = 1$; the action φ is *transitive* iff for any two elements $x, y \in X$, there is some $g \in G$ so that $g \cdot x = y$.

Given an action, $\varphi\colon G \times X \to X$, for every $g \in G$, we have a function, $\varphi_g\colon X \to X$, defined by

$$\varphi_g(x) = g \cdot x, \quad \text{for all } x \in X.$$

Observe that φ_g has $\varphi_{g^{-1}}$ as inverse, since

$$\varphi_{g^{-1}}(\varphi_g(x)) = \varphi_{g^{-1}}(g \cdot x) = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = 1 \cdot x = x,$$

and similarly, $\varphi_g \circ \varphi_{g^{-1}} = \mathrm{id}$. Therefore, φ_g is a bijection of X, i.e., a permutation of X. Moreover, we check immediately that

$$\varphi_g \circ \varphi_h = \varphi_{gh},$$

so, the map $g \mapsto \varphi_g$ is a group homomorphism from G to $\mathfrak{S}_X$, the group of permutations of X. With a slight abuse of notation, this group homomorphism $G \longrightarrow \mathfrak{S}_X$ is also denoted φ.

Conversely, it is easy to see that any group homomorphism, $\varphi\colon G \to \mathfrak{S}_X$, yields a group action, $\cdot\colon G \times X \longrightarrow X$, by setting

$$g \cdot x = \varphi(g)(x).$$

Observe that an action, φ, is faithful iff the group homomorphism, $\varphi\colon G \to \mathfrak{S}_X$, is injective. Also, we have $g \cdot x = y$ iff $g^{-1} \cdot y = x$, since $(gh) \cdot x = g \cdot (h \cdot x)$ and $1 \cdot x = x$, for all $g, h \in G$ and all $x \in X$.

Definition 2.6 Given two G-sets, X and Y, a function, $f\colon X \to Y$, is said to be *equivariant*, or a *G-map* iff for all $x \in X$ and all $g \in G$, we have

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

Remark: We can also define a *right action*, $\cdot\colon X \times G \to X$, of a group G on a set X, as a map satisfying the conditions

(1) For all $g, h \in G$ and all $x \in X$,

$$(x \cdot g) \cdot h = x \cdot gh,$$

(2) For all $x \in X$,

$$x \cdot 1 = x.$$

Every notion defined for left actions is also defined for right actions, in the obvious way.

Here are some examples of (left) group actions.

Example 1: The unit sphere S^2 (more generally, S^{n-1}).

Recall that for any $n \geq 1$, the *(real) unit sphere*, S^{n-1}, is the set of points in $\mathbb{R}^n$ given by

$$S^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 = 1\}.$$

In particular, S^2 is the usual sphere in $\mathbb{R}^3$. Since the group $\mathbf{SO}(3) = \mathbf{SO}(3, \mathbb{R})$ consists of (orientation preserving) linear isometries, i.e., *linear* maps that are distance preserving (and of determinant $+1$), and every linear map leaves the origin fixed, we see that any rotation maps S^2 into itself.

Beware that this would be false if we considered the group of *affine* isometries, $\mathbf{SE}(3)$, of $\mathbb{E}^3$. For example, a screw motion does *not* map S^2 into itself, even though it is distance preserving, because the origin is translated.

Thus, we have an action, $\cdot \colon \mathbf{SO}(3) \times S^2 \to S^2$, given by

$$R \cdot x = Rx.$$

The verification that the above is indeed an action is trivial. This action is transitive. This is because, for any two points x, y on the sphere S^2, there is a rotation whose axis is perpendicular to the plane containing x, y and the center, O, of the sphere (this plane is not unique when x and y are antipodal, i.e., on a diameter) mapping x to y.

Similarly, for any $n \geq 1$, we get an action, $\cdot \colon \mathbf{SO}(n) \times S^{n-1} \to S^{n-1}$. It is easy to show that this action is transitive.

Analogously, we can define the *(complex) unit sphere*, Σ^{n-1}, as the set of points in $\mathbb{C}^n$ given by

$$\Sigma^{n-1} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1\overline{z}_1 + \cdots + z_n\overline{z}_n = 1\}.$$

If we write $z_j = x_j + iy_j$, with $x_j, y_j \in \mathbb{R}$, then

$$\Sigma^{n-1} = \{(x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^{2n} \mid x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 = 1\}.$$

Therefore, we can view the complex sphere, Σ^{n-1} (in $\mathbb{C}^n$), as the real sphere, S^{2n-1} (in $\mathbb{R}^{2n}$). By analogy with the real case, we can define an action, $\cdot \colon \mathbf{SU}(n) \times \Sigma^{n-1} \to \Sigma^{n-1}$, of the group, $\mathbf{SU}(n)$, of *linear* maps of $\mathbb{C}^n$ preserving the hermitian inner product (and the origin, as all linear maps do) and this action is transitive.

One should not confuse the unit sphere, Σ^{n-1}, with the hypersurface, $S^{n-1}_{\mathbb{C}}$, given by

$$S^{n-1}_{\mathbb{C}} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1^2 + \cdots + z_n^2 = 1\}.$$

For instance, one should check that a line, L, through the origin intersects Σ^{n-1} in a circle, whereas it intersects $S^{n-1}_{\mathbb{C}}$ in exactly two points!

Example 2: The upper half-plane.

The *upper half-plane*, H, is the open subset of $\mathbb{R}^2$ consisting of all points, $(x, y) \in \mathbb{R}^2$, with $y > 0$. It is convenient to identify H with the set of complex numbers, $z \in \mathbb{C}$, such that $\Im z > 0$. Then, we can define an action, $\cdot\colon \mathbf{SL}(2, \mathbb{R}) \times H \to H$, of the group $\mathbf{SL}(2, \mathbb{R})$ on H, as follows: For any $z \in H$, for any $A \in \mathbf{SL}(2, \mathbb{R})$,

$$A \cdot z = \frac{az + b}{cz + d},$$

where

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

with $ad - bc = 1$. It is easily verified that $A \cdot z$ is indeed always well defined and in H when $z \in H$. This action is transitive (check this).

Maps of the form

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

where $z \in \mathbb{C}$ and $ad - bc = 1$, are called *Möbius transformations*. Here, $a, b, c, d \in \mathbb{R}$, but in general, we allow $a, b, c, d \in \mathbb{C}$. Actually, these transformations are not necessarily defined everywhere on $\mathbb{C}$, for example, for $z = -d/c$ if $c \neq 0$. To fix this problem, we add a "point at infinity", ∞, to $\mathbb{C}$ and define Möbius transformations as functions $\mathbb{C} \cup \{\infty\} \longrightarrow \mathbb{C} \cup \{\infty\}$. If $c = 0$, the Möbius transformation sends ∞ to itself, otherwise, $-d/c \mapsto \infty$ and $\infty \mapsto a/c$. The space $\mathbb{C} \cup \{\infty\}$ can be viewed as the plane, $\mathbb{R}^2$, extended with a point at infinity. Using a stereographic projection from the sphere S^2 to the plane, (say from the north pole to the equatorial plane), we see that there is a bijection between the sphere, S^2, and $\mathbb{C} \cup \{\infty\}$. More precisely, the *stereographic projection* of the sphere S^2 from the north pole, $N = (0, 0, 1)$, to the plane $z = 0$ (extended with the point at infinity, ∞) is given by

$$(x, y, z) \in S^2 - \{(0, 0, 1)\} \mapsto \left(\frac{x}{1 - z}, \frac{y}{1 - z}\right) = \frac{x + iy}{1 - z} \in \mathbb{C}, \quad \text{with} \quad (0, 0, 1) \mapsto \infty.$$

The inverse stereographic projection is given by

$$(x, y) \mapsto \left(\frac{2x}{x^2 + y^2 + 1}, \frac{2y}{x^2 + y^2 + 1}, \frac{x^2 + y^2 - 1}{x^2 + y^2 + 1}\right), \quad \text{with} \quad \infty \mapsto (0, 0, 1).$$

Intuitively, the inverse stereographic projection "wraps" the equatorial plane around the sphere. The space $\mathbb{C} \cup \{\infty\}$ is known as the *Riemann sphere*. We will see shortly that

$\mathbb{C} \cup \{\infty\} \cong S^2$ is also the complex projective line, $\mathbb{CP}^1$. In summary, Möbius transformations are bijections of the Riemann sphere. It is easy to check that these transformations form a group under composition for all $a, b, c, d \in \mathbb{C}$, with $ad - bc = 1$. This is the *Möbius group*, denoted $\mathbf{Möb}^+$. The Möbius transformations corresponding to the case $a, b, c, d \in \mathbb{R}$, with $ad - bc = 1$ form a subgroup of $\mathbf{Möb}^+$ denoted $\mathbf{Möb}^+_{\mathbb{R}}$. The map from $\mathbf{SL}(2, \mathbb{C})$ to $\mathbf{Möb}^+$ that sends $A \in \mathbf{SL}(2, \mathbb{C})$ to the corresponding Möbius transformation is a surjective group homomorphism and one checks easily that its kernel is $\{-I, I\}$ (where I is the 2×2 identity matrix). Therefore, the Möbius group $\mathbf{Möb}^+$ is isomorphic to the quotient group $\mathbf{SL}(2, \mathbb{C})/\{-I, I\}$, denoted $\mathbf{PSL}(2, \mathbb{C})$. This latter group turns out to be the group of projective transformations of the projective space $\mathbb{CP}^1$. The same reasoning shows that the subgroup $\mathbf{Möb}^+_{\mathbb{R}}$ is isomorphic to $\mathbf{SL}(2, \mathbb{R})/\{-I, I\}$, denoted $\mathbf{PSL}(2, \mathbb{R})$.

The group $\mathbf{SL}(2, \mathbb{C})$ acts on $\mathbb{C} \cup \{\infty\} \cong S^2$ the same way that $\mathbf{SL}(2, \mathbb{R})$ acts on H, namely: For any $A \in \mathbf{SL}(2, \mathbb{C})$, for any $z \in \mathbb{C} \cup \{\infty\}$,

$$A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with} \quad ad - bc = 1.$$

This action is clearly transitive.

One may recall from complex analysis that the (complex) Möbius transformation

$$z \mapsto \frac{z - i}{z + i}$$

is a biholomorphic isomorphism between the upper half plane, H, and the open unit disk,

$$D = \{z \in \mathbb{C} \mid |z| < 1\}.$$

As a consequence, it is possible to define a transitive action of $\mathbf{SL}(2, \mathbb{R})$ on D. This can be done in a more direct fashion, using a group isomorphic to $\mathbf{SL}(2, \mathbb{R})$, namely, $\mathbf{SU}(1, 1)$ (a group of complex matrices), but we don't want to do this right now.

Example 3: The set of $n \times n$ symmetric, positive, definite matrices, $\mathbf{SPD}(n)$.

The group $\mathbf{GL}(n) = \mathbf{GL}(n, \mathbb{R})$ acts on $\mathbf{SPD}(n)$ as follows: For all $A \in \mathbf{GL}(n)$ and all $S \in \mathbf{SPD}(n)$,

$$A \cdot S = ASA^\top.$$

It is easily checked that $ASA^\top$ is in $\mathbf{SPD}(n)$ if S is in $\mathbf{SPD}(n)$. This action is transitive because every SPD matrix, S, can be written as $S = AA^\top$, for some invertible matrix, A (prove this as an exercise).

Example 4: The projective spaces $\mathbb{RP}^n$ and $\mathbb{CP}^n$.

The *(real) projective space*, $\mathbb{RP}^n$, is the set of all lines through the origin in $\mathbb{R}^{n+1}$, i.e., the set of one-dimensional subspaces of $\mathbb{R}^{n+1}$ (where $n \geq 0$). Since a one-dimensional subspace, $L \subseteq \mathbb{R}^{n+1}$, is spanned by any nonzero vector, $u \in L$, we can view $\mathbb{RP}^n$ as the set of equivalence classes of nonzero vectors in $\mathbb{R}^{n+1} - \{0\}$ modulo the equivalence relation,

$$u \sim v \quad \text{iff} \quad v = \lambda u, \quad \text{for some} \quad \lambda \in \mathbb{R},\ \lambda \neq 0.$$

In terms of this definition, there is a projection, $pr\colon (\mathbb{R}^{n+1} - \{0\}) \to \mathbb{RP}^n$, given by $pr(u) = [u]_\sim$, the equivalence class of u modulo $\sim$. Write $[u]$ for the line defined by the nonzero vector, u. Since every line, L, in $\mathbb{R}^{n+1}$ intersects the sphere S^n in two antipodal points, we can view $\mathbb{RP}^n$ as the quotient of the sphere S^n by identification of antipodal points. We write

$$S^n/\{I, -I\} \cong \mathbb{RP}^n.$$

We define an action of $\mathbf{SO}(n+1)$ on $\mathbb{RP}^n$ as follows: For any line, $L = [u]$, for any $R \in \mathbf{SO}(n+1)$,

$$R \cdot L = [Ru].$$

Since R is linear, the line $[Ru]$ is well defined, i.e., does not depend on the choice of $u \in L$. It is clear that this action is transitive.

The *(complex) projective space*, $\mathbb{CP}^n$, is defined analogously as the set of all lines through the origin in $\mathbb{C}^{n+1}$, i.e., the set of one-dimensional subspaces of $\mathbb{C}^{n+1}$ (where $n \geq 0$). This time, we can view $\mathbb{CP}^n$ as the set of equivalence classes of vectors in $\mathbb{C}^{n+1} - \{0\}$ modulo the equivalence relation,

$$u \sim v \quad \text{iff} \quad v = \lambda u, \quad \text{for some} \quad \lambda\ \neq 0 \in \mathbb{C}.$$

We have the projection, $pr\colon \mathbb{C}^{n+1} - \{0\} \to \mathbb{CP}^n$, given by $pr(u) = [u]_\sim$, the equivalence class of u modulo $\sim$. Again, write $[u]$ for the line defined by the nonzero vector, u.

Remark: Algebraic geometers write $\mathbb{P}^n_{\mathbb{R}}$ for $\mathbb{RP}^n$ and $\mathbb{P}^n_{\mathbb{C}}$ (or even $\mathbb{P}^n$) for $\mathbb{CP}^n$.

