Some notes on Lie groups

Richard Hartley

Australian National University

Je n'ai fait celle-ci plus longue que parce ce que je n'ai pas eu le loisir de la faire plus courte.

– Blaise Pascal, Lettres provinciales.

Contents

1 Definitions

A Lie group is defined as a group G, which is at the same time a manifold, such that the group operations are smooth operations in the manifold topology. A Lie group is a specialization of the concept of a *topological group* which is at the same time a topological space and a group, for which the group operations are smooth.

Many of the most important examples of Lie groups (at least finite-dimensional ones) may be represented as subgroups of the matrix group $GL(n, \mathbb{C})$ or $GL(n, \mathbb{R})$. In these notes, we shall be mainly concerned with these so-called matrix Lie groups (to be defined below). The advantage of this approach is that it is more concrete, and the intuition is clearer than in the abstract setting. Often, things are easier to prove in the setting of matrix Lie groups. As a meta-mathematical statement, one can say that most statements that one can make about matrix Lie groups, for which the statement does not specifically make use of matrix concepts, are probably true in the case of abstract Lie groups in general.

Matrix Lie groups. The general linear group $GL(n, \mathbb{R})$ is the group of all non-singular $n \times n$ matrices. It forms a group under the usual matrix multiplication. It has the topology given by a standard norm. $d(A, B) = \|A - B\|_\infty = \max_{ij} |A_{ij} - B_{ij}|$. Alternatively, (and equivalently) one may use the Frobenius or L_2 distance

$$\begin{aligned} d(A, B)^2 &= \|A - B\|_F & (1) \\ &= \sum_{ij} |A_{ij} - B_{ij}|^2 & (2) \\ &= \mathrm{Tr}(A^\top B) & (3) \end{aligned}$$

Definition 1.1. A matrix Lie group is a closed subgroup of $GL(n, \mathbb{R})$.

Clarification The condition that $G < GL(n, \mathbb{R})$ is closed subgroup means that if (A_i) is a sequence of matrices in G, converging to a matrix $A_\infty \in GL(n, \mathbb{R})$, then A_∞ is in G.

Example 1.2. Look at the matrices of the form

$$A_\alpha = \begin{bmatrix} 1 & 0 \\ 0 & \alpha \end{bmatrix}$$

for $\alpha > 0$. These form a group G. The matrix $\mathrm{diag}(1, 0)$ is a limit of elements in this group. However, this matrix is not an element of $GL(2, \mathbb{R})$. The group is a Lie group. Note, that G is closed as a subset of $GL(2, \mathbb{R})$, but **not** as a subset of $M(2, \mathbb{R})$, the set of 2×2 real matrices.

In general, $GL(n, \mathbb{R})$ itself has limits that are not in $GL(n, R)$, so $GL(n, \mathbb{R})$ is not closed as a subset of $M(n, \mathbb{R})$. However, this is not relevant, since it $GL(n, \mathbb{R})$ is closed as a subset of itself, which is all that is required.

Example 1.3. Look at the matrices

$$A_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} .$$

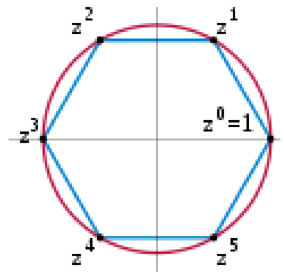

Figure 1: *Dihedral group.*

These form a Lie group, homeomorphic (topologically) to the circle S^1. It is called $SL(2, \mathbb{R})$. Now consider the matrices

$$\begin{bmatrix} A_\theta & 0 \\ 0 & A_{k\theta} \ , \end{bmatrix} \tag{4}$$

where k is a non-zero real number. The ensemble of all such matrices (for varing k an θ) forms a group G, isomorphic to $SL(2, \mathbb{R}) \times SL(2, \mathbb{R})$.

However, if k is a *fixed* irrational number, the set of matrices so obtained (although a group, call it G_k is not closed as a subgroup of $GL(4, \mathbb{R})$. Hence, G is not a Lie group.

Observe that G_k forms a subgroup of $S^1 \times S^1$, the torus, and is related to the irrational flow on a torus. (In fact, it forms one orbit of the irrational flow). Denote $S^1 \times S^1$ to mean the group of matrices of the form

$$\begin{bmatrix} A_{\alpha\theta} & 0 \\ 0 & A_{\beta\theta} \ , \end{bmatrix}$$

for α and β real numbers. Then G_k is strictly smaller than $S^1 \times S^1$, and is not a closed subset of $S^1 \times S^1$. In fact $S^1 \times S^1$ is the closure of G, meaning that any element in $S^1 \times S^1$ is the limit of some sequence of elements of G.

With the topology induced from being a (dense) subset of $S^1 \times S^1$, the group G_k, with k irrational is a topological group, isomorphic and homeomorphic to $\mathbb{R}$, where $\mathbb{R}$ has a somewhat curious topology – not the usual one.

Note, however, that for k rational, this forms a closed subgroup of $GL(4, \mathbb{R})$, and hence a Lie group.

Various groups

1. Discrete Lie groups (for instance, the groups of reflections, dihedral groups (D_p) (fig 1), groups of symmetries (fig 3). These groups have dimension 0, and are not very interesting for us in these lectures.