Recall that $\Sigma^n \subseteq \mathbb{C}^{n+1}$, the unit sphere in $\mathbb{C}^{n+1}$, is defined by

$$\Sigma^n = \{(z_1, \ldots, z_{n+1}) \in \mathbb{C}^{n+1} \mid z_1\overline{z}_1 + \cdots + z_{n+1}\overline{z}_{n+1} = 1\}.$$

For any line, $L = [u]$, where $u \in \mathbb{C}^{n+1}$ is a nonzero vector, writing $u = (u_1, \ldots, u_{n+1})$, a point $z \in \mathbb{C}^{n+1}$ belongs to L iff $z = \lambda(u_1, \ldots, u_{n+1})$, for some $\lambda \in \mathbb{C}$. Therefore, the intersection, $L \cap \Sigma^n$, of the line L and the sphere Σ^n is given by

$$L \cap \Sigma^n = \{\lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \mid \lambda \in \mathbb{C},\ \lambda\overline{\lambda}(u_1\overline{u}_1 + \cdots + u_{n+1}\overline{u}_{n+1}) = 1\},$$

i.e.,

$$L \cap \Sigma^n = \left\{ \lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \;\middle|\; \lambda \in \mathbb{C},\ |\lambda| = \frac{1}{\sqrt{|u_1|^2 + \cdots + |u_{n+1}|^2}} \right\}.$$

Thus, we see that there is a bijection between $L \cap \Sigma^n$ and the circle, S^1, i.e., geometrically, $L \cap \Sigma^n$ is a circle. Moreover, since any line, L, through the origin is determined by just one other point, we see that for any two lines L_1 and L_2 through the origin,

$$L_1 \neq L_2 \quad \text{iff} \quad (L_1 \cap \Sigma^n) \cap (L_2 \cap \Sigma^n) = \emptyset.$$

However, Σ^n is the sphere S^{2n+1} in $\mathbb{R}^{2n+2}$. It follows that $\mathbb{CP}^n$ is the quotient of S^{2n+1} by the equivalence relation, $\sim$, defined such that

$$y \sim z \quad \text{iff} \quad y, z \in L \cap \Sigma^n, \quad \text{for some line, } L, \text{ through the origin.}$$

Therefore, we can write

$$S^{2n+1}/S^1 \cong \mathbb{CP}^n.$$

Observe that $\mathbb{CP}^n$ can also be viewed as the orbit space of the action, $\cdot\colon S^1 \times S^{2n+1} \to S^{2n+1}$, given by

$$\lambda \cdot (z_1, \ldots, z_{n+1}) = (\lambda z_1, \ldots, \lambda z_{n+1}),$$

where $S^1 = \mathbf{U}(1)$ (the group of complex numbers of modulus 1) and S^{2n+1} is identified with Σ^n. The case $n = 1$ is particularly interesting, as it turns out that

$$S^3/S^1 \cong S^2.$$

This is the famous *Hopf fibration*. To show this, proceed as follows: As

$$S^3 \cong \Sigma^1 = \{(z, z') \in \mathbb{C}^2 \mid |z|^2 + |z'|^2 = 1\},$$

define a map, $\mathrm{HF}\colon S^3 \to S^2$, by

$$\mathrm{HF}((z, z')) = (2z\overline{z'}, |z|^2 - |z'|^2).$$

We leave as a homework exercise to prove that this map has range S^2 and that

$$\mathrm{HF}((z_1, z_1')) = \mathrm{HF}((z_2, z_2')) \quad \text{iff} \quad (z_1, z_1') = \lambda(z_2, z_2'), \quad \text{for some } \lambda \text{ with } |\lambda| = 1.$$

In other words, for any point, $p \in S^2$, the inverse image, $\mathrm{HF}^{-1}(p)$ (also called *fibre* over p), is a circle on S^3. Consequently, S^3 can be viewed as the union of a family of disjoint circles. This is the *Hopf fibration*. It is possible to visualize the Hopf fibration using the stereographic projection from S^3 onto $\mathbb{R}^3$. This is a beautiful and puzzling picture. For example, see Berger [15]. Therefore, HF induces a bijection from $\mathbb{CP}^1$ to S^2, and it is a homeomorphism.

We define an action of $\mathbf{SU}(n+1)$ on $\mathbb{CP}^n$ as follows: For any line, $L = [u]$, for any $R \in \mathbf{SU}(n+1)$,

$$R \cdot L = [Ru].$$

Again, this action is well defined and it is transitive.

Example 5: Affine spaces.

If E is any (real) vector space and X is any set, a transitive and faithful action, $\cdot\colon E \times X \to X$, of the additive group of E on X makes X into an *affine space*. The intuition is that the members of E are translations.

Those familiar with affine spaces as in Gallier [58] (Chapter 2) or Berger [15] will point out that if X is an affine space, then, not only is the action of E on X transitive, but more is true: For any two points, $a, b \in E$, there is a *unique* vector, $u \in E$, such that $u \cdot a = b$. By the way, the action of E on X is usually considered to be a right action and is written additively, so $u \cdot a$ is written $a + u$ (the result of translating a by u). Thus, it would seem that we have to require more of our action. However, this is not necessary because E (under addition) is *abelian*. More precisely, we have the proposition

Proposition 2.1 *If G is an abelian group acting on a set X and the action $\cdot\colon G \times X \to X$ is transitive and faithful, then for any two elements $x, y \in X$, there is a unique $g \in G$ so that $g \cdot x = y$ (the action is simply transitive).*

Proof. Since our action is transitive, there is at least some $g \in G$ so that $g \cdot x = y$. Assume that we have $g_1, g_2 \in G$ with

$$g_1 \cdot x = g_2 \cdot x = y.$$

We shall prove that, actually,

$$g_1 \cdot z = g_2 \cdot z, \quad \text{for all } z \in X.$$

As our action is faithful we must have $g_1 = g_2$, and this proves our proposition.

Pick any $z \in X$. As our action is transitive, there is some $h \in G$ so that $z = h \cdot x$. Then, we have

$$\begin{aligned}
g_1 \cdot z &= g_1 \cdot (h \cdot x) \\
&= (g_1 h) \cdot x \\
&= (h g_1) \cdot x \qquad \text{(since } G \text{ is abelian)} \\
&= h \cdot (g_1 \cdot x) \\
&= h \cdot (g_2 \cdot x) \qquad \text{(since } \quad g_1 \cdot x = g_2 \cdot x) \\
&= (h g_2) \cdot x \\
&= (g_2 h) \cdot x \qquad \text{(since } G \text{ is abelian)} \\
&= g_2 \cdot (h \cdot x) \\
&= g_2 \cdot z.
\end{aligned}$$

Therefore, $g_1 \cdot z = g_2 \cdot z$, for all $z \in X$, as claimed. □

More examples will be considered later.

The subset of group elements that leave some given element $x \in X$ fixed plays an important role.

Definition 2.7 Given an action, $\cdot\colon G \times X \to X$, of a group G on a set X, for any $x \in X$, the group G_x (also denoted $\mathrm{Stab}_G(x)$), called the *stabilizer* of x or *isotropy group at* x is given by

$$G_x = \{g \in G \mid g \cdot x = x\}.$$

We have to verify that G_x is indeed a subgroup of G, but this is easy. Indeed, if $g \cdot x = x$ and $h \cdot x = x$, then we also have $h^{-1} \cdot x = x$ and so, we get $gh^{-1} \cdot x = x$, proving that G_x is a subgroup of G. In general, G_x is **not** a normal subgroup.

Observe that

$$G_{g\cdot x} = gG_xg^{-1},$$

for all $g \in G$ and all $x \in X$.

Indeed,

$$\begin{aligned} G_{g\cdot x} &= \{h \in G \mid h \cdot (g \cdot x) = g \cdot x\} \\ &= \{h \in G \mid hg \cdot x = g \cdot x\} \\ &= \{h \in G \mid g^{-1}hg \cdot x = x\} \\ &= gG_xg^{-1}. \end{aligned}$$

Therefore, the stabilizers of x and $g \cdot x$ are conjugate of each other.

When the action of G on X is transitive, for any fixed $x \in G$, the set X is a quotient (as set, not as group) of G by G_x. Indeed, we can define the map, $\pi_x\colon G \to X$, by

$$\pi_x(g) = g \cdot x, \quad \text{for all } g \in G.$$

Observe that

$$\pi_x(gG_x) = (gG_x) \cdot x = g \cdot (G_x \cdot x) = g \cdot x = \pi_x(g).$$

This shows that $\pi_x\colon G \to X$ induces a quotient map, $\overline{\pi}_x\colon G/G_x \to X$, from the set, G/G_x, of (left) cosets of G_x to X, defined by

$$\overline{\pi}_x(gG_x) = g \cdot x.$$

Since

$$\pi_x(g) = \pi_x(h) \quad \text{iff} \quad g \cdot x = h \cdot x \quad \text{iff} \quad g^{-1}h \cdot x = x \quad \text{iff} \quad g^{-1}h \in G_x \quad \text{iff} \quad gG_x = hG_x,$$

we deduce that $\overline{\pi}_x\colon G/G_x \to X$ is injective. However, since our action is transitive, for every $y \in X$, there is some $g \in G$ so that $g \cdot x = y$ and so, $\overline{\pi}_x(gG_x) = g \cdot x = y$, i.e., the map $\overline{\pi}_x$ is also surjective. Therefore, the map $\overline{\pi}_x\colon G/G_x \to X$ is a bijection (of sets, not groups). The map $\pi_x\colon G \to X$ is also surjective. Let us record this important fact as

Proposition 2.2 *If $\cdot\colon G \times X \to X$ is a transitive action of a group G on a set X, for every fixed $x \in X$, the surjection, $\pi\colon G \to X$, given by*

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

induces a bijection

$$\overline{\pi}\colon G/G_x \to X,$$

where G_x is the stabilizer of x.

The map $\pi\colon G \to X$ (corresponding to a fixed $x \in X$) is sometimes called a *projection* of G onto X. Proposition 2.2 shows that for every $y \in X$, the subset, $\pi^{-1}(y)$, of G (called the *fibre above* y) is equal to some coset, gG_x, of G and thus, is in bijection with the group G_x itself. We can think of G as a moving family of fibres, G_x, parametrized by X. This point of view of viewing a space as a moving family of simpler spaces is typical in (algebraic) geometry, and underlies the notion of (principal) fibre bundle.

Note that if the action $\cdot\colon G \times X \to X$ is transitive, then the stabilizers G_x and G_y of any two elements $x, y \in X$ are isomorphic, as they as conjugates. Thus, in this case, it is enough to compute one of these stabilizers for a "convenient" x.

As the situation of Proposition 2.2 is of particular interest, we make the following definition:

Definition 2.8 A set, X, is said to be a *homogeneous space* if there is a transitive action, $\cdot\colon G \times X \to X$, of some group, G, on X.

We see that all the spaces of Example 1–5 are homogeneous spaces. Another example that will play an important role when we deal with Lie groups is the situation where we have a group, G, a subgroup, H, of G (not necessarily normal) and where $X = G/H$, the set of left cosets of G modulo H. The group G acts on G/H by left multiplication:

$$a \cdot (gH) = (ag)H,$$

where $a, g \in G$. This action is clearly transitive and one checks that the stabilizer of gH is gHg^{-1}. If G is a topological group and H is a closed subgroup of G (see later for an explanation), it turns out that G/H is Hausdorff (Recall that a topological space, X, is *Hausdorff* iff for any two distinct points $x \neq y \in X$, there exists two disjoint open subsets, U and V, with $x \in U$ and $y \in V$.) If G is a Lie group, we obtain a manifold.

Even if G and X are topological spaces and the action, $\cdot\colon G \times X \to X$, is continuous, the space G/G_x under the quotient topology is, in general, **not** homeomorphic to X.

We will give later sufficient conditions that insure that X is indeed a topological space or even a manifold. In particular, X will be a manifold when G is a Lie group.

In general, an action $\cdot\colon G \times X \to X$ is not transitive on X, but for every $x \in X$, it is transitive on the set

$$O(x) = G \cdot x = \{g \cdot x \mid g \in G\}.$$

Such a set is called the *orbit* of x. The orbits are the equivalence classes of the following equivalence relation:

Definition 2.9 Given an action, $\cdot\colon G \times X \to X$, of some group, G, on X, the equivalence relation, $\sim$, on X is defined so that, for all $x, y \in X$,

$$x \sim y \quad \text{iff} \quad y = g \cdot x, \quad \text{for some } g \in G.$$

For every $x \in X$, the equivalence class of x is the *orbit of* x, denoted $O(x)$ or $\mathrm{Orb}_G(x)$, with

$$O(x) = \{g \cdot x \mid g \in G\}.$$

The set of orbits is denoted X/G.

The orbit space, X/G, is obtained from X by an identification (or merging) process: For every orbit, all points in that orbit are merged into a single point. For example, if $X = S^2$ and G is the group consisting of the restrictions of the two linear maps I and $-I$ of $\mathbb{R}^3$ to S^2 (where $-I(x, y, z) = (-x, -y, -z)$), then

$$X/G = S^2/\{I, -I\} \cong \mathbb{RP}^2.$$

Many manifolds can be obtained in this fashion, including the torus, the Klein bottle, the Möbius band, etc.

Since the action of G is transitive on $O(x)$, by Proposition 2.2, we see that for every $x \in X$, we have a bijection

$$O(x) \cong G/G_x.$$

As a corollary, if both X and G are finite, for any set, $A \subseteq X$, of representatives from every orbit, we have the *orbit formula*:

$$|X| = \sum_{a \in A} [G\colon G_x] = \sum_{a \in A} |G|/|G_x|.$$

Even if a group action, $\cdot\colon G \times X \to X$, is not transitive, when X is a manifold, we can consider the set of orbits, X/G, and if the action of G on X satisfies certain conditions, X/G is actually a manifold. Manifolds arising in this fashion are often called *orbifolds*. In summary, we see that manifolds arise in at least two ways from a group action:

(1) As homogeneous spaces, G/G_x, if the action is transitive.

(2) As orbifolds, X/G.

Of course, in both cases, the action must satisfy some additional properties.

Let us now determine some stabilizers for the actions of Examples 1–4, and for more examples of homogeneous spaces.