2. Invertible matrices $GL(n, \mathbb{R})$ or $GL(n, \mathbb{C})$.

3. The group $GL^+(n, \mathbb{R})$ consisting of those elements of $GL(n, \mathbb{R})$ with positive determinant.

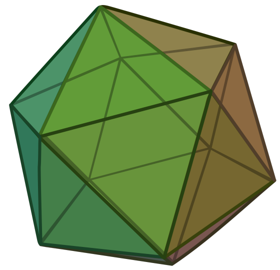

Figure 2: *The icosahedron is the group of symmetries of an icosahedron.*

4. In the case $n = 1$, these are the same as the groups $\mathbb{R}^*$ and $\mathbb{C}^*$, the non-zero elements of $\mathbb{R}$ and $\mathbb{C}$ respectively, which are groups under multiplication. The group $\mathbb{R}^*$ consists of both positive and negative elements. The positive elements of $\mathbb{R}$ themselves form a group, denoted by $\mathbb{R}^+$, which is once again a subgroup of $GL(1,\mathbb{R})$. (Note that one cannot do a similar type of thing for $\mathbb{C}^*$.)

5. Matrices of determinant 1, namely $SL(n,\mathbb{R})$ and $SL(n,\mathbb{C})$.

6. Orthogonal groups $O(n,\mathbb{R})$ and $O(n,\mathbb{C})$, square matrices satisfying the condition $A^\top A = I$.

7. Unitary groups, $U(n)$ (or $U(n,\mathbb{C})$, the same thing) satisfying the condition $A^*A = I$, where A^* represents the conjugate transpose.

8. Projective linear groups, $PGL(n,\mathbb{R})$, and $PGL(n,\mathbb{C})$, which is the quotient group of $GL(n,\mathbb{R})$ modulo multiplication by a constant. Thus $PGL(n,\mathbb{R}) \equiv GL(n,R)/Z(GL(n,\mathbb{R})$, where Z represents the *centre*, namely the set of *scalar matrices*.

9. Projective special linear group, $PSL(n,\mathbb{R})$ and $PSL(n,\mathbb{C})$, which are the quotients of $SL(n,\mathbb{R})$ and $SL(n,\mathbb{C})$.

Exercise 1.4. $PSL(n,\mathbb{R})$ is the same as $PGL(n,\mathbb{R})$, for n odd, since -1 has an n-th root. In the case where n is even, $PGL(n,\mathbb{R})$ is twice as big as $PSL(n,\mathbb{R})$. Matrices with determinant $+1$ and -1 are different in $PGL(n,\mathbb{R})$, when n is even, but the same when n is odd (since multiplying by -1 changes the determinant. Thus, for n odd, $PGL(n,\mathbb{R})$ has two disconnected components, those with negative and those with positive determinant. The subgroup of $PGL(n,\mathbb{R})$, with n odd, with determinant positive, is isomorphic to $PSL(n,\mathbb{R})$. This is known as the identity component of $PGL(n,\mathbb{R})$.

However, since -1 has an n-th root in $\mathbb{C}$, the groups $PSL(n,\mathbb{C})$ and $PGL(n,\mathbb{C})$ are isomorphic. (Multiplication of a matrix by an n-th root of -1 changes the sign of the determinant).

Additive Lie groups. Under addition, both $\mathbb{R}$ and $\mathbb{C}$ are groups. They are clearly groups, and they both have a topology (the usual topology) compatible with the group operation. They are examples of *topological groups*.

However, they are not exactly subgroups of any $GL(n,\mathbb{R})$ (most particularly $GL(1,\mathbb{R})$), since the group operation (addition) in $\mathbb{R}$ does not correspond with the group operation (multiplication) in $GL(1,\mathbb{R})$.

However, there exists a mapping $\phi : (\mathbb{R}, +) \to (\mathbb{R}^+, \times)$, defined by $\phi(x) = e^x$. It is easy to see that ϕ is a group isomorphism, and moreover it is a diffeomorphism with respect to the usual topology of $\mathbb{R}$ and $\mathbb{R}^+$. Identifying $\mathbb{R}^+$ as a subgroup of $GL(1, \mathbb{R})$, one may verify that the image of ϕ is a Lie group (one must verify that it is a closed subgroup of $GL(1, \mathbb{R})$, which is trivial). Such a mapping ϕ a *faithful representation* of the topological group $(R, +)$ as a Lie subgroup of $GL(1, \mathbb{R})$.[1]

Via a faithful representation, a topological group may be identified with a matrix Lie group. and all that will be said about matrix Lie groups applies equally well to those topological groups with a faithful representation as a matrix Lie group.

The above discussion can be extended to $\mathbb{R}^m$ and $\mathbb{C}^m$, as groups under addition. They have faithful representations as Lie subgroups of $GL(m, \mathbb{R})$ and $GL(m,\mathbb{C})$ respectively (as groups of diagonal matrices) via the exponential mapping.

Thus, for $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ in $\mathbb{R}^m$ or $\mathbb{C}^m$) there is a faithful representation:

$$\mathbf{x} \mapsto \begin{bmatrix} e^{x_1} & & \\ & \ddots & \\ & & e^{x_m} \end{bmatrix} .$$

Although this is a possible faithful representation of $\mathbb{R}^m$ as a subgroup of $GL(m, \mathbb{R})$, there is a different representation, sometimes more useful, where $\mathbb{R}^m$ is represented as a subgroup of $GL(m + 1, \mathbb{R})$, as

$$\mathbf{x} \mapsto \begin{bmatrix} 1 & & & x_1 \\ & \ddots & & \vdots \\ & & 1 & x_m \\ & & & 1 \end{bmatrix} \tag{5}$$

One may verify that this is a faithful representation of $(\mathbb{R}^m, +)$ as a matrix Lie group.

1.1 Groups of geometric transformations

In addition, there are several Lie groups consisting of different transformations of a Euclidean space (that is, one-to-one continuous transformations of $\mathbb{R}^m$). In applications, this is commonly the $3D$ or $2D$ Euclidean space $\mathbb{R}^2$ or $\mathbb{R}^n$. In the cases described, the sets of transformations clearly form a group, but the topology will be defined by identifying them with groups of matrices.