(a) Consider the action, $\cdot\colon \mathbf{SO}(n) \times S^{n-1} \to S^{n-1}$, of $\mathbf{SO}(n)$ on the sphere S^{n-1} $(n \geq 1)$ defined in Example 1. Since this action is transitive, we can determine the stabilizer of any convenient element of S^{n-1}, say $e_1 = (1, 0, \ldots, 0)$. In order for any $R \in \mathbf{SO}(n)$ to leave e_1 fixed, the first column of R must be e_1, so R is an orthogonal matrix of the form

$$R = \begin{pmatrix} 1 & U \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \det(S) = 1.$$

As the rows of R must be unit vector, we see that $U = 0$ and $S \in \mathbf{SO}(n-1)$. Therefore, the stabilizer of e_1 is isomorphic to $\mathbf{SO}(n-1)$, and we deduce the bijection

$$\mathbf{SO}(n)/\mathbf{SO}(n-1) \cong S^{n-1}.$$

Strictly speaking, $\mathbf{SO}(n-1)$ is not a subgroup of $\mathbf{SO}(n)$ and in all rigor, we should consider the subgroup, $\widetilde{\mathbf{SO}}(n-1)$, of $\mathbf{SO}(n)$ consisting of all matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \det(S) = 1$$

and write

$$\mathbf{SO}(n)/\widetilde{\mathbf{SO}}(n-1) \cong S^{n-1}.$$

However, it is common practice to identify $\mathbf{SO}(n-1)$ with $\widetilde{\mathbf{SO}}(n-1)$.

When $n = 2$, as $\mathbf{SO}(1) = \{1\}$, we find that $\mathbf{SO}(2) \cong S^1$, a circle, a fact that we already knew. When $n = 3$, we find that $\mathbf{SO}(3)/\mathbf{SO}(2) \cong S^2$. This says that $\mathbf{SO}(3)$ is somehow the result of glueing circles to the surface of a sphere (in $\mathbb{R}^3$), in such a way that these circles do not intersect. This is hard to visualize!

A similar argument for the complex unit sphere, Σ^{n-1}, shows that

$$\mathbf{SU}(n)/\mathbf{SU}(n-1) \cong \Sigma^{n-1} \cong S^{2n-1}.$$

Again, we identify $\mathbf{SU}(n-1)$ with a subgroup of $\mathbf{SU}(n)$, as in the real case. In particular, when $n = 2$, as $\mathbf{SU}(1) = \{1\}$, we find that

$$\mathbf{SU}(2) \cong S^3,$$

i.e., the group $\mathbf{SU}(2)$ is topologically the sphere S^3! Actually, this is not surprising if we remember that $\mathbf{SU}(2)$ is in fact the group of unit quaternions.

(b) We saw in Example 2 that the action, $\cdot\colon \mathbf{SL}(2,\mathbb{R}) \times H \to H$, of the group $\mathbf{SL}(2,\mathbb{R})$ on the upper half plane is transitive. Let us find out what the stabilizer of $z = i$ is. We should have

$$\frac{ai+b}{ci+d} = i,$$

that is, $ai + b = -c + di$, i.e.,

$$(d-a)i = b + c.$$

Since a, b, c, d are real, we must have $d = a$ and $b = -c$. Moreover, $ad - bc = 1$, so we get $a^2 + b^2 = 1$. We conclude that a matrix in $\mathbf{SL}(2,\mathbb{R})$ fixes i iff it is of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad \text{with} \quad a^2 + b^2 = 1.$$

Clearly, these are the rotation matrices in $\mathbf{SO}(2)$ and so, the stabilizer of i is $\mathbf{SO}(2)$. We conclude that

$$\mathbf{SL}(2,\mathbb{R})/\mathbf{SO}(2) \cong H.$$

This time, we can view $\mathbf{SL}(2,\mathbb{R})$ as the result of glueing circles to the upper half plane. This is not so easy to visualize. There is a better way to visualize the topology of $\mathbf{SL}(2,\mathbb{R})$ by making it act on the open disk, D. We will return to this action in a little while.

Now, consider the action of $\mathbf{SL}(2,\mathbb{C})$ on $\mathbb{C} \cup \{\infty\} \cong S^2$. As it is transitive, let us find the stabilizer of $z = 0$. We must have

$$\frac{b}{d} = 0,$$

and as $ad - bc = 1$, we must have $b = 0$ and $ad = 1$. Thus, the stabilizer of 0 is the subgroup, $\mathbf{SL}(2,\mathbb{C})_0$, of $\mathbf{SL}(2,\mathbb{C})$ consisting of all matrices of the form

$$\begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix}, \quad \text{where} \quad a \in \mathbb{C} - \{0\} \quad \text{and} \quad c \in \mathbb{C}.$$

We get

$$\mathbf{SL}(2,\mathbb{C})/\mathbf{SL}(2,\mathbb{C})_0 \cong \mathbb{C} \cup \{\infty\} \cong S^2,$$

but this is not very illuminating.

(c) In Example 3, we considered the action, $\cdot\colon \mathbf{GL}(n) \times \mathbf{SPD}(n) \to \mathbf{SPD}(n)$, of $\mathbf{GL}(n)$ on $\mathbf{SPD}(n)$, the set of symmetric positive definite matrices. As this action is transitive, let us find the stabilizer of I. For any $A \in \mathbf{GL}(n)$, the matrix A stabilizes I iff

$$AIA^\top = AA^\top = I.$$

Therefore, the stabilizer of I is $\mathbf{O}(n)$ and we find that

$$\mathbf{GL}(n)/\mathbf{O}(n) = \mathbf{SPD}(n).$$

Observe that if $\mathbf{GL}^+(n)$ denotes the subgroup of $\mathbf{GL}(n)$ consisting of all matrices with a strictly positive determinant, then we have an action $\cdot\colon \mathbf{GL}^+(n) \times \mathbf{SPD}(n) \to \mathbf{SPD}(n)$ of $\mathbf{GL}^+(n)$ on $\mathbf{SPD}(n)$. This action is transtive and we find that the stabilizer of I is $\mathbf{SO}(n)$; consequently, we get

$$\mathbf{GL}^+(n)/\mathbf{SO}(n) = \mathbf{SPD}(n).$$

(d) In Example 4, we considered the action, $\cdot\colon \mathbf{SO}(n+1) \times \mathbb{RP}^n \to \mathbb{RP}^n$, of $\mathbf{SO}(n+1)$ on the (real) projective space, $\mathbb{RP}^n$. As this action is transitive, let us find the stabilizer of the line, $L = [e_1]$, where $e_1 = (1, 0, \ldots, 0)$. For any $R \in \mathbf{SO}(n+1)$, the line L is fixed iff either $R(e_1) = e_1$ or $R(e_1) = -e_1$, since e_1 and $-e_1$ define the same line. As R is orthogonal with $\det(R) = 1$, this means that R is of the form

$$R = \begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \alpha = \pm 1 \quad \text{and} \quad \det(S) = \alpha.$$

But, S must be orthogonal, so we conclude $S \in \mathbf{O}(n)$. Therefore, the stabilizer of $L = [e_1]$ is isomorphic to the group $\mathbf{O}(n)$ and we find that

$$\mathbf{SO}(n+1)/\mathbf{O}(n) \cong \mathbb{RP}^n.$$

Strictly speaking, $\mathbf{O}(n)$ is not a subgroup of $\mathbf{SO}(n+1)$, so the above equation does not make sense. We should write

$$\mathbf{SO}(n+1)/\widetilde{\mathbf{O}}(n) \cong \mathbb{RP}^n,$$

where $\widetilde{\mathbf{O}}(n)$ is the subgroup of $\mathbf{SO}(n+1)$ consisting of all matrices of the form

$$\begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad S \in \mathbf{O}(n),\ \alpha = \pm 1 \quad \text{and} \quad \det(S) = \alpha.$$

However, the common practice is to write $\mathbf{O}(n)$ instead of $\widetilde{\mathbf{O}}(n)$.

We should mention that $\mathbb{RP}^3$ and $\mathbf{SO}(3)$ are homeomorphic spaces. This is shown using the quaternions, for example, see Gallier [58], Chapter 8.

A similar argument applies to the action, $\cdot\colon \mathbf{SU}(n+1) \times \mathbb{CP}^n \to \mathbb{CP}^n$, of $\mathbf{SU}(n+1)$ on the (complex) projective space, $\mathbb{CP}^n$. We find that

$$\mathbf{SU}(n+1)/\mathbf{U}(n) \cong \mathbb{CP}^n.$$

Again, the above is a bit sloppy as $\mathbf{U}(n)$ is not a subgroup of $\mathbf{SU}(n+1)$. To be rigorous, we should use the subgroup, $\widetilde{\mathbf{U}}(n)$, consisting of all matrices of the form

$$\begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad S \in \mathbf{U}(n),\ |\alpha| = 1 \quad \text{and} \quad \det(S) = \overline{\alpha}.$$

The common practice is to write $\mathbf{U}(n)$ instead of $\widetilde{\mathbf{U}}(n)$. In particular, when $n = 1$, we find that

$$\mathbf{SU}(2)/\mathbf{U}(1) \cong \mathbb{CP}^1.$$

But, we know that $\mathbf{SU}(2) \cong S^3$ and, clearly, $\mathbf{U}(1) \cong S^1$. So, again, we find that $S^3/S^1 \cong \mathbb{CP}^1$ (but we know, more, namely, $S^3/S^1 \cong S^2 \cong \mathbb{CP}^1$.)

(e) We now consider a generalization of projective spaces (real and complex). First, consider the real case. Given any $n \geq 1$, for any k, with $0 \leq k \leq n$, let $G(k, n)$ be the set of all linear k-dimensional subspaces of $\mathbb{R}^n$ (also called *k-planes*). Any k-dimensional subspace, U, of $\mathbb{R}$ is spanned by k linearly independent vectors, $u_1, \ldots, u_k$, in $\mathbb{R}^n$; write $U = \text{span}(u_1, \ldots, u_k)$. We can define an action, $\cdot\colon \mathbf{O}(n) \times G(k, n) \to G(k, n)$, as follows: For any $R \in \mathbf{O}(n)$, for any $U = \text{span}(u_1, \ldots, u_k)$, let

$$R \cdot U = \text{span}(Ru_1, \ldots, Ru_k).$$

We have to check that the above is well defined. If $U = \text{span}(v_1, \ldots, v_k)$ for any other k linearly independent vectors, $v_1, \ldots, v_k$, we have

$$v_i = \sum_{j=1}^{k} a_{ij} u_j, \quad 1 \leq i \leq k,$$

for some $a_{ij} \in \mathbb{R}$, and so,

$$Rv_i = \sum_{j=1}^{k} a_{ij} Ru_j, \quad 1 \leq i \leq k,$$

which shows that

$$\text{span}(Ru_1, \ldots, Ru_k) = \text{span}(Rv_1, \ldots, Rv_k),$$

i.e., the above action is well defined. This action is transitive. This is because if U and V are any two k-planes, we may assume that $U = \text{span}(u_1, \ldots, u_k)$ and $V = \text{span}(v_1, \ldots, v_k)$, where the u_i's form an orthonormal family and similarly for the v_i's. Then, we can extend these families to orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ or $\mathbb{R}^n$, and w.r.t. the orthonormal basis $(u_1, \ldots, u_n)$, the matrix of the linear map sending u_i to v_i is orthogonal. Thus, it is enough to find the stabilizer of any k-plane. Pick $U = \text{span}(e_1, \ldots, e_k)$, where $(e_1, \ldots, e_n)$ is the canonical basis of $\mathbb{R}^n$ (i.e., $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, with the 1 in the ith position). Now, any $R \in \mathbf{O}(n)$ stabilizes U iff R maps $e_1, \ldots, e_k$ to k linearly independent vectors in the subspace $U = \text{span}(e_1, \ldots, e_k)$, i.e., R is of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

where S is $k \times k$ and T is $(n-k) \times (n-k)$. Moreover, as R is orthogonal, S and T must be orthogonal, i.e., $S \in \mathbf{O}(k)$ and $T \in \mathbf{O}(n-k)$. We deduce that the stabilizer of U is isomorphic to $\mathbf{O}(k) \times \mathbf{O}(n-k)$ and we find that

$$\mathbf{O}(n)/(\mathbf{O}(k) \times \mathbf{O}(n-k)) \cong G(k, n).$$

It turns out that this makes $G(k,n)$ into a smooth manifold of dimension $k(n-k)$ called a *Grassmannian*.

The restriction of the action of $\mathbf{O}(n)$ on $G(k,n)$ to $\mathbf{SO}(n)$ yields an action, $\cdot\colon \mathbf{SO}(n) \times G(k,n) \to G(k,n)$, of $\mathbf{SO}(n)$ on $G(k,n)$. Then, it is easy to see that the stabilizer of the subspace U is isomorphic to the subgroup, $S(\mathbf{O}(k) \times \mathbf{O}(n-k))$, of $\mathbf{SO}(n)$ consisting of the rotations of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

with $S \in \mathbf{O}(k)$, $T \in \mathbf{O}(n-k)$ and $\det(S)\det(T) = 1$. Thus, we also have

$$\mathbf{SO}(n)/S(\mathbf{O}(k) \times \mathbf{O}(n-k)) \cong G(k,n).$$

If we recall the projection $pr\colon \mathbb{R}^{n+1} - \{0\} \to \mathbb{RP}^n$, by definition, a *$k$-plane* in $\mathbb{RP}^n$ is the image under pr of any $(k+1)$-plane in $\mathbb{R}^{n+1}$. So, for example, a line in $\mathbb{RP}^n$ is the image of a 2-plane in $\mathbb{R}^{n+1}$, and a hyperplane in $\mathbb{RP}^n$ is the image of a hyperplane in $\mathbb{R}^{n+1}$. The advantage of this point of view is that the k-planes in $\mathbb{RP}^n$ are arbitrary, i.e., they do not have to go through "the origin" (which does not make sense, anyway!). Then, we see that we can interpret the Grassmannian, $G(k+1,n+1)$, as a space of "parameters" for the k-planes in $\mathbb{RP}^n$. For example, $G(2,n+1)$ parametrizes the lines in $\mathbb{RP}^n$. In this viewpoint, $G(k+1,n+1)$ is usually denoted $\mathbb{G}(k,n)$.

It can be proved (using some exterior algebra) that $G(k,n)$ can be embedded in $\mathbb{RP}^{\binom{n}{k}-1}$. Much more is true. For example, $G(k,n)$ is a projective variety, which means that it can be defined as a subset of $\mathbb{RP}^{\binom{n}{k}-1}$ equal to the zero locus of a set of homogeneous equations. There is even a set of quadratic equations, known as the *Plücker equations*, defining $G(k,n)$. In particular, when $n = 4$ and $k = 2$, we have $G(2,4) \subseteq \mathbb{RP}^5$ and $G(2,4)$ is defined by a single equation of degree 2. The Grassmannian $G(2,4) = \mathbb{G}(1,3)$ is known as the *Klein quadric*. This hypersurface in $\mathbb{RP}^5$ parametrizes the lines in $\mathbb{RP}^3$.