First, the groups will be listed below. Later, it will be shown that they may be represented as matrix Lie groups.

1. The group of linear transformations, equal to (or represented by) $GL(n, \mathbb{R})$.
2. The group of rigid linear transformations, represented by to $O(n, \mathbb{R})$. Note that the group $O(n, \mathbb{R})$ includes reflections, so by rigid linear transformations, in this context, we include reflections. To exclude reflections, one uses the group $SO(n, \mathbb{R})$ instead (as mentioned again below).

[1]A subgroup $H < G$ is a Lie subgroup if it is a closed in the topology of G.

3. The group of scaled rigid linear transformations, This may be represented by the set of matrices kA where $A \in O(n, \mathbb{R})$ and $k \neq 0$. This is clearly a matrix Lie group. An alternative representation of the group of scaled rigid linear transformations is as the group of matrices of the form $\left[\begin{smallmatrix} A & 0 \\ 0 & 1/k \end{smallmatrix}\right]$, where $A \in O(m, \mathbb{R})$ and $k \neq 0$. This shows the group of scaled rigid linear transformations to be isomorphic to $O(m, \mathbb{R}) \times \mathbb{R}^*$, and is a faithful representation as a subgroup of $GL(m+1, \mathbb{R})$.

4. The group of translations of $\mathbb{R}^m$. This may be identified with the group $(\mathbb{R}^m, +)$ itself, so a faithful representation of this group is given by (5).

5. The group of rigid transformations (or scaled rigid transformations) of $\mathbb{R}^m$. A rigid transformation consists of a rigid linear transformation plus a translation. The standard way to represent a rigid transformation as a group is to represent an element $\mathbf{x} = (x_1, \ldots, x_m)^\top \in \mathbb{R}^m$ in homogeneous coordinates.[2] An element $\mathbf{x} \in \mathbb{R}^m$ is represented by the vector $\tilde(\mathbf{x}) \in \mathbb{R}^{m+1}$ equal to $(x_1, \ldots, x_m, 1)^\top$. Then a rigid transformation is represented by a matrix

$$\widetilde{A} = \begin{bmatrix} A & \mathbf{t} \\ 0^\top & 1 \end{bmatrix} \tag{6}$$

where A is an orthogonal matrix and $\mathbf{t} \in \mathbb{R}^m$. Then, the product $\tilde{\mathbf{y}} = \widetilde{A}\tilde{\mathbf{x}}$ produces a further homogeneous vector $\tilde{\mathbf{y}} = (y_1, \ldots, y_m, 1)^\top$ which is the homogeneous coordinate vector such that

$$\mathbf{y} = (y_1, \ldots, y_m)^\top = A\mathbf{x} + \mathbf{t} .$$

Thus the association of the rigid transformation with the matrix $\widetilde{A}$ is a faithful Lie group representation of the group of rigid transformations.

6. Scaled rigid transformations. In a similar way, a scaled rigid transformation can be represented by the matrix of the form (6) where the matrix A is of the form kU, where U is orthogonal and $k \neq 0$.

7. Affine transformations. An affine transformation is a combination of a linear transformation followed by a translation. An affine transformation is represented by a matrix of the form (6) where A is an arbitrary element of $GL(m, \mathbb{R})$.

8. All these transformations can be defined also in their "orientation preserving" forms, in which $O(n, \mathbb{R})$ or $GL(n, \mathbb{R})$ are replaced by $SO(n, \mathbb{R})$ and $GL^+(n, \mathbb{R})$. The group of orientation preserving rigid transformations of $\mathbb{R}^m$ is called $SE(m)$. The group $SE(3)$ is of particular importance as representing the group of all possible "motions" of a rigid object. Similarly, the group $SO(3)$ is the group of all possible rotations of an object.

1.2 The group of homographies

A homography, or projective transform is a mapping between geometric projective spaces, represented by an invertible linear mapping on homogeneous coordinates. The projective

[2] By our convention, vectors represented by (for instance) $\mathbf{x}$ are *column vectors*. When the elements are listed, such as $(x_1, x_2, \ldots, x_m)$, this is a row vector, since the elements are listed horizontally. Consequently, to obtain a column vector, we transpose; so $(x_1, \ldots, x_m)^\top$ is the column vector equal to $\mathbf{x}$.

space $\mathcal{P}^n$ consists of equivalence classes of non-zero vectors, modulo multiplication by a non-zero contant. Thus, $\mathbf{v} \equiv k\mathbf{v}$ for all k. A homography is a mapping that takes $\mathbf{v}$ to $H\mathbf{v}$ (more exactly, it takes the equivalence class represented by $\mathbf{v}$ to the equivalence class represented by $H\mathbf{v}$. Since multiplication by a constant does not matter, the matrix H and kH represent the same homography. Since only non-singular matrices are allowed, the group of homographies $\mathcal{P}^n \rightarrow \mathcal{P}^n$ forms a group. The group of homographies of this dimension, which we shall denote by $H(n, \mathbb{R})$ is therefore, the same thing as the group $PGL(n+1, \mathbb{R})$. (Sometimes, $PGL(n, \mathbb{R})$ is defined to mean equivalence classes of $(n+1) \times (n+1)$ matrices (unlike here, where we mean $n \times n$ matrices, in line with the Wikipedia page). In this case, what we have called $H(n, \mathbb{R})$ is the same thing as $PGL(n, \mathbb{R})$, equivalence classes of $n+1$ dimensional matrices.

Just to avoid any ambiguity, we use the symbol $H(n, \mathbb{R})$ to represent the group of homographies, equal to $PGL(n+1, \mathbb{R})$.