Complex Grassmannians are defined in a similar way, by replacing $\mathbb{R}$ by $\mathbb{C}$ and $\mathbf{O}(n)$ by $\mathbf{U}(n)$ throughout. The complex Grassmannian, $G_{\mathbb{C}}(k,n)$, is a complex manifold as well as a real manifold and we have

$$\mathbf{U}(n)/(\mathbf{U}(k) \times \mathbf{U}(n-k)) \cong G_{\mathbb{C}}(k,n).$$

As in the case of the real Grassmannians, the action of $\mathbf{U}(n)$ on $G_{\mathbb{C}}(k,n)$ yields an action of $\mathbf{SU}(n)$ on $G_{\mathbb{C}}(k,n)$ and we get

$$\mathbf{SU}(n)/S(\mathbf{U}(k) \times \mathbf{U}(n-k)) \cong G_{\mathbb{C}}(k,n),$$

where $S(\mathbf{U}(k) \times \mathbf{U}(n-k))$ is the subgroup of $\mathbf{SU}(n)$ consisting of all matrices, $R \in \mathbf{SU}(n)$, of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

with $S \in \mathbf{U}(k)$, $T \in \mathbf{U}(n-k)$ and $\det(S)\det(T) = 1$.

We now return to case (b) to give a better picture of $\mathbf{SL}(2,\mathbb{R})$. Instead of having $\mathbf{SL}(2,\mathbb{R})$ act on the upper half plane we define an action of $\mathbf{SL}(2,\mathbb{R})$ on the open unit disk, D. Technically, it is easier to consider the group, $\mathbf{SU}(1,1)$, which is isomorphic to $\mathbf{SL}(2,\mathbb{R})$, and to make $\mathbf{SU}(1,1)$ act on D. The group $\mathbf{SU}(1,1)$ is the group of 2×2 complex matrices of the form

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix}, \quad \text{with} \quad a\overline{a} - b\overline{b} = 1.$$

The reader should check that if we let

$$g = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix},$$

then the map from $\mathbf{SL}(2,\mathbb{R})$ to $\mathbf{SU}(1,1)$ given by

$$A \mapsto gAg^{-1}$$

is an isomorphism. Observe that the Möbius transformation associated with g is

$$z \mapsto \frac{z-i}{z+1},$$

which is the holomorphic isomorphism mapping H to D mentionned earlier! Now, we can define a bijection between $\mathbf{SU}(1,1)$ and $S^1 \times D$ given by

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \mapsto (a/|a|, b/a).$$

We conclude that $\mathbf{SL}(2,\mathbb{R}) \cong \mathbf{SU}(1,1)$ is topologically an open solid torus (i.e., with the surface of the torus removed). It is possible to further classify the elements of $\mathbf{SL}(2,\mathbb{R})$ into three categories and to have geometric interpretations of these as certain regions of the torus. For details, the reader should consult Carter, Segal and Macdonald [31] or Duistermatt and Kolk [53] (Chapter 1, Section 1.2).

The group $\mathbf{SU}(1,1)$ acts on D by interpreting any matrix in $\mathbf{SU}(1,1)$ as a Möbius tranformation, i.e.,

$$\begin{pmatrix} a & b \\ \overline{b} & \overline{a} \end{pmatrix} \mapsto \left(z \mapsto \frac{az+b}{\overline{b}z + \overline{a}} \right).$$

The reader should check that these transformations preserve D. Both the upper half-plane and the open disk are models of Lobachevsky's non-Euclidean geometry (where the parallel postulate fails). They are also models of hyperbolic spaces (Riemannian manifolds with constant negative curvature, see Gallot, Hulin and Lafontaine [60], Chapter III). According to Dubrovin, Fomenko, and Novikov [51] (Chapter 2, Section 13.2), the open disk model is due to Poincaré and the upper half-plane model to Klein, although Poincaré was the first to realize that the upper half-plane is a hyperbolic space.

2.3 The Lorentz Groups $\mathbf{O}(n,1)$, $\mathbf{SO}(n,1)$ and $\mathbf{SO}_0(n,1)$

The Lorentz group provides another interesting example. Moreover, the Lorentz group $\mathbf{SO}(3,1)$ shows up in an interesting way in computer vision.

Denote the $p \times p$-identity matrix by I_p, for $p, q, \geq 1$, and define

$$I_{p,q} = \begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}.$$

If $n = p + q$, the matrix $I_{p,q}$ is associated with the nondegenerate symmetric bilinear form

$$\varphi_{p,q}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \sum_{i=1}^{p} x_i y_i - \sum_{j=p+1}^{n} x_j y_j$$

with associated quadratic form

$$\Phi_{p,q}((x_1, \ldots, x_n)) = \sum_{i=1}^{p} x_i^2 - \sum_{j=p+1}^{n} x_j^2.$$

In particular, when $p = 1$ and $q = 3$, we have the *Lorentz metric*

$$x_1^2 - x_2^2 - x_3^2 - x_4^2.$$

In physics, x_1 is interpreted as time and written t and x_2, x_3, x_4 as coordinates in $\mathbb{R}^3$ and written x, y, z. Thus, the Lozentz metric is usually written a

$$t^2 - x^2 - y^2 - z^2,$$

although it also appears as

$$x^2 + y^2 + z^2 - t^2,$$

which is equivalent but slightly less convenient for certain purposes, as we will see later. The space $\mathbb{R}^4$ with the Lorentz metric is called *Minkowski space*. It plays an important role in Einstein's theory of special relativity.

The group $\mathbf{O}(p, q)$ is the set of all $n \times n$-matrices

$$\mathbf{O}(p,q) = \{A \in \mathbf{GL}(n, \mathbb{R}) \mid A^\top I_{p,q} A = I_{p,q}\}.$$

This is the group of all invertible linear maps of $\mathbb{R}^n$ that preserve the quadratic form, $\Phi_{p,q}$, i.e., the group of isometries of $\Phi_{p,q}$. Clearly, $I_{p,q}^2 = I$, so the condition $A^\top I_{p,q} A = I_{p,q}$ is equivalent to $I_{p,q} A^\top I_{p,q} A = I$, which means that

$$A^{-1} = I_{p,q} A^\top I_{p,q}.$$

Thus, $AI_{p,q}A^\top = I_{p,q}$ also holds, which shows that $\mathbf{O}(p,q)$ is closed under transposition (i.e., if $A \in \mathbf{O}(p,q)$, then $A^\top \in \mathbf{O}(p,q)$). We have the subgroup

$$\mathbf{SO}(p,q) = \{A \in \mathbf{O}(p,q) \mid \det(A) = 1\}$$

consisting of the isometries of $(\mathbb{R}^n, \Phi_{p,q})$ with determinant $+1$. It is clear that $\mathbf{SO}(p,q)$ is also closed under transposition. The condition $A^\top I_{p,q} A = I_{p,q}$ has an interpretation in terms of the inner product $\varphi_{p,q}$ and the columns (and rows) of A. Indeed, if we denote the jth column of A by A_j, then

$$A^\top I_{p,q} A = (\varphi_{p,q}(A_i, A_j)),$$

so $A \in \mathbf{O}(p,q)$ iff the columns of A form an "orthonormal basis" w.r.t. $\varphi_{p,q}$, i.e.,

$$\varphi_{p,q}(A_i, A_j) = \begin{cases} \delta_{ij} & \text{if } 1 \leq i, j \leq p; \\ -\delta_{ij} & \text{if } p+1 \leq i, j \leq p+q. \end{cases}$$

The difference with the usual orthogonal matrices is that $\varphi_{p,q}(A_i, A_i) = -1$, if $p+1 \leq i \leq p+q$. As $\mathbf{O}(p,q)$ is closed under transposition, the rows of A also form an orthonormal basis w.r.t. $\varphi_{p,q}$.

It turns out that $\mathbf{SO}(p,q)$ has two connected components and the component containing the identity is a subgroup of $\mathbf{SO}(p,q)$ denoted $\mathbf{SO}_0(p,q)$. The group $\mathbf{SO}_0(p,q)$ turns out to be homeomorphic to $\mathbf{SO}(p) \times \mathbf{SO}(q) \times \mathbb{R}^{pq}$, but this is not easy to prove. (One way to prove it is to use results on pseudo-algebraic subgroups of $\mathbf{GL}(n,\mathbb{C})$, see Knapp [89] or Gallier's notes on Clifford algebras (on the web)).

We will now determine the polar decomposition and the SVD decomposition of matrices in the Lorentz groups $\mathbf{O}(n,1)$ and $\mathbf{SO}(n,1)$. Write $J = I_{n,1}$ and, given any $A \in \mathbf{O}(n,1)$, write

$$A = \begin{pmatrix} B & u \\ v^\top & c \end{pmatrix},$$

where B is an $n \times n$ matrix, u, v are (column) vectors in $\mathbb{R}^n$ and $c \in \mathbb{R}$. We begin with the polar decomposition of matrices in the Lorentz groups $\mathbf{O}(n,1)$.

Proposition 2.3 *Every matrix $A \in \mathbf{O}(n,1)$ has a polar decomposition of the form*

$$A = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix},$$

where $Q \in \mathbf{O}(n)$ and $c = \sqrt{\|v\|^2 + 1}$.

Proof. Write A in block form as above. As the condition for A to be in $\mathbf{O}(n,1)$ is $A^\top J A = J$, we get

$$\begin{pmatrix} B^\top & v \\ u^\top & c \end{pmatrix} \begin{pmatrix} B & u \\ -v^\top & -c \end{pmatrix} = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix},$$

i.e,.

$$\begin{aligned} B^\top B &= I + vv^\top \\ u^\top u &= c^2 - 1 \\ B^\top u &= cv. \end{aligned}$$

If we remember that we also have $AJA^\top = J$, then

$$Bv = cu,$$

which can also be deduced from the three equations above. From $u^\top u = \|u\|^2 = c^2 - 1$, we deduce that $|c| \geq 1$, and from $B^\top B = I + vv^\top$, we deduce that $B^\top B$ is symmetric, positive definite. Now, geometrically, it is well known that $vv^\top / v^\top v$ is the orthogonal projection onto the line determined by v. Consequently, the kernel of $vv^\top$ is the orthogonal complement of v and $vv^\top$ has the eigenvalue 0 with multiplicity $n-1$ and the eigenvalue $c^2 - 1 = \|v\|^2 = v^\top v$ with multiplicity 1. The eigenvectors associated with 0 are orthogonal to v and the eigenvectors associated with $c^2 - 1$ are proportional with v. It follows that $I + vv^\top$ has the eigenvalue 1 with multiplicity $n-1$ and the eigenvalue c^2 with multiplicity 1, the eigenvectors being as before. Now, B has polar form $B = QS_1$, where Q is orthogonal and S_1 is symmetric positive definite and $S_1^2 = B^\top B = I + vv^\top$. Therefore, if $c > 0$, then $S_1 = \sqrt{I + vv^\top}$ is a symmetric positive definite matrix with eigenvalue 1 with multiplicity $n-1$ and eigenvalue c with multiplicity 1, the eigenvectors being as before. If $c < 0$, then change c to $-c$.

Case 1: $c > 0$. Then, v is an eigenvector of S_1 for c and we must also have $Bv = cu$, which implies

$$Bv = QS_1 v = Q(cv) = cQv = cu,$$

so

$$Qv = u.$$

It follows that

$$A = \begin{pmatrix} B & u \\ v^\top & c \end{pmatrix} = \begin{pmatrix} QS_1 & Qv \\ v^\top & c \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix}.$$

Therefore, the polar decomposition of $A \in \mathbf{O}(n,1)$ is

$$A = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix},$$

where $Q \in \mathbf{O}(n)$ and $c = \sqrt{\|v\|^2 + 1}$.

Case 2: $c < 0$. Then, v is an eigenvector of S_1 for $-c$ and we must also have $Bv = cu$, which implies

$$Bv = QS_1 v = Q(-cv) = cQ(-v) = cu,$$

so

$$Q(-v) = u.$$

It follows that

$$A = \begin{pmatrix} B & u \\ v^\top & c \end{pmatrix} = \begin{pmatrix} QS_1 & Q(-v) \\ v^\top & c \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & -v \\ -v^\top & -c \end{pmatrix}.$$

In this case, the polar decomposition of $A \in \mathbf{O}(n,1)$ is

$$A = \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & -v \\ -v^\top & -c \end{pmatrix},$$

where $Q \in \mathbf{O}(n)$ and $c = -\sqrt{\|v\|^2 + 1}$. Therefore, we conclude that any $A \in \mathbf{O}(n,1)$ has a polar decomposition of the form

$$A = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix} \quad \text{or} \quad A = \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix},$$

where $Q \in \mathbf{O}(n)$ and $c = \sqrt{\|v\|^2 + 1}$. $\square$

Thus, we see that $\mathbf{O}(n,1)$ has four components corresponding to the cases:

(1) $Q \in \mathbf{O}(n)$; $\det(Q) < 0$; $+1$ as the lower right entry of the orthogonal matrix;

(2) $Q \in \mathbf{SO}(n)$; -1 as the lower right entry of the orthogonal matrix;

(3) $Q \in \mathbf{O}(n)$; $\det(Q) < 0$; -1 as the lower right entry of the orthogonal matrix;

(4) $Q \in \mathbf{SO}(n)$; $+1$ as the lower right entry of the orthogonal matrix.

Observe that $\det(A) = -1$ in cases (1) and (2) and that $\det(A) = +1$ in cases (3) and (4). Thus, (3) and (4) correspond to the group $\mathbf{SO}(n,1)$, in which case the polar decomposition is of the form

$$A = \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix},$$

where $Q \in \mathbf{O}(n)$, with $\det(Q) = -1$ and $c = \sqrt{\|v\|^2 + 1}$ or

$$A = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix}$$

where $Q \in \mathbf{SO}(n)$ and $c = \sqrt{\|v\|^2 + 1}$. The components in (1) and (2) are not groups. We will show later that all four components are connected and that case (4) corresponds to a group (Proposition 2.8). This group is the connected component of the identity and it is denoted $\mathbf{SO}_0(n,1)$ (see Corollary 2.27). For the time being, note that $A \in \mathbf{SO}_0(n,1)$ iff

$A \in \mathbf{SO}(n,1)$ and $a_{n+1\,n+1}$ $(= c) > 0$ (here, $A = (a_{i\,j})$.) In fact, we proved above that if $a_{n+1\,n+1} > 0$, then $a_{n+1\,n+1} \geq 1$.