Homographies as a matrix Lie group. Since the group of homographies is defined as a quotient group of the matrix Lie group $GL(n, \mathbb{R})$, it is not immediately obvious that it is a matrix Lie group. What is in question here is whether there is a matrix Lie group isomorphic to $GL(n, \mathbb{R})$. In other words, is there a homography from $GL(n, \mathbb{R})$ to some matrix Lie group, having kernel $Z(GL(n, \mathbb{R}))$, the subgroup of scalar matrices.

In the case where n is odd, $PGL(n, \mathbb{R})$ is isomorphic to $SO(n, \mathbb{R})$. The mapping $\tau : GL(n, \mathbb{R}) \rightarrow SO(n, \mathbb{R})$, defined by $A \mapsto A/\det(A)^{1/n}$, takes A to a matrix with determinant 1, and hence an element of $SO(n)$. Here $\det(A)$ has an n-th root $\det(A)^{1/n}$, because n is odd. The kernel of this mapping is clearly the set of scalar matrices kI. Hence, when n is odd, $PGL(n, \mathbb{R})$ is a matrix Lie group. In particular, the group of homographies of $\mathcal{P}^2$, i.e. planar homographies, is a matrix Lie group.

In the case where n is even, the situation is not so clear. One argues as follows. Let $M(n, \mathbb{R})$ be the set of $n \times n$ matrices. An element A of $GL(n, \mathbb{R})$ acts on an element $X \in M(n, \mathbb{R})$ according to $\phi_A : X \mapsto A^{-1}XA$. One sees that $M(n, \mathbb{R})$ is a vector space of dimension n^2, and that ϕ_A is a linear mapping of $M(n, \mathbb{R})$. The association of A with the mapping ϕ_A defines a mapping (denoted by ϕ) from $GL(n, \mathbb{R})$ to the group $GL(n^2, \mathbb{R})$, which one may verify is a homomorphism. In other words, $\phi_{AB} = \phi_A \phi_B$, which one checks by observing that

$$\begin{aligned}
\phi_{AB}(X) &= (AB)X(AB)^{-1} \\
&= A(BXB^{-1})A^{-1} \\
&= \phi_A \circ \phi_B(X)
\end{aligned}$$

The kernel of ϕ is equal to the set of scalar matrices. So the image of ϕ can be seen as a subgroup of $GL(n^2, \mathbb{R})$ isomorphic to $PGL(n, \mathbb{R})$. This is not in itself sufficient to show that $PGL(n, \mathbb{R})$ is a matrix Lie group, since it is necessary to show that the image of ϕ is a *closed* subgroup of $GL(n^2, \mathbb{R})$, hence a matrix Lie group. The proof of this is omitted.

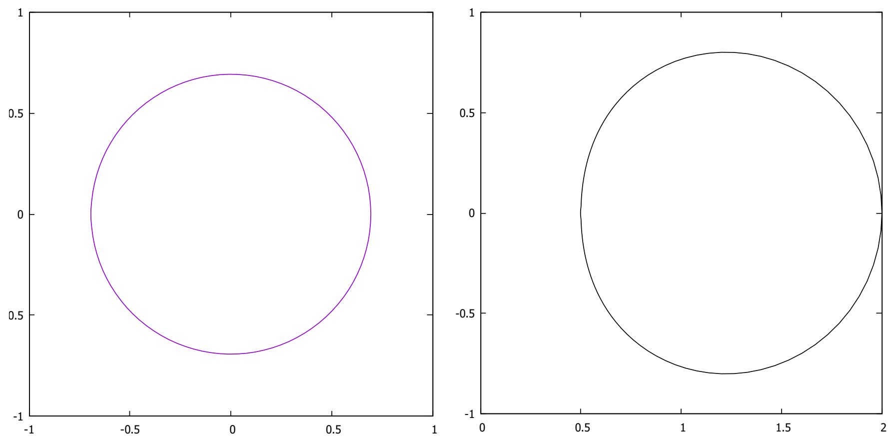

Figure 3: *The exponential and logarithm maps are inverse mappings on suffiently small domains, exponential (left) and logarithm (right).*

2 Exponential map

Given a matrix $A \in M(n, \mathbb{C})$, the exponential is defined as

$$\exp(A) = I + A + A^2/2! + \ldots + A^n/n! + \ldots \tag{7}$$

This series converges for all A.

It is easily seen that if matrices A and B commute, then

$$\exp(A + B) = \exp(A)\exp(B) \ .$$

A proof would follow the proof that the same relationship holds for the usual exponential defined on the real or complex field.

An inverse of this operation is the logarithm map.

$$\log(A) = \sum_{m=1}^{\infty} (-1)^{m+1} \frac{(A - I)^m}{m} \ . \tag{8}$$

This does not converge for all matrices A. However, the following is true.

Theorem 2.5. *If* $\|A - I\| < 1$*, then the sequence converges, and furthermore,* $\exp(\log(A)) = A$*. Furthermore, if* $\|A\| < \log(2)$*, then* $\|\exp(A) - I\| < 1$ *and* $\log(\exp(A)) = A$*.*

Note. The exponential map satisfies the identity

$$\det(\exp(A)) = e^{\mathrm{Tr}(A)} \ . \tag{9}$$

Since the right-hand side is always positive, the exponential map: $\exp : M(n, \mathbb{R}) \to GL(n, \mathbb{R})$ is not surjective. However, $\exp : M(n, \mathbb{C}) \to GL(n, \mathbb{C})$ is surjective.

2.1 Calculating the exponential map

Since the exponential map is defined as an infinite sum, it is not obvious how to compute it directly.