Remark: If we let

$$\Lambda_P = \begin{pmatrix} I_{n-1,1} & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \Lambda_T = I_{n,1}, \quad \text{where} \quad I_{n,1} = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix},$$

then we have the disjoint union

$$\mathbf{O}(n,1) = \mathbf{SO}_0(n,1) \cup \Lambda_P \mathbf{SO}_0(n,1) \cup \Lambda_T \mathbf{SO}_0(n,1) \cup \Lambda_P \Lambda_T \mathbf{SO}_0(n,1).$$

In order to determine the SVD of matrices in $\mathbf{SO}_0(n,1)$, we analyze the eigenvectors and the eigenvalues of the positive definite symmetric matrix

$$S = \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix}$$

involved in Proposition 2.3. Such a matrix is called a *Lorentz boost*. Observe that if $v = 0$, then $c = 1$ and $S = I_{n+1}$.

Proposition 2.4 *Assume $v \neq 0$. The eigenvalues of the symmetric positive definite matrix*

$$S = \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix},$$

where $c = \sqrt{\|v\|^2 + 1}$, are 1 with multiplicity $n-1$, and e^α and $e^{-\alpha}$ each with multiplicity 1 (for some $\alpha \geq 0$). An orthonormal basis of eigenvectors of S consists of vectors of the form

$$\begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} u_{n-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{v}{\sqrt{2}\|v\|} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{v}{\sqrt{2}\|v\|} \\ -\frac{1}{\sqrt{2}} \end{pmatrix},$$

where the $u_i \in \mathbb{R}^n$ are all orthogonal to v and pairwise orthogonal.

Proof. Let us solve the linear system

$$\begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix} \begin{pmatrix} v \\ d \end{pmatrix} = \lambda \begin{pmatrix} v \\ d \end{pmatrix}.$$

We get

$$\begin{aligned} \sqrt{I + vv^\top}(v) + dv &= \lambda v \\ v^\top v + cd &= \lambda d, \end{aligned}$$

that is (since $c = \sqrt{\|v\|^2 + 1}$ and $\sqrt{I + vv^\top}(v) = cv$),

$$\begin{aligned}(c+d)v &= \lambda v\\ c^2 - 1 + cd &= \lambda d.\end{aligned}$$

Since $v \neq 0$, we get $\lambda = c + d$. Substituting in the second equation, we get

$$c^2 - 1 + cd = (c+d)d,$$

that is,

$$d^2 = c^2 - 1.$$

Thus, either $\lambda_1 = c + \sqrt{c^2-1}$ and $d = \sqrt{c^2-1}$, or $\lambda_2 = c - \sqrt{c^2-1}$ and $d = -\sqrt{c^2-1}$. Since $c \geq 1$ and $\lambda_1\lambda_2 = 1$, set $\alpha = \log(c + \sqrt{c^2-1}) \geq 0$, so that $-\alpha = \log(c - \sqrt{c^2-1})$ and then, $\lambda_1 = e^\alpha$ and $\lambda_2 = e^{-\alpha}$. On the other hand, if u is orthogonal to v, observe that

$$\begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix} \begin{pmatrix} u \\ 0 \end{pmatrix} = \begin{pmatrix} u \\ 0 \end{pmatrix},$$

since the kernel of $vv^\top$ is the orthogonal complement of v. The rest is clear. □

Corollary 2.5 *The singular values of any matrix $A \in \mathbf{O}(n,1)$ are 1 with multiplicity $n-1$, e^α, and $e^{-\alpha}$, for some $\alpha \geq 0$.*

Note that the case $\alpha = 0$ is possible, in which case, A is an orthogonal matrix of the form

$$\begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} Q & 0 \\ 0 & -1 \end{pmatrix},$$

with $Q \in \mathbf{O}(n)$. The two singular values e^α and $e^{-\alpha}$ tell us how much A deviates from being orthogonal.

We can now determine a convenient form for the SVD of matrices in $\mathbf{O}(n,1)$.

Theorem 2.6 *Every matrix $A \in \mathbf{O}(n,1)$ can be written as*

$$A = \begin{pmatrix} P & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cosh\alpha & \sinh\alpha \\ 0 & \cdots & 0 & \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} Q^\top & 0 \\ 0 & 1 \end{pmatrix}$$

with $\epsilon = \pm 1$, $P \in \mathbf{O}(n)$ and $Q \in \mathbf{SO}(n)$. When $A \in \mathbf{SO}(n,1)$, we have $\det(P)\epsilon = +1$, and when $A \in \mathbf{SO}_0(n,1)$, we have $\epsilon = +1$ and $P \in \mathbf{SO}(n)$, that is,

$$A = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cosh\alpha & \sinh\alpha \\ 0 & \cdots & 0 & \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} Q^\top & 0 \\ 0 & 1 \end{pmatrix}$$

with $P \in \mathbf{SO}(n)$ and $Q \in \mathbf{SO}(n)$.

Proof. By Proposition 2.3, any matrix $A \in \mathbf{O}(n)$ can be written as

$$A = \begin{pmatrix} R & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} \sqrt{I + vv^\top} & v \\ v^\top & c \end{pmatrix}$$

where $\epsilon = \pm 1$, $R \in \mathbf{O}(n)$ and $c = \sqrt{\|v\|^2 + 1}$. The case where $c = 1$ is trivial, so assume $c > 1$, which means that α from Proposition 2.4 is such that $\alpha > 0$. The key fact is that the eigenvalues of the matrix

$$\begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}$$

are e^α and $e^{-\alpha}$ and that

$$\begin{pmatrix} e^\alpha & 0 \\ 0 & e^{-\alpha} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

From this fact, we see that the diagonal matrix

$$D = \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & e^\alpha & 0 \\ 0 & \cdots & 0 & 0 & e^{-\alpha} \end{pmatrix}$$

of eigenvalues of S is given by

$$D = \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cosh\alpha & \sinh\alpha \\ 0 & \cdots & 0 & \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}.$$

By Proposition 2.4, an orthonormal basis of eigenvectors of S consists of vectors of the form

$$\begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} u_{n-1} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{v}{\sqrt{2}\|v\|} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \begin{pmatrix} \frac{v}{\sqrt{2}\|v\|} \\ -\frac{1}{\sqrt{2}} \end{pmatrix},$$

where the $u_i \in \mathbb{R}^n$ are all orthogonal to v and pairwise orthogonal. Now, if we multiply the matrices

$$\begin{pmatrix} u_1 & \cdots & u_{n-1} & \frac{v}{\sqrt{2}\|v\|} & \frac{v}{\sqrt{2}\|v\|} \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 0 & \cdots & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix},$$

we get an orthogonal matrix of the form

$$\begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}$$

where the columns of Q are the vectors

$$u_1, \cdots, u_{n-1}, \frac{v}{\|v\|}.$$

By flipping u_1 to $-u_1$ if necessary, we can make sure that this matrix has determinant $+1$. Consequently,

$$S = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cosh\alpha & \sinh\alpha \\ 0 & \cdots & 0 & \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} Q^\top & 0 \\ 0 & 1 \end{pmatrix},$$

so

$$A = \begin{pmatrix} R & 0 \\ 0 & \epsilon \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & \cdots & 0 & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 & 0 \\ 0 & \cdots & 0 & \cosh\alpha & \sinh\alpha \\ 0 & \cdots & 0 & \sinh\alpha & \cosh\alpha \end{pmatrix} \begin{pmatrix} Q^\top & 0 \\ 0 & 1 \end{pmatrix},$$

and if we let $P = RQ$, we get the desired decomposition. □

Remark: We warn our readers about Chapter 6 of Baker's book [13]. Indeed, this chapter is seriously flawed. The main two Theorems (Theorem 6.9 and Theorem 6.10) are false and as consequence, the proof of Theorem 6.11 is wrong too. Theorem 6.11 states that the exponential map $\exp\colon \mathfrak{so}(n,1) \to \mathbf{SO}_0(n,1)$ is surjective, which is correct, but known proofs are nontrivial and quite lengthy (see Section 5.5). The proof of Theorem 6.12 is also false, although the theorem itself is correct (this is our Theorem 5.22, see Section 5.5). The main problem with Theorem 6.9 (in Baker) is that the existence of the normal form for matrices in $\mathbf{SO}_0(n,1)$ claimed by this theorem is unfortunately false on several accounts. Firstly, it would imply that every matrix in $\mathbf{SO}_0(n,1)$ can be diagonalized, but this is false for $n \geq 2$. Secondly, even if a matrix $A \in \mathbf{SO}_0(n,1)$ is diagonalizable as $A = PDP^{-1}$, Theorem 6.9 (and Theorem 6.10) miss some possible eigenvalues and the matrix P is not necessarily in $\mathbf{SO}_0(n,1)$ (as the case $n = 1$ already shows). For a thorough analysis of the eigenvalues of Lorentz isometries (and much more), one should consult Riesz [126] (Chapter III).

Clearly, a result similar to Theorem 2.6 also holds for the matrices in the groups $\mathbf{O}(1,n)$,

$\mathbf{SO}(1,n)$ and $\mathbf{SO}_0(1,n)$. For example, every matrix $A \in \mathbf{SO}_0(1,n)$ can be written as

$$A = \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix} \begin{pmatrix} \cosh\alpha & \sinh\alpha & 0 & \cdots & 0 \\ \sinh\alpha & \cosh\alpha & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & Q^\top \end{pmatrix},$$

where $P, Q \in \mathbf{SO}(n)$.

In the case $n = 3$, we obtain the *proper orthochronous Lorentz group*, $\mathbf{SO}_0(1,3)$, also denoted $\mathbf{Lor}(1,3)$. By the way, $\mathbf{O}(1,3)$ is called the *(full) Lorentz group* and $\mathbf{SO}(1,3)$ is the *special Lorentz group*.

Theorem 2.6 (really, the version for $\mathbf{SO}_0(1,n)$) shows that the Lorentz group $\mathbf{SO}_0(1,3)$ is generated by the matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix} \quad \text{with } P \in \mathbf{SO}(3)$$

and the matrices of the form

$$\begin{pmatrix} \cosh\alpha & \sinh\alpha & 0 & 0 \\ \sinh\alpha & \cosh\alpha & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

This fact will be useful when we prove that the homomorphism $\varphi\colon \mathbf{SL}(2,\mathbb{C}) \to \mathbf{SO}_0(1,3)$ is surjective.

Remark: Unfortunately, unlike orthogonal matrices which can always be diagonalized over $\mathbb{C}$, **not** every matrix in $\mathbf{SO}(1,n)$ can be diagonalized for $n \geq 2$. This has to do with the fact that the Lie algebra $\mathfrak{so}(1,n)$ has non-zero idempotents (see Section 5.5).

It turns out that the group $\mathbf{SO}_0(1,3)$ admits another interesting characterization involving the hypersurface

$$\mathcal{H} = \{(t,x,y,z) \in \mathbb{R}^4 \mid t^2 - x^2 - y^2 - z^2 = 1\}.$$

This surface has two sheets and it is not hard to show that $\mathbf{SO}_0(1,3)$ is the subgroup of $\mathbf{SO}(1,3)$ that preserves these two sheets (does not swap them). Actually, we will prove this fact for any n. In preparation for this we need some definitions and a few propositions.

Let us switch back to $\mathbf{SO}(n,1)$. First, as a matter of notation, we write every $u \in \mathbb{R}^{n+1}$ as $u = (\mathbf{u}, t)$, where $\mathbf{u} \in \mathbb{R}^n$ and $t \in \mathbb{R}$, so that the Lorentz inner product can be expressed as

$$\langle u, v\rangle = \langle(\mathbf{u},t),(\mathbf{v},s)\rangle = \mathbf{u}\cdot\mathbf{v} - ts,$$

where $\mathbf{u}\cdot\mathbf{v}$ is the standard Euclidean inner product (the Euclidean norm of x is denoted $\|x\|$). Then, we can classify the vectors in $\mathbb{R}^{n+1}$ as follows:

Definition 2.10 A nonzero vector, $u = (\mathbf{u}, t) \in \mathbb{R}^{n+1}$ is called

(a) *spacelike* iff $\langle u, u\rangle > 0$, i.e., iff $\|\mathbf{u}\|^2 > t^2$;

(b) *timelike* iff $\langle u, u\rangle < 0$, i.e., iff $\|\mathbf{u}\|^2 < t^2$;

(c) *lightlike* or *isotropic* iff $\langle u, u\rangle = 0$, i.e., iff $\|\mathbf{u}\|^2 = t^2$.

A spacelike (resp. timelike, resp. lightlike) vector is said to be *positive* iff $t > 0$ and *negative* iff $t < 0$. The set of all isotropic vectors

$$\mathcal{H}_n(0) = \{u = (\mathbf{u}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{u}\|^2 = t^2\}$$

is called the *light cone*. For every $r > 0$, let

$$\mathcal{H}_n(r) = \{u = (\mathbf{u}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{u}\|^2 - t^2 = -r\},$$

a hyperboloid of two sheets.

It is easy to check that $\mathcal{H}_n(r)$ has two connected components as follows: First, since $r > 0$ and

$$\|\mathbf{u}\|^2 + r = t^2,$$

we have $|t| \geq \sqrt{r}$. Now, for any $x = (x_1, \ldots, x_n, t) \in \mathcal{H}_n(r)$ with $t \geq \sqrt{r}$, we have the continuous path from $(0, \ldots, 0, \sqrt{r})$ to x given by

$$\lambda \mapsto (\lambda x_1, \ldots, \lambda x_n, \sqrt{r + \lambda^2(t^2 - r)}),$$

where $\lambda \in [0, 1]$, proving that the component of $(0, \ldots, 0, \sqrt{r})$ is connected. Similarly, when $t \leq -\sqrt{r}$, we have the continuous path from $(0, \ldots, 0, -\sqrt{r})$ to x given by

$$\lambda \mapsto (\lambda x_1, \ldots, \lambda x_n, -\sqrt{r + \lambda^2(t^2 - r)}),$$

where $\lambda \in [0, 1]$, proving that the component of $(0, \ldots, 0, -\sqrt{r})$ is connected. We denote the sheet containing $(0, \ldots, 0, \sqrt{r})$ by $\mathcal{H}_n^+(r)$ and sheet containing $(0, \ldots, 0, -\sqrt{r})$ by $\mathcal{H}_n^-(r)$

Since every Lorentz isometry, $A \in \mathbf{SO}(n, 1)$, preserves the Lorentz inner product, we conclude that A globally preserves every hyperboloid, $\mathcal{H}_n(r)$, for $r > 0$. We claim that every $A \in \mathbf{SO}_0(n, 1)$ preserves both $\mathcal{H}_n^+(r)$ and $\mathcal{H}_n^-(r)$. This follows immediately from

Proposition 2.7 *If $a_{n+1\,n+1} > 0$, then every isometry, $A \in \mathbf{O}(n, 1)$, preserves all positive (resp. negative) timelike vectors and all positive (resp. negative) lightlike vectors. Moreover, if $A \in \mathbf{O}(n, 1)$ preserves all positive timelike vectors, then $a_{n+1\,n+1} > 0$.*

Proof. Let $u = (\mathbf{u}, t)$ be a nonzero timelike or lightlike vector. This means that

$$\|\mathbf{u}\|^2 \leq t^2 \quad \text{and} \quad t \neq 0.$$

Since $A \in \mathbf{O}(n,1)$, the matrix A preserves the inner product; if $\langle u, u\rangle = \|\mathbf{u}\|^2 - t^2 < 0$, we get $\langle Au, Au\rangle < 0$, which shows that Au is also timelike. Similarly, if $\langle u, u\rangle = 0$, then $\langle Au, Au\rangle = 0$. As $A \in \mathbf{O}(n,1)$, we know that

$$\langle A_{n+1}, A_{n+1}\rangle = -1,$$

that is,

$$\|\mathbf{A}_{n+1}\|^2 - a^2_{n+1,\,n+1} = -1,$$

where $A_{n+1} = (\mathbf{A}_{n+1}, a_{n+1,\,n+1})$ is the $(n+1)$th row of the matrix A. The $(n+1)$th component of the vector Au is

$$\mathbf{u} \cdot \mathbf{A}_{n+1} + a_{n+1,\,n+1} t.$$

By Cauchy-Schwarz,

$$(\mathbf{u} \cdot \mathbf{A}_{n+1})^2 \leq \|\mathbf{u}\|^2 \|\mathbf{A}_{n+1}\|^2,$$

so we get,

$$\begin{aligned}(\mathbf{u} \cdot \mathbf{A}_{n+1})^2 &\leq \|\mathbf{u}\|^2 \|\mathbf{A}_{n+1}\|^2 \\ &\leq t^2(a^2_{n+1,\,n+1} - 1) = t^2 a^2_{n+1,\,n+1} - t^2 \\ &< t^2 a^2_{n+1,\,n+1},\end{aligned}$$

since $t \neq 0$. It follows that $\mathbf{u} \cdot \mathbf{A}_{n+1} + a_{n+1,\,n+1} t$ has the same sign as t, since $a_{n+1,\,n+1} > 0$. Consequently, if $a_{n+1,\,n+1} > 0$, we see that A maps positive timelike (resp. lightlike) vectors to positive timelike (resp. lightlike) vectors and similarly with negative timelight (resp. lightlike) vectors.

Conversely, as $e_{n+1} = (0, \ldots, 0, 1)$ is timelike and positive, if A preserves all positive timelike vectors, then Ae_{n+1} is timelike positive, which implies $a_{n+1,\,n+1} > 0$. □

Let $\mathbf{O}^+(n,1)$ denote the subset of $\mathbf{O}(n,1)$ consisting of all matrices, $A = (a_{i\,j})$, such that $a_{n+1\,n+1} > 0$. Using Proposition 2.7, we can now show that $\mathbf{O}^+(n,1)$ is a subgroup of $\mathbf{O}(n,1)$ and that $\mathbf{SO}_0(n,1)$ is a subgroup of $\mathbf{SO}(n,1)$. Recall that

$$\mathbf{SO}_0(n,1) = \{A \in \mathbf{SO}(n,1) \mid a_{n+1\,n+1} > 0\}.$$

Note that $\mathbf{SO}_0(n,1) = \mathbf{O}^+(n,1) \cap \mathbf{SO}(n,1)$.

Proposition 2.8 *The set* $\mathbf{O}^+(n,1)$ *is a subgroup of* $\mathbf{O}(n,1)$ *and the set* $\mathbf{SO}_0(n,1)$ *is a subgroup of* $\mathbf{SO}(n,1)$.

Proof. Let $A \in \mathbf{O}^+(n,1) \subseteq \mathbf{O}(n,1)$, so that $a_{n+1\,n+1} > 0$. The inverse of A in $\mathbf{O}(n,1)$ is $JA^\top J$, where

$$J = \begin{pmatrix} I_n & 0 \\ 0 & -1 \end{pmatrix},$$

which implies that $a^{-1}_{n+1\,n+1} = a_{n+1\,n+1} > 0$ and so, $A^{-1} \in \mathbf{O}^+(n,1)$. If $A, B \in \mathbf{O}^+(n,1)$, then, by Proposition 2.7, both A and B preserve all positive timelike vectors, so AB preserve all positive timelike vectors. By Proposition 2.7, again, $AB \in \mathbf{O}^+(n,1)$. Therefore, $\mathbf{O}^+(n,1)$ is a group. But then, $\mathbf{SO}_0(n,1) = \mathbf{O}^+(n,1) \cap \mathbf{SO}(n,1)$ is also a group. □

Since any matrix, $A \in \mathbf{SO}_0(n,1)$, preserves the Lorentz inner product and all positive timelike vectors and since $\mathcal{H}^+_n(1)$ consists of timelike vectors, we see that every $A \in \mathbf{SO}_0(n,1)$ maps $\mathcal{H}^+_n(1)$ into itself. Similarly, every $A \in \mathbf{SO}_0(n,1)$ maps $\mathcal{H}^-_n(1)$ into itself. Thus, we can define an action $\cdot\colon \mathbf{SO}_0(n,1) \times \mathcal{H}^+_n(1) \longrightarrow \mathcal{H}^+_n(1)$ by

$$A \cdot u = Au$$

and similarly, we have an action $\cdot\colon \mathbf{SO}_0(n,1) \times \mathcal{H}^-_n(1) \longrightarrow \mathcal{H}^-_n(1)$.

Proposition 2.9 *The group $\mathbf{SO}_0(n,1)$ is the subgroup of $\mathbf{SO}(n,1)$ that preserves $\mathcal{H}^+_n(1)$ (and $\mathcal{H}^-_n(1)$) i.e.,*

$$\mathbf{SO}_0(n,1) = \{A \in \mathbf{SO}(n,1) \mid A(\mathcal{H}^+_n(1)) = \mathcal{H}^+_n(1) \quad \textit{and} \quad A(\mathcal{H}^-_n(1)) = \mathcal{H}^-_n(1)\}.$$

Proof. We already observed that $A(\mathcal{H}^+_n(1)) = \mathcal{H}^+_n(1)$ if $A \in \mathbf{SO}_0(n,1)$ (and similarly, $A(\mathcal{H}^-_n(1)) = \mathcal{H}^-_n(1)$). Conversely, for any $A \in \mathbf{SO}(n,1)$ such that $A(\mathcal{H}^+_n(1)) = \mathcal{H}^+_n(1)$, as $e_{n+1} = (0, \ldots, 0, 1) \in \mathcal{H}^+_n(1)$, the vector Ae_{n+1} must be positive timelike, but this says that $a_{n+1,\,n+1} > 0$, i.e., $A \in \mathbf{SO}_0(n,1)$. □

Next, we wish to prove that the action $\mathbf{SO}_0(n,1) \times \mathcal{H}^+_n(1) \longrightarrow \mathcal{H}^+_n(1)$ is transitive. For this, we need the next two propositions.

Proposition 2.10 *Let $u = (\mathbf{u}, t)$ and $v = (\mathbf{v}, s)$ be nonzero vectors in $\mathbb{R}^{n+1}$ with $\langle u, v\rangle = 0$. If u is timelike, then v is spacelike (i.e., $\langle v, v\rangle > 0$).*

Proof. We have $\|\mathbf{u}\|^2 < t^2$, so $t \neq 0$. Since $\mathbf{u} \cdot \mathbf{v} - ts = 0$, we get

$$\langle v, v\rangle = \|\mathbf{v}\|^2 - s^2 = \|\mathbf{v}\|^2 - \frac{(\mathbf{u} \cdot \mathbf{v})^2}{t^2}.$$

But, Cauchy-Schwarz implies that $(\mathbf{u} \cdot \mathbf{v})^2 \leq \|\mathbf{u}\|^2 \|\mathbf{v}\|^2$, so we get

$$\langle v, v\rangle = \|\mathbf{v}\|^2 - \frac{(\mathbf{u} \cdot \mathbf{v})^2}{t^2} > \|\mathbf{v}\|^2 - \frac{(\mathbf{u} \cdot \mathbf{v})^2}{\|\mathbf{u}\|^2} \geq 0,$$

as $\|\mathbf{u}\|^2 < t^2$. □

Lemma 2.10 also holds if $u = (\mathbf{u}, t)$ is a nonzero isotropic vector and $v = (\mathbf{v}, s)$ is a nonzero vector that is not collinear with u: If $\langle u, v\rangle = 0$, then v is spacelike (i.e., $\langle v, v\rangle > 0$). The proof is left as an exercise to the reader.

Proposition 2.11 *The action* $\mathbf{SO}_0(n,1) \times \mathcal{H}_n^+(1) \longrightarrow \mathcal{H}_n^+(1)$ *is transitive.*

Proof. Let $e_{n+1} = (0, \ldots, 0, 1) \in \mathcal{H}_n^+(1)$. It is enough to prove that for every $u = (\mathbf{u}, t) \in \mathcal{H}_n^+(1)$, there is some $A \in \mathbf{SO}_0(n,1)$ such that $Ae_{n+1} = u$. By hypothesis,

$$\langle u, u\rangle = \|\mathbf{u}\|^2 - t^2 = -1.$$

We show that we can construct an orthonormal basis, $e_1, \ldots, e_n, u$, with respect to the Lorentz inner product. Consider the hyperplane

$$H = \{v \in \mathbb{R}^{n+1} \mid \langle u, v\rangle = 0\}.$$

Since u is timelike, by Proposition 2.10, every nonzero vector $v \in H$ is spacelike, i.e., $\langle v, v\rangle > 0$. Let $v_1, \ldots, v_n$ be a basis of H. Since all (nonzero) vectors in H are spacelike, we can apply the Gramm-Schmidt orthonormalization procedure and we get a basis $e_1, \ldots, e_n$, of H, such that

$$\langle e_i, e_j\rangle = \delta_{i,j}, \quad 1 \leq i, j \leq n.$$

Now, by construction, we also have

$$\langle e_i, u\rangle = 0, \quad 1 \leq i \leq n, \quad \text{and} \quad \langle u, u\rangle = -1.$$

Therefore, $e_1, \ldots, e_n, u$ are the column vectors of a Lorentz matrix, A, such that $Ae_{n+1} = u$, proving our assertion. □

Let us find the stabilizer of $e_{n+1} = (0, \ldots, 0, 1)$. We must have $Ae_{n+1} = e_{n+1}$, and the polar form implies that

$$A = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{with} \quad P \in \mathbf{SO}(n).$$

Therefore, the stabilizer of e_{n+1} is isomorphic to $\mathbf{SO}(n)$ and we conclude that $\mathcal{H}_n^+(1)$, as a homogeneous space, is

$$\mathcal{H}_n^+(1) \cong \mathbf{SO}_0(n,1)/\mathbf{SO}(n).$$

We will show in Section 2.5 that $\mathbf{SO}_0(n,1)$ is connected.

2.4 More on $\mathbf{O}(p,q)$

Recall from Section 2.3 that the group $\mathbf{O}(p,q)$ is the set of all $n \times n$-matrices

$$\mathbf{O}(p,q) = \{A \in \mathbf{GL}(n,\mathbb{R}) \mid A^\top I_{p,q} A = I_{p,q}\}.$$

We deduce immediately that $|\det(A)| = 1$ and we also know that $AI_{p,q}A^\top = I_{p,q}$ holds. Unfortunately, when $p \neq 0, 1$ and $q \neq 0, 1$, it does not seem possible to obtain a formula as nice as that given in Proposition 2.3. Nevertheless, we can obtain a formula for the polar form of matrices in $\mathbf{O}(p,q)$. First, recall (for example, see Gallier [58], Chapter 12) that if

S is a symmetric positive definite matrix, then there is a unique symmetric positive definite matrix, T, so that

$$S = T^2.$$

We denote T by $S^{\frac{1}{2}}$ or $\sqrt{S}$. By $S^{-\frac{1}{2}}$, we mean the inverse of $S^{\frac{1}{2}}$. In order to obtain the polar form of a matrix in $\mathbf{O}(p,q)$, we begin with the following proposition:

Proposition 2.12 *Every matrix $X \in \mathbf{O}(p,q)$ can be written as*

$$X = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} \alpha^{\frac{1}{2}} & \alpha^{\frac{1}{2}} Z^\top \\ \delta^{\frac{1}{2}} Z & \delta^{\frac{1}{2}} \end{pmatrix},$$

where $\alpha = (I - Z^\top Z)^{-1}$ and $\delta = (I - ZZ^\top)^{-1}$, for some orthogonal matrices $U \in \mathbf{O}(p)$, $V \in \mathbf{O}(q)$ and for some $q \times p$ matrix, Z, such that $I - Z^\top Z$ and $I - ZZ^\top$ are symmetric positive definite matrices. Moreover, U, V, Z are uniquely determined by X.

Proof. If we write

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

with A a $p \times p$ matrix, D a $q \times q$ matrix, B a $p \times q$ matrix and C a $q \times p$ matrix, then the equations $A^\top I_{p,q} A = I_{p,q}$ and $A I_{p,q} A^\top = I_{p,q}$ yield the (not independent) conditions

$$\begin{aligned} A^\top A &= I + C^\top C \\ D^\top D &= I + B^\top B \\ A^\top B &= C^\top D \\ AA^\top &= I + BB^\top \\ DD^\top &= I + CC^\top \\ AC^\top &= BD^\top. \end{aligned}$$

Since $C^\top C$ is symmetric and since it is easy to show that $C^\top C$ has nonnegative eigenvalues, we deduce that $A^\top A$ is symmetric positive definite and similarly for $D^\top D$. If we assume that the above decomposition of X holds, we deduce that

$$\begin{aligned} A &= U(I - Z^\top Z)^{-\frac{1}{2}} \\ B &= U(I - Z^\top Z)^{-\frac{1}{2}} Z^\top \\ C &= V(I - ZZ^\top)^{-\frac{1}{2}} Z \\ D &= V(I - ZZ^\top)^{-\frac{1}{2}}, \end{aligned}$$

which implies

$$Z = D^{-1}C \quad \text{and} \quad Z^\top = A^{-1}B.$$

Thus, we must check that

$$(D^{-1}C)^\top = A^{-1}B$$

i.e.,

$$C^\top (D^\top)^{-1} = A^{-1}B,$$

namely,

$$AC^\top = BD^\top,$$

which is indeed the last of our identities. Thus, we must have $Z = D^{-1}C = (A^{-1}B)^\top$. The above expressions for A and D also imply that

$$A^\top A = (I - Z^\top Z)^{-1} \quad \text{and} \quad D^\top D = (I - ZZ^\top)^{-1},$$

so we must check that the choice $Z = D^{-1}C = (A^{-1}B)^\top$ yields the above equations.