If A is symmetric, then its exponential can be computed by using the eigenvalue decomposition. Observe first that if D is a diagonal matrix $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$, then $\exp(D) = \mathrm{diag}(e^{d_1}, e^{d_2}, \ldots, e^{d_n})$, so computing the exponential is easy enough. An arbitrary real symmetric matrix can be written as $A = UDU^\top$, where U is orthogonal. Similarly, a complex Hermitian matrix can be written as $A = UDU^*$. In the real case, it follows that i

$$\exp(A) = U \exp(D) U^\top$$

and a similar thing happens in the complex Hermitian case. Thus, the exponential of symmetric or Hermitian matrices are easy to compute.

Similarly, if A is a matrix with distinct eigenvalues, then it can be diagonalized as $A = PDP^{-1}$ where P is a non-singular matrix. The exponential of this matrix is $\exp(A) = P \exp(D) P^{-1}$.

In the case where A is not symmetric, or Hermitian, it cannot usually be diagonalized. However, one can use the Jordan normal form in this case. Every (square) matrix can be written as

$$A = P\,\mathrm{diag}(B_1, \ldots, B_n)\,P^{-1}$$

where each B_i is a square, and P is some nonsingular matrix. In this case, it is easily seen that

$$\exp(A) = P\,\mathrm{diag}(\exp(B_1), \ldots, \exp(B_n)) P^{-1}\ .$$

In the Jordan normal form (see Wikipedia "Jordan normal form"), each matrix B_i in this decomposition is of the form

$$B_i = \begin{bmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{bmatrix}\ .$$

Note that the diagonal entries are all equal (they are the eigenvalues of A), and there are entries, equal to 1 on the super-diagonal. All other entries are zero. (If some eigenvalue appears once only, then the corresponding block is 1×1 and there is no superdiagonal.)

The matrix B_i can be written as $B_i = \lambda_i I + E$, where E is the matrix with entries 1 on the superdiagonal. Note that $\lambda_i I$ and E commute, so that

$$\begin{aligned} \exp(B_i) &= \exp(\lambda_i I + E) \\ &= \exp(\lambda_i I)\ \exp(E) \\ &= e^{\lambda_i} \exp(E) \end{aligned}$$

It is easily seen that if E has dimension $m \times m$, then $E^m = 0$, so that the infinite sum expansion of $\exp(E)$ terminates after m terms. Consequently, $\exp(E)$ is easy to compute, and can be written down in closed form. **Exercise:** work out what $\exp(E)$ looks like exactly.

This gives a method to compute the exponential of an arbitrary matrix. (Note, however, the warning on the Wikipedia page about numerical stability of the Jordan normal form. Note also that if the matrix has distinct eigenvalues, then it is diagonalizable, and that a given matrix can be approximated arbitrarily closely by matrices with distinct eigenvalues.)

For a more about ways to compute the matrix exponential, see the Wikipedia pages on "Matrix exponential" and "Jordan-Chevalley decomposition".

3 The Lie Algebra

Definition 3.6. The Lie Algebra $\mathfrak{g}$ of a matrix Lie group G, a subgroup of $GL(n, \mathbb{R})$, is the set of all matrices A such that $\exp(tA) \in G$ for all $t \in \mathbb{R}$.

In this case, $f_A(t) = \exp(tA)$ forms a curve in G, defined for all time t. The curve must lie in G for all t, not just for some t. Note however, that it is sufficient to insist that $\exp(tA)$ lies in G for all sufficiently small t.

Exercise 3.7. If $\exp(tA)$ lies in group G for all $0 \leq t < \epsilon$, then $\exp(tA) \in G$ for all t.

This is a curious definition of the Lie Algebra, so we shall look at it a bit more. The first (and perhaps most important) remark is that the set of all such matrices A forms a real vector space.

Theorem 3.8. *If G is a matrix Lie group in $GL(n,\mathbb{C})$ and $\mathfrak{g}$ is its Lie algebra, then $\mathfrak{g}$ has the following properties, for all, $k \in \mathbb{R}$, $X, Y \in \mathfrak{g}$ and $A \in G$.*

1. $A^{-1}XA^{-1} \in \mathfrak{g}$.
2. $kX \in \mathfrak{g}$.
3. $X + Y \in \mathfrak{g}$
4. $XY - YX \in \mathfrak{g}$.

The first statement is very easy, and the proof that $kX \in \mathfrak{g}$ is trivial. **Note** however, this relationship holds for $k \in R$, not $k \in \mathbb{C}$, even though G is a subgroup of $GL(n,\mathbb{C})$).

To prove that $X+Y$ is in $\mathfrak{g}$ is a little more tricky, since one does **not** have the relationship $\exp(t(X + Y)) = \exp(tX) \exp(tY)$. The result relies on the fact that this relationship is approximately true for small value of t, in particular one has the relationship

$$\lim_{m \to 0} (\exp(mX) \exp(mY))^{1/m} = \exp(X + Y) \ .$$

from which the result follows. The final condition is also relatively simple, and is shown as follows.

As a function of t, the trajectory $\exp(tX)Y \exp(-tX)$ is a smooth curve lying in $\mathfrak{g}$. Its derivative therefore lies in $\mathfrak{g}$, since $\mathfrak{g}$ is a linear subspace of $GL(n, F)$. Calculation shows that the derivative, evaluated at $t = 0$ is exactly $XY - YX$.

The second and third conditions show that $\mathfrak{g}$ is a **real** vector space. The final condition defines the so-called bracket operator on the Lie group, giving it the structure of a real Lie algebra, in the algebraic sense.