Since $Z^\top = A^{-1}B$, we have

$$\begin{aligned} Z^\top Z &= A^{-1}BB^\top (A^\top)^{-1} \\ &= A^{-1}(AA^\top - I)(A^\top)^{-1} \\ &= I - A^{-1}(A^\top)^{-1} \\ &= I - (A^\top A)^{-1}. \end{aligned}$$

Therefore,

$$(A^\top A)^{-1} = I - Z^\top Z,$$

i.e.,

$$A^\top A = (I - Z^\top Z)^{-1},$$

as desired. We also have, this time, with $Z = D^{-1}C$,

$$\begin{aligned} ZZ^\top &= D^{-1}CC^\top (D^\top)^{-1} \\ &= D^{-1}(DD^\top - I)(D^\top)^{-1} \\ &= I - D^{-1}(D^\top)^{-1} \\ &= I - (D^\top D)^{-1}. \end{aligned}$$

Therefore,

$$(D^\top D)^{-1} = I - ZZ^\top,$$

i.e.,

$$D^\top D = (I - ZZ^\top)^{-1},$$

as desired. Now, since $A^\top A$ and $D^\top D$ are positive definite, the polar form implies that

$$A = U(A^\top A)^{\frac{1}{2}} = U(I - Z^\top Z)^{-\frac{1}{2}}$$

and

$$D = V(D^\top D)^{\frac{1}{2}} = V(I - ZZ^\top)^{-\frac{1}{2}},$$

for some unique matrices, $U \in \mathbf{O}(p)$ and $V \in \mathbf{O}(q)$. Since $Z = D^{-1}C$ and $Z^\top = A^{-1}B$, we get $C = DZ$ and $B = AZ^\top$, but this is

$$\begin{aligned} B &= U(I - Z^\top Z)^{-\frac{1}{2}} Z^\top \\ C &= V(I - ZZ^\top)^{-\frac{1}{2}} Z, \end{aligned}$$

as required. Therefore, the unique choice of $Z = D^{-1}C = (A^{-1}B)^\top$, U and V does yield the formula of the proposition. □

It remains to show that the matrix

$$\begin{pmatrix} \alpha^{\frac{1}{2}} & \alpha^{\frac{1}{2}} Z^\top \\ \delta^{\frac{1}{2}} Z & \delta^{\frac{1}{2}} \end{pmatrix} = \begin{pmatrix} (I - Z^\top Z)^{-\frac{1}{2}} & (I - Z^\top Z)^{-\frac{1}{2}} Z^\top \\ (I - ZZ^\top)^{-\frac{1}{2}} Z & (I - ZZ^\top)^{-\frac{1}{2}} \end{pmatrix}$$

is symmetric. To prove this, we will use power series and a continuity argument.

Proposition 2.13 *For any $q \times p$ matrix, Z, such that $I - Z^\top Z$ and $I - ZZ^\top$ are symmetric positive definite, the matrix*

$$S = \begin{pmatrix} \alpha^{\frac{1}{2}} & \alpha^{\frac{1}{2}} Z^\top \\ \delta^{\frac{1}{2}} Z & \delta^{\frac{1}{2}} \end{pmatrix}$$

is symmetric, where $\alpha = (I - Z^\top Z)^{-1}$ and $\delta = (I - ZZ^\top)^{-1}$.

Proof. The matrix S is symmetric iff

$$Z\alpha^{\frac{1}{2}} = \delta^{\frac{1}{2}} Z,$$

i.e., iff

$$Z(I - Z^\top Z)^{-\frac{1}{2}} = (I - ZZ^\top)^{-\frac{1}{2}} Z.$$

Consider the matrices

$$\beta(t) = (I - tZ^\top Z)^{-\frac{1}{2}} \quad \text{and} \quad \gamma(t) = (I - tZZ^\top)^{-\frac{1}{2}},$$

for any t with $0 \leq t \leq 1$. We claim that these matrices make sense. Indeed, since $Z^\top Z$ is symmetric, we can write

$$Z^\top Z = PDP^\top$$

where P is orthogonal and D is a diagonal matrix with nonnegative entries. Moreover, as

$$I - Z^\top Z = P(I - D)P^\top$$

and $I - Z^\top Z$ is positive definite, $0 \leq \lambda < 1$, for every eigenvalue in D. But then, as

$$I - tZ^\top Z = P(I - tD)P^\top,$$

we have $1 - t\lambda > 0$ for every λ in D and for all t with $0 \leq t \leq 1$, so that $I - tZ^\top Z$ is positive definite and thus, $(I - tZ^\top Z)^{-\frac{1}{2}}$ is also well defined. A similar argument applies to $(I - tZZ^\top)^{-\frac{1}{2}}$. Observe that

$$\lim_{t\to 1} \beta(t) = \alpha^{\frac{1}{2}}$$

since

$$\beta(t) = (I - tZ^\top Z)^{-\frac{1}{2}} = P(I - tD)^{-\frac{1}{2}} P^\top,$$

where $(I-tD)^{-\frac{1}{2}}$ is a diagonal matrix with entries of the form $(1-t\lambda)^{-\frac{1}{2}}$ and these eigenvalues are continuous functions of t for $t \in [0, 1]$. A similar argument shows that

$$\lim_{t\to 1} \gamma(t) = \delta^{\frac{1}{2}}.$$

Therefore, it is enough to show that

$$Z\beta(t) = \gamma(t)Z,$$

with $0 \leq t < 1$ and our result will follow by continuity. However, when $0 \leq t < 1$, the power series for $\beta(t)$ and $\gamma(t)$ converge. Thus, we have

$$\beta(t) = 1 + \frac{1}{2}tZ^\top Z - \frac{1}{8}t^2(Z^\top Z)^2 + \cdots + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}t^k(Z^\top Z)^k + \cdots$$

and

$$\gamma(t) = 1 + \frac{1}{2}tZZ^\top - \frac{1}{8}t^2(ZZ^\top)^2 + \cdots + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}t^k(ZZ^\top)^k + \cdots$$

and we get

$$Z\beta(t) = Z + \frac{1}{2}tZZ^\top Z - \frac{1}{8}t^2Z(Z^\top Z)^2 + \cdots + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}t^kZ(Z^\top Z)^k + \cdots$$

and

$$\gamma(t)Z = Z + \frac{1}{2}tZZ^\top Z - \frac{1}{8}t^2(ZZ^\top)^2Z + \cdots + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\cdots\left(\frac{1}{2}-k+1\right)}{k!}t^k(ZZ^\top)^kZ + \cdots.$$

However

$$Z(Z^\top Z)^k = Z\underbrace{Z^\top Z\cdots Z^\top Z}_{k} = \underbrace{ZZ^\top\cdots ZZ^\top}_{k}Z = (ZZ^\top)^kZ,$$

which proves that $Z\beta(t) = \gamma(t)Z$, as required. □

Another proof of Proposition 2.13 can be given using the SVD of Z. Indeed, we can write

$$Z = PDQ^\top$$

where P is a $q \times q$ orthogonal matrix, Q is a $p \times p$ orthogonal matrix and D is a $q \times p$ matrix whose diagonal entries are (strictly) positive and all other entries zero. Then,

$$I - Z^\top Z = I - QD^\top P^\top PDQ^\top = Q(I - D^\top D)Q^\top,$$

a symmetric positive definite matrix. We also have

$$I - ZZ^\top = I - PDQ^\top QD^\top P^\top = P(I - DD^\top)P^\top,$$

another symmetric positive definite matrix. Then,

$$Z(I - Z^\top Z)^{-\frac{1}{2}} = PDQ^\top Q(I - D^\top D)^{-\frac{1}{2}}Q^\top = PD(I - D^\top D)^{-\frac{1}{2}}Q^\top$$

and

$$(I - ZZ^\top)^{-\frac{1}{2}} = P(I - DD^\top)^{-\frac{1}{2}}P^\top PDQ^\top = P(I - DD^\top)^{-\frac{1}{2}}DQ^\top,$$

so it suffices to prove that

$$D(I - D^\top D)^{-\frac{1}{2}} = (I - DD^\top)^{-\frac{1}{2}}D.$$

However, D is essentially a diagonal matrix and the above is easily verified, as the reader should check.

Remark: The polar form can also be obtained *via* the exponential map and the Lie algebra, $\mathfrak{o}(p,q)$, of $\mathbf{O}(p,q)$, see Section 5.6.

We also have the following amusing property of the determinants of A and D:

Proposition 2.14 *For any matrix $X \in \mathbf{O}(p,q)$, if we write*

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

then

$$\det(X) = \det(A)\det(D)^{-1} \quad \textit{and} \quad |\det(A)| = |\det(D)| \geq 1.$$

Proof. Using the identities $A^\top B = C^\top D$ and $D^\top D = I + B^\top B$ proved earlier, observe that

$$\begin{pmatrix} A^\top & 0 \\ B^\top & -D^\top \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} A^\top A & A^\top B \\ B^\top A - D^\top C & B^\top B - D^\top D \end{pmatrix} = \begin{pmatrix} A^\top A & A^\top B \\ 0 & -I_q \end{pmatrix}.$$

If we compute determinants, we get

$$\det(A)(-1)^q \det(D)\det(X) = \det(A)^2(-1)^q.$$

It follows that

$$\det(X) = \det(A)\det(D)^{-1}.$$

From $A^\top A = I + C^\top C$ and $D^\top D = I + B^\top B$, we conclude that $\det(A) \geq 1$ and $\det(D) \geq 1$. Since $|\det(X)| = 1$, we have $|\det(A)| = |\det(D)| \geq 1$. □

Remark: It is easy to see that the equations relating A, B, C, D established in the proof of Proposition 2.12 imply that

$$\det(A) = \pm 1 \quad \text{iff} \quad C = 0 \quad \text{iff} \quad B = 0 \quad \text{iff} \quad \det(D) = \pm 1.$$

2.5 Topological Groups

Since Lie groups are topological groups (and manifolds), it is useful to gather a few basic facts about topological groups.

Definition 2.11 A set, G, is a *topological group* iff

(a) G is a Hausdorff topological space;

(b) G is a group (with identity 1);

(c) Multiplication, $\cdot\colon G \times G \to G$, and the inverse operation, $G \longrightarrow G\colon g \mapsto g^{-1}$, are continuous, where $G \times G$ has the product topology.

It is easy to see that the two requirements of condition (c) are equivalent to

(c′) The map $G \times G \longrightarrow G\colon (g, h) \mapsto gh^{-1}$ is continuous.

Given a topological 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 $b \in G$. Observe that $L_{a^{-1}}$ is the inverse of L_a and similarly, $R_{a^{-1}}$ is the inverse of R_a. As multiplication is continuous, we see that L_a and R_a are continuous. Moreover, since they have a continuous inverse, they are homeomorphisms. As a consequence, if U is an open subset of G, then so is $gU = L_g(U)$ (resp. $Ug = R_gU$), for all $g \in G$. Therefore, the topology of a topological group (i.e., its family of open sets) is *determined* by the knowledge of the open subsets containing the identity, 1.

Given any subset, $S \subseteq G$, let $S^{-1} = \{s^{-1} \mid s \in S\}$; let $S^0 = \{1\}$ and $S^{n+1} = S^nS$, for all $n \geq 0$. Property (c) of Definition 2.11 has the following useful consequences:

Proposition 2.15 *If G is a topological group and U is any open subset containing 1, then there is some open subset, $V \subseteq U$, with $1 \in V$, so that $V = V^{-1}$ and $V^2 \subseteq U$. Furthermore, $\overline{V} \subseteq U$.*

Proof. Since multiplication $G \times G \longrightarrow G$ is continuous and $G \times G$ is given the product topology, there are open subsets, U_1 and U_2, with $1 \in U_1$ and $1 \in U_2$, so that $U_1U_2 \subseteq U$. Ley $W = U_1 \cap U_2$ and $V = W \cap W^{-1}$. Then, V is an open set containing 1 and, clearly, $V = V^{-1}$ and $V^2 \subseteq U_1U_2 \subseteq U$. If $g \in \overline{V}$, then gV is an open set containing g (since $1 \in V$) and thus, $gV \cap V \neq \emptyset$. This means that there are some $h_1, h_2 \in V$ so that $gh_1 = h_2$, but then, $g = h_2h_1^{-1} \in VV^{-1} = VV \subseteq U$. □

A subset, U, containing 1 and such that $U = U^{-1}$, is called *symmetric*. Using Proposition 2.15, we can give a very convenient characterization of the Hausdorff separation property in a topological group.

Proposition 2.16 *If G is a topological group, then the following properties are equivalent:*

(1) G is Hausdorff;

(2) The set $\{1\}$ is closed;

(3) The set $\{g\}$ is closed, for every $g \in G$.

Proof. The implication (1) $\longrightarrow$ (2) is true in any Hausdorff topological space. We just have to prove that $G - \{1\}$ is open, which goes as follows: For any $g \neq 1$, since G is Hausdorff, there exists disjoint open subsets U_g and V_g, with $g \in U_g$ and $1 \in V_g$. Thus, $\bigcup U_g = G - \{1\}$, showing that $G - \{1\}$ is open. Since L_g is a homeomorphism, (2) and (3) are equivalent. Let us prove that (3) $\longrightarrow$ (1). Let $g_1, g_2 \in G$ with $g_1 \neq g_2$. Then, $g_1^{-1}g_2 \neq 1$ and if U and V are distinct open subsets such that $1 \in U$ and $g_1^{-1}g_2 \in V$, then $g_1 \in g_1U$ and $g_2 \in g_1V$, where g_1U and g_1V are still open and disjoint. Thus, it is enough to separate 1 and $g \neq 1$. Pick any $g \neq 1$. If every open subset containing 1 also contained g, then 1 would be in the closure of $\{g\}$, which is absurd, since $\{g\}$ is closed and $g \neq 1$. Therefore, there is some open subset, U, such that $1 \in U$ and $g \notin U$. By Proposition 2.15, we can find an open subset, V, containing 1, so that $VV \subseteq U$ and $V = V^{-1}$. We claim that V and Vg are disjoint open sets with $1 \in V$ and $g \in gV$.