An (algebraic) Lie algebra is a vector space with a binary operator (the bracket operator) $[\cdot,\cdot]$, which is bilinear, antisymmetric, and satisfies the Jacobi identity

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

Under some circumstances, $\mathfrak{g}$ will also be a complex vector space, meaning that iX is in $\mathfrak{g}$ whenever X is. In this case G is called a *complex Lie group*. If G is a complex Lie group, so that $\mathfrak{g}$ is a complex vector space, hence isomorphic to $\mathbb{C}^n$, then the exponential map provides charts making G into a *complex* manifold.

Exercise 3.9. Investigate when the examples of Lie groups given above are complex Lie groups. In particular, are the groups $O(n,\mathbb{C})$ (complex orthogonal group) and $U(n)$ (the unitary group) complex Lie groups?

3.1 The Lie algebra and the tangent space

The mapping exp takes a sufficiently small neighbourhood of the zero-matrix 0 in $M(n,\mathbb{C})$ and maps it homeomorphically onto a neighbourhood of the identity I. Restricted to the Lie algebra of some Lie group G, the exponential maps a small neighbourhood of $0\in\mathfrak{g}$ to a smooth embedded neighbourhood of the identity in G. More is true.

Theorem 3.10. *Given a matrix Lie group G with Lie algebra $\mathfrak{g}$, there exists a (sufficiently) small neighbourhood U of $0\in\mathfrak{g}$, which is mapped smoothly and bijectively onto its image $V=\exp(U)$, and such that $U\cap\mathfrak{g}$ maps smoothly and bijectively onto $G\cap V$.*

The important point here is that the exponential map takes a neighbourhood of 0 **onto** a neighbourhood of $I\in G$.

This mapping makes G into a smooth (real) manifold. The exponential map itself is a chart for a neighbourhood of the identity in G. Moreover, for any other point $g\in G$, the mapping $X\mapsto g\exp(X)$ maps a small neighbourhood $U\subset\mathfrak{g}$ epimorphically to a neigbourhood of $g\in G$. Together, these maps form a complete set of charts for G.

Tangent space at the identity. The tangent space to the smooth manifold G[3] at a point $g\in G$ may be viewed as the set of derivatives of all smooth curves passing through g.

Consider the curve $f(t)=\exp(tX)$ for $X\in\mathfrak{g}$. The derivative of this curve, is computed to be equal to $X\exp(tX)=\exp(tX)X$. When $t=0$, this is equal to X. Thus, X is in the tangent space at the identity. Conversely, if $A(t)$ is a smooth curve in G with $A(0)=I$, let $a(t)=\log(A(t))$. In a small neighbourhood, the logarithm is defined and smooth, so the curve $a(t)$ is smooth. Furthermore $A(t)=\exp(a(t))$. Taking derivatives gives $A'(t)=a'(t)\exp(a(t))$, and when $t=0$, this gives $A'(t)=a'(t)$. However, $a(t)$ is a

[3]Since we are chiefly concerned with Lie groups, which have been shown to be (or are by definition) manifolds, we use the symbol G to represent a manifold, or a Lie group.

curve lying in $\mathfrak{g}$, which is a linear vector space. Hence $a'(t)$ is in $\mathfrak{g}$, and so is $A'(t)$. Thus, the tangent space at the identity is a subset of $\mathfrak{g}$.

This proves that $\mathfrak{g}$ is equal to the tangent space $T_I(G)$ of the Lie group at the identity.

Tangent space at other points. At a point $A \in G$, the tangent space is the set of all tangents $f'(0)$, where $f(t)$ is a smooth curve in G with $f(0) = A$. Given a curve $f_0(t)$ with $f_0(t) = I$, the curve $f(t) = A\, f_0(t)$ is a smooth curve with $f(0) = A$, and every smooth curve is obtained in this way. Taking derivatives gives $f'(0) = A\, f_0'(0)$, which shows that the tangent space $T_A(G)$ is equal to $A\mathfrak{g}$.

Exercise 3.11. What (if anything) is special about left multiplication here? The tangent space at A is equal to $A\mathfrak{g}$, which is the set of all products AB with $B \in \mathfrak{g}$. Why not $\mathfrak{g}A$? Show that for $A \in G$, the sets $A\mathfrak{g}$ and $\mathfrak{g}A$ are the same.

For reference, we state the result as a theorem.

Theorem 3.12. *Given a matrix Lie group G with Lie algebra $\mathfrak{g}$, the tangent space to G at a point A is equal to the set $T_A(G) = A\mathfrak{g}$.*

Exercise 3.13. The set $T_A(G) = A\mathfrak{g}$ is a real vector subspace of $GL(n,\mathbb{C})$.

Finding the Lie algebra of a matrix Lie group. A useful method to find the Lie algebra is to find tangents of curves in $GL(n, \mathbb{R})$ at the identity. Consider a point $I + \Delta$, where Δ is an incremental adjustment to I. The question is to find those Δ such that $I + \Delta$ lies in G. The way to do this is to find a defining condition for G, and apply it to $I + \Delta$, treating Δ as in infinitessimal adjustment, and ignoring second-order terms.

Thus, for example, consider the group $O(3, \mathbb{R})$ of all orthogonal 3×3 matrices. The defining condition for an element in $O(3, \mathbb{R})$ is that $A^\top A = I$. Applying this to a matrix $A = I + \Delta$ gives

$$\begin{aligned} A^\top A &= I + \Delta + \Delta^\top + \Delta^\top \Delta \\ &= I + \Delta + \Delta^\top \ . \end{aligned}$$

where $\Delta^\top \Delta$ is deleted, since it consists of elements of second order in the entries of Δ.

The condition that $A + A^\top = I$ leads to $\Delta + \Delta^\top = 0$. Hence, Δ is skew-symmetric. This shows that the Lie algebra of $O(n, \mathbb{R})$ is the vector space of skew-symmetrix matrices of dimension n.