Since $1 \in V$, it is clear that $1 \in V$ and $g \in gV$. If we had $V \cap gV \neq \emptyset$, then we would have $g \in VV^{-1} = VV \subseteq U$, a contradiction. □

If H is a subgroup of G (not necessarily normal), we can form the set of left cosets, G/H and we have the projection, $p\colon G \to G/H$, where $p(g) = gH = \overline{g}$. If G is a topological group, then G/H can be given the *quotient topology*, where a subset $U \subseteq G/H$ is open iff $p^{-1}(U)$ is open in G. With this topology, p is continuous. The trouble is that G/H is not necessarily Hausdorff. However, we can neatly characterize when this happens.

Proposition 2.17 *If G is a topological group and H is a subgroup of G then the following properties hold:*

(1) The map $p\colon G \to G/H$ is an open map, which means that $p(V)$ is open in G/H whenever V is open in G.

(2) The space G/H is Hausdorff iff H is closed in G.

(3) If H is open, then H is closed and G/H has the discrete topology (every subset is open).

(4) The subgroup H is open iff $1 \in \overset{\circ}{H}$ (i.e., there is some open subset, U, so that $1 \in U \subseteq H$).

Proof. (1) Observe that if V is open in G, then $VH = \bigcup_{h \in H} Vh$ is open, since each Vh is open (as right translation is a homeomorphism). However, it is clear that

$$p^{-1}(p(V)) = VH,$$

i.e., $p^{-1}(p(V))$ is open, which, by definition, means that $p(V)$ is open.

(2) If G/H is Hausdorff, then by Proposition 2.16, every point of G/H is closed, i.e., each coset gH is closed, so H is closed. Conversely, assume H is closed. Let $\overline{x}$ and $\overline{y}$ be two distinct point in G/H and let $x, y \in G$ be some elements with $p(x) = \overline{x}$ and $p(y) = \overline{y}$. As $\overline{x} \neq \overline{y}$, the elements x and y are not in the same coset, so $x \notin yH$. As H is closed, so is yH, and since $x \notin yH$, there is some open containing x which is disjoint from yH, and we may assume (by translation) that it is of the form Ux, where U is an open containing 1. By Proposition 2.15, there is some open V containing 1 so that $VV \subseteq U$ and $V = V^{-1}$. Thus, we have

$$V^2 x \cap yH = \emptyset$$

and in fact,

$$V^2 xH \cap yH = \emptyset,$$

since H is a group. Since $V = V^{-1}$, we get

$$VxH \cap VyH = \emptyset,$$

and then, since V is open, both VxH and VyH are disjoint, open, so $p(VxH)$ and $p(VyH)$ are open sets (by (1)) containing $\overline{x}$ and $\overline{y}$ respectively and $p(VxH)$ and $p(VyH)$ are disjoint (because $p^{-1}(p(VxH)) = VxHH = VxH$ and $p^{-1}(p(VyH)) = VyHH = VyH$ and $VxH \cap VyH = \emptyset$).

(3) If H is open, then every coset gH is open, so every point of G/H is open and G/H is discrete. Also, $\bigcup_{g \notin H} gH$ is open, i.e., H is closed.

(4) Say U is an open subset such that $1 \in U \subseteq H$. Then, for every $h \in H$, the set hU is an open subset of H with $h \in hU$, which shows that H is open. The converse is trivial. □

Proposition 2.18 *If G is a connected topological group, then G is generated by any symmetric neighborhood, V, of 1. In fact,*

$$G = \bigcup_{n \geq 1} V^n.$$

Proof. Since $V = V^{-1}$, it is immediately checked that $H = \bigcup_{n \geq 1} V^n$ is the group generated by V. As V is a neighborhood of 1, there is some open subset, $U \subseteq V$, with $1 \in U$, and so $1 \in \overset{\circ}{H}$. From Proposition 2.17, the subgroup H is open and closed and since G is connected, $H = G$. □

A subgroup, H, of a topological group G is *discrete* iff the induced topology on H is discrete, i.e., for every $h \in H$, there is some open subset, U, of G so that $U \cap H = \{h\}$.

Proposition 2.19 *If G is a topological group and H is discrete subgroup of G, then H is closed.*

Proof. As H is discrete, there is an open subset, U, of G so that $U \cap H = \{1\}$, and by Proposition 2.15, we may assume that $U = U^{-1}$. If $g \in \overline{H}$, as gU is an open set containing g, we have $gU \cap H \neq \emptyset$. Consequently, there is some $y \in gU \cap H = gU^{-1} \cap H$, so $g \in yU$ with $y \in H$. Thus, we have

$$g \in yU \cap \overline{H} \subseteq \overline{yU \cap H} = \overline{\{y\}} = \{y\},$$

since $U \cap H = \{1\}$, $y \in H$ and G is Hausdorff. Therefore, $g = y \in H$. □

Proposition 2.20 *If G is a topological group and H is any subgroup of G, then the closure, $\overline{H}$, of H is a subgroup of G.*

Proof. This follows easily from the continuity of multiplication and of the inverse operation, the details are left as an exercise to the reader. □

Proposition 2.21 *Let G be a topological group and H be any subgroup of G. If H and G/H are connected, then G is connected.*

Proof. It is a standard fact of topology that a space G is connected iff every continuous function, f, from G to the discrete space $\{0, 1\}$ is constant. Pick any continuous function, f, from G to $\{0, 1\}$. As H is connected and left translations are homeomorphisms, all cosets, gH, are connected. Thus, f is constant on every coset, gH. Thus, the function $f\colon G \to \{0, 1\}$ induces a continuous function, $\overline{f}\colon G/H \to \{0, 1\}$, such that $f = \overline{f} \circ p$ (where $p\colon G \to G/H$; the continuity of $\overline{f}$ follows immediately from the definition of the quotient topology on G/H). As G/H is connected, $\overline{f}$ is constant and so, $f = \overline{f} \circ p$ is constant. □

Proposition 2.22 *Let G be a topological group and let V be any connected symmetric open subset containing 1. Then, if G_0 is the connected component of the identity, we have*

$$G_0 = \bigcup_{n \geq 1} V^n$$

and G_0 is a normal subgroup of G. Moreover, the group G/G_0 is discrete.

Proof. First, as V is open, every V^n is open, so the group $\bigcup_{n \geq 1} V^n$ is open, and thus closed, by Proposition 2.17 (3). For every $n \geq 1$, we have the continuous map

$$\underbrace{V \times \cdots \times V}_{n} \longrightarrow V^n : (g_1, \ldots, g_n) \mapsto g_1 \cdots g_n.$$

As V is connected, $V \times \cdots \times V$ is connected and so, V^n is connected. Since $1 \in V^n$ for all $n \geq 1$, and every V^n is connected, we conclude that $\bigcup_{n \geq 1} V^n$ is connected. Now, $\bigcup_{n \geq 1} V^n$ is connected, open and closed, so it is the connected component of 1. Finally, for every $g \in G$, the group gG_0g^{-1} is connected and contains 1, so it is contained in G_0, which proves that G_0 is normal. Since G_0 is open, the group G/G_0 is discrete. □

A topological space, X, is *locally compact* iff for every point $p \in X$, there is a compact neighborhood, C of p, i.e., there is a compact, C, and an open, U, with $p \in U \subseteq C$. For example, manifolds are locally compact.

Proposition 2.23 *Let G be a topological group and assume that G is connected and locally compact. Then, G is countable at infinity, which means that G is the union of a countable family of compact subsets. In fact, if V is any symmetric compact neighborhood of 1, then*

$$G = \bigcup_{n \geq 1} V^n.$$

Proof. Since G is locally compact, there is some compact neighborhood, K, of 1. Then, $V = K \cap K^{-1}$ is also compact and a symmetric neigborhood of 1. By Proposition 2.18, we have

$$G = \bigcup_{n \geq 1} V^n.$$

An argument similar to the one used in the proof of Proposition 2.22 to show that V^n is connected if V is connected proves that each V^n compact if V is compact. □

If a topological group, G acts on a topological space, X, and the action $\cdot\colon G \times X \to X$ is continuous, we say that G acts *continuously on* X. Under some mild assumptions on G and X, the quotient space, G/G_x, is homeomorphic to X. For example, this happens if X is a Baire space.

Recall that a *Baire space*, X, is a topological space with the property that if $\{F\}_{i \geq 1}$ is any countable family of closed sets, F_i, such that each F_i has empty interior, then $\bigcup_{i \geq 1} F_i$ also has empty interior. By complementation, this is equivalent to the fact that for every countable family of open sets, U_i, such that each U_i is dense in X (i.e., $\overline{U}_i = X$), then $\bigcap_{i \geq 1} U_i$ is also dense in X.

Remark: A subset, $A \subseteq X$, is *rare* if its closure, $\overline{A}$, has empty interior. A subset, $Y \subseteq X$, is *meager* if it is a countable union of rare sets. Then, it is immediately verified that a space, X, is a Baire space iff every nonempty open subset of X is not meager.

The following theorem shows that there are plenty of Baire spaces:

Theorem 2.24 *(Baire) (1) Every locally compact topological space is a Baire space.*

(2) Every complete metric space is a Baire space.

A proof of Theorem 2.24 can be found in Bourbaki [24], Chapter IX, Section 5, Theorem 1.

We can now greatly improve Proposition 2.2 when G and X are topological spaces having some "nice" properties.

Theorem 2.25 *Let G be a topological group which is locally compact and countable at infinity, X a Hausdorff topological space which is a Baire space and assume that G acts transitively and continuously on X. Then, for any $x \in X$, the map $\varphi\colon G/G_x \to X$ is a homeomorphism.*

By Theorem 2.24, we get the following important corollary:

Theorem 2.26 *Let G be a topological group which is locally compact and countable at infinity, X a Hausdorff locally compact topological space and assume that G acts transitively and continuously on X. Then, for any $x \in X$, the map $\varphi\colon G/G_x \to X$ is a homeomorphism.*

Proof of Theorem 2.25. We follow the proof given in Bourbaki [24], Chapter IX, Section 5, Proposition 6 (Essentially the same proof can be found in Mneimné and Testard [111], Chapter 2). First, observe that if a topological group acts continuously and transitively on a Hausdorff topological space, then for every $x \in X$, the stabilizer, G_x, is a closed subgroup of G. This is because, as the action is continuous, the projection $\pi\colon G \longrightarrow X\colon g \mapsto g \cdot x$ is continuous, and $G_x = \pi^{-1}(\{x\})$, with $\{x\}$ closed. Therefore, by Proposition 2.17, the quotient space, G/G_x, is Hausdorff. As the map $\pi\colon G \longrightarrow X$ is continuous, the induced map $\varphi\colon G/G_x \to X$ is continuous and by Proposition 2.2, it is a bijection. Therefore, to prove that φ is a homeomorphism, it is enough to prove that φ is an open map. For this, it suffices to show that π is an open map. Given any open, U, in G, we will prove that for any $g \in U$, the element $\pi(g) = g \cdot x$ is contained in the interior of $U \cdot x$. However, observe that this is equivalent to proving that x belongs to the interior of $(g^{-1} \cdot U) \cdot x$. Therefore, we are reduced to the case: If U is any open subset of G containing 1, then x belongs to the interior of $U \cdot x$.

Since G is locally compact, using Proposition 2.15, we can find a compact neighborhood of the form $W = \overline{V}$, such that $1 \in W$, $W = W^{-1}$ and $W^2 \subseteq U$, where V is open with $1 \in V \subseteq U$. As G is countable at infinity, $G = \bigcup_{i \geq 1} K_i$, where each K_i is compact. Since V is open, all the cosets gV are open, and as each K_i is covered by the gV's, by compactness of K_i, finitely many cosets gV cover each K_i and so,

$$G = \bigcup_{i \geq 1} g_i V = \bigcup_{i \geq 1} g_i W,$$

for countably many $g_i \in G$, where each $g_i W$ is compact. As our action is transitive, we deduce that

$$X = \bigcup_{i \geq 1} g_i W \cdot x,$$

where each $g_i W \cdot x$ is compact, since our action is continuous and the $g_i W$ are compact. As X is Hausdorff, each $g_i W \cdot x$ is closed and as X is a Baire space expressed as a union of closed sets, one of the $g_i W \cdot x$ must have nonempty interior, i.e., there is some $w \in W$, with $g_i w \cdot x$ in the interior of $g_i W \cdot x$, for some i. But then, as the map $y \mapsto g \cdot y$ is a homeomorphism for any given $g \in G$ (where $y \in X$), we see that x is in the interior of

$$w^{-1} g_i^{-1} \cdot (g_i W \cdot x) = w^{-1} W \cdot x \subseteq W^{-1} W \cdot x = W^2 \cdot x \subseteq U \cdot x,$$

as desired. □

As an application of Theorem 2.26 and Proposition 2.21, we show that the Lorentz group $\mathbf{SO}_0(n,1)$ is connected. Firstly, it is easy to check that $\mathbf{SO}_0(n,1)$ and $\mathcal{H}_n^+(1)$ satisfy the assumptions of Theorem 2.26 because they are both manifolds, although this notion has not been discussed yet (but will be in Chapter 3). Also, we saw at the end of Section 2.3 that the action $\cdot\colon \mathbf{SO}_0(n,1) \times \mathcal{H}_n^+(1) \longrightarrow \mathcal{H}_n^+(1)$ of $\mathbf{SO}_0(n,1)$ on $\mathcal{H}_n^+(1)$ is transitive, so that, as topological spaces

$$\mathbf{SO}_0(n,1)/\mathbf{SO}(n) \cong \mathcal{H}_n^+(1).$$

Now, we already showed that $\mathcal{H}_n^+(1)$ is connected so, by Proposition 2.21, the connectivity of $\mathbf{SO}_0(n,1)$ follows from the connectivity of $\mathbf{SO}(n)$ for $n \geq 1$. The connectivity of $\mathbf{SO}(n)$ is a consequence of the surjectivity of the exponential map (for instance, see Gallier [58], Chapter 14) but we can also give a quick proof using Proposition 2.21. Indeed, $\mathbf{SO}(n+1)$ and S^n are both manifolds and we saw in Section 2.2 that

$$\mathbf{SO}(n+1)/\mathbf{SO}(n) \cong S^n.$$

Now, S^n is connected for $n \geq 1$ and $\mathbf{SO}(1) \cong S^1$ is connected. We finish the proof by induction on n.

Corollary 2.27 *The Lorentz group $\mathbf{SO}_0(n,1)$ is connected; it is the component of the identity in $\mathbf{O}(n,1)$.*

Readers who wish to learn more about topological groups may consult Sagle and Walde [129] and Chevalley [34] for an introductory account, and Bourbaki [23], Weil [149] and Pontryagin [122, 123], for a more comprehensive account (especially the last two references).