As a second example, consider the group $SL(n, \mathbb{R})$, which has the defining condition that $\det(A) = 1$. Consider an element $I + \Delta$, and look at $\det(I + \Delta)$. It is easily seen that to first order, this is equal to $1 + \mathrm{Tr}(\Delta)$. Hence, the Lie algebra of $SL(n, \mathbb{R})$ is the subspace of $M(n,\, \mathbb{R})$ consisting of all matrices with trace 0.

The exponential map does not map onto G. Consider a matrix Lie group G with Lie algebra $\mathfrak{g}$ (the tangent space at the identity). Since a curve $\exp(tA)$ is smooth and continuous, the set of points $\exp(A)$ with A in $\mathfrak{g}$ must be a path-connected, and hence connected set. Consequently, the exponential map cannot map $\mathfrak{g}$ onto the group G

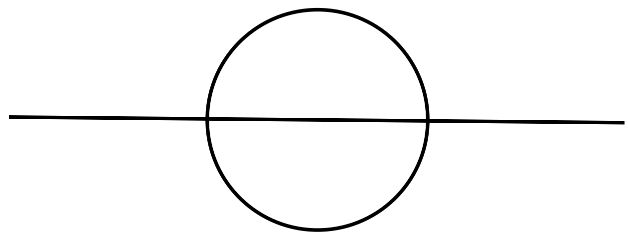

Figure 4: *Eigenvalues of an element in* $SL(2, \mathbb{R})$*. The eigenvalues come in inverse pairs. An element of* $SL(2, \mathbb{R})$ *is in the image of the exponential map if and only if both eigenvalues satisfy* $\mathrm{Re}(\lambda) \geq -1$*.*

unless G is connected. The group $GL(n, \mathbb{R})$ is not connected since the determinant map takes it onto the set $\mathbb{R}\backslash\{0\}$, which is not connected. There is no path from the identity to a matrix with negative determinant. Hence, the exponential map is not onto.

Consider the set of elements $g \in G$ that are in the same connected component of G as the identity. This is a subgroup of G, called the identity component.

Exercise 3.14. The identity component of a Lie group G is a normal subgroup of G.

One may surmise that the exponential map for a matrix Lie group maps the Lie algebra **onto** the identity component of G. However, this is not the case, as the following example shows.

Example 3.15. Let $G = SL(2, \mathbb{R})$. This is a connected Lie group. The Lie algebra $\mathfrak{g}$ is the set of 2×2 matrices with trace 0, namely the set

$$\mathfrak{g} = \left\{ \begin{bmatrix} a & b \\ c & -a \end{bmatrix} \right\} .$$

Now, if $X \in \mathfrak{g}$ is a matrix of this form, then $X^2 = -\det(X)I$. Setting $D = \det(X)$, the exponential expands as

$$\begin{aligned} \exp(X) &= \cos(\sqrt{D})I + \sin(\sqrt{D})/\sqrt{D}X \quad \text{if } D > 0 \\ &= \cosh(\sqrt{-D})I + \sinh(\sqrt{-D})/\sqrt{-D}X \quad \text{if } D < 0 \end{aligned}$$

Since trace of X is zero, this shows that $\mathrm{Tr}(\exp(X)) \geq -2$. However, there are certainly matrices in $SL(2, \mathbb{R})$ with trace less than -2, for instance, the matrix $X = \mathrm{diag}(-2, -1/2)$. This is illustrated in fig 4.

4 More about the Lie bracket

The Lie algebra was defined in the previous section, and was seen to be identifiable with the tangent space of the Lie group at the identity. An important property of the Lie algebra is the existence of the bracket operator $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$, which is a bilinear, antisymmetric product satisfying the Jacobi identity.

However, the Lie bracket can be defined in the context of arbitrary differential manifolds, as is well know. In the case of differential manifolds, however, the Lie bracket is defined on vector fields, rather than (in the case of Lie groups) a single tangent space $\mathfrak{g}$, the tangent space at the identity.

A (smooth) vector field on a differential manifold M is a smooth assignment of a vector $\mathbf{v} \in T_{\mathbf{m}}(M)$ at each point $\mathbf{m} \in M$. In other words, it is a smooth mapping $X : M \rightarrow TM$ such that $X(\mathbf{m}) \in T_{\mathbf{m}}(M)$ for all $\mathbf{m} \in M$.

Given a vector field X, and a function $f : M \rightarrow \mathbb{R}$, defined on the manifold, one may define the derivative of f in the direction X, denoted by Xf, which is defined at each point $\mathbf{m} \in M$. More specifically, at each point $\mathbf{m} \in M$, there is defined a vector $X(\mathbf{m})$ in the tangent space at $\mathbf{m}$. Then Xf denotes the derivative of f in the direction $X(m)$ at $\mathbf{m} \in M$. Thus, Xf is a real number, for each point $\mathbf{m} \in M$. In other words, Xf is also a function $Xf : M \rightarrow \mathbb{R}$.

Notation. We denote the set (vector space) of all vector fields on a manifold M by $T^1(M)$, and the set of functions defined in M by $T^0(M)$. Thus for $f \in T^0(M)$ and $X \in T^1(M)$, the directional derivative Xf is an element of $T^0(M)$.

Now, given two vector fields, X and Y, since $Xf \in T^0(M)$, one may apply Y to this giving $Y(Xf)$, or simply YXf, which is again an element of $T^0(M)$. Doing this in the opposite direction, one can consider $XYf - YXf$, which is again an element of $T^0(M)$. The important observation one makes here is that, given X and Y, there exists a (unique) $Z \in T^1(M)$ such that $XYf - YXf = Zf$. (This is easily shown by expanding $XYf - YXf$ in a local coordinate system, using the properties of the directional derivative.) Thus, X and Y uniquely determine Z; one defines the Lie bracket operation by $[X, Y] = Z$. Clearly, the bracket operator so defined is bilinear a anti-symmetric. Once again, in a local coordinate system, one can verify the Jacobi identity. Therefore, the bracket operator makes $T^1(M)$ into a Lie algebra.

The question is, what does this have to do with the Lie algebra of a Lie group, which we have seen is a single vector space, the tangent space at the identity.

First, it is important to see that in the differential manifold setting, one cannot define a Lie bracket structure on the tangent space at a single point. Why is this? Given $X \in T^1(M)$ and $f \in T^0(M)$, on intuitive grounds one sees that the directional derivative at a point $\mathbf{m}$ depends on the value of X only at the point $\mathbf{m}$ (defining a direction in which f varies), but it requires the value of f to be defined *in a neighbourhood* of $\mathbf{m}$. To repeat, X needs to be known only at the point $\mathbf{m}$, not in a neighbourhood. **However,** when this operation is repeated by applying Y to Xf, it is necessary that Xf be defined in a neighbourhood of $\mathbf{m}$. For Xf to be defined in a neighbourhood of $\mathbf{m}$, it is necessary, therefore, for X to be defined in that neighbourhood. In a similar way, to make any sense of XYf at a point $\mathbf{m}$, it is necessary for Y to be defined in a neighbourhood of $\mathbf{m}$. To form $XYf - YXf$, therefore, both X and Y, as well as f must be defined in a neighbourhood of $\mathbf{m}$.

The question, then is what is different in the Lie group case. The answer has to do with translation of tangentvectors by the group action. To look at this, let M be a smooth manifold, and G a group. In the case of Lie groups, both M and G are the same thing, but we use different symbols M and G to emphasize their different roles in the following discussion.

Actions of groups. A (left) action of G on M is a mapping $G \times M \to M$ denoted by $(g, \mathbf{m}) \mapsto g \circ \mathbf{m}$, or simply $g\mathbf{m}$, where $g \circ \mathbf{m}$ (or simply $g\mathbf{m}$) is an element of M, such that the mapping $\mathbf{m} \mapsto g\mathbf{m}$ is smooth for all g, and such that $g(h\mathbf{m}) = (gh)\mathbf{m}$. A *right* action is the same, with the condition $g(h\mathbf{m}) = (hg)\mathbf{m}$, written more conveniently as $(\mathbf{m}h)g = \mathbf{m}(hg)$. We assume that the action is *transitive* meaning that the action of G on M takes any point to any other.

Denote the mapping $\mathbf{m} \mapsto g\mathbf{m}$ by λ_g (transformation by g), which is a smooth map from M to M. Let $\mathbf{m}_0$ be a base point in M (any point). If M is the Lie group G, then $\mathbf{m}_0$ should be the group identity. Then λ_g induces a linear transform $\lambda_g^* : T_{\mathbf{m}_0}(M) \to T_{g\mathbf{m}_0}(M)$.

Now, given a vector X in $T_{\mathbf{m}_0}$, one can define a vector $\lambda_g^*(X)$ lying in $T_{\mathbf{m}_0}$, for any g. If G acts transitively on M, then this defines a vector in the tangent space at any point. One thing needs to be checked, namely that this vector is well defined, namely that if there are two elements $g, h \in G$ such that $g\mathbf{m}_0 = h\mathbf{m}_0$, then the mappings λ_g^* and λ_h^*, taking $T_{\mathbf{m}_0}(M)$ to $T_{g\mathbf{m}_0}$ are the same, so that which vector X is mapped to does not depend on whether one transforms it via λ_g^* or λ_h^*. For the result to be the same, one requires that $\lambda_h^* \lambda_g^{*-1}$, which is a linear transformation from $T_{\mathbf{m}_0}(M)$ to itself, is equal to the identity. This is equivalent to the condition that $\lambda_{hg^{-1}}$ is the identity. Defining $\mathrm{stab}(\mathbf{m}_0)$, the *stabilizer* of the point $\mathbf{m}_0$ under the group action to be the subgroup of G that fixes $\mathbf{m}_0$, the required condition that the left translation of the vector X be well defined is that if $g \in \mathrm{stab}(\mathbf{m}_0)$, then λ_g^* is the identity map on $T_{\mathbf{m}_0}(M)$.

This is obviously true if the stabilizer of any point $\mathbf{m}_0$ is trivial, namely that G acts without fixed points. This is certainly the case of the action of a Lie group G on itself.

Getting back to the definition of the Lie bracket on the Lie algebra of a Lie group, let X and Y be vectors in $\mathfrak{g} = T_I(G)$. These can be extended to smooth vector fields (also called X and Y here) on the whole of G. From this, we may define a Lie bracket, as in section 4.

The key fact here is that the Lie bracket defined in this way for left-invariant vector fields on differential manifolds, and the Lie bracket defined as a commutator $XY - YX$ on the Lie algebra of a matrix Lie group are the same.

Exercise 4.16. Given X and Y in $\mathfrak{g}$, one may define their Lie bracket, as seen above, by extending them to left-invariant vector fields X_L and Y_L (where L stands for left) on the whole of G, and then taking the Lie bracket of vector fields.

What happens if instead of extending X and Y by left multiplication in G, we extend to two vector fields X_R and Y_R by right multiplication by elements of the group? Note that it will in general not be true that $X_L = X_R$ or $Y_L = Y_R$. However, verify that the Lie brackets $[X_L, Y_L]$ and $[X_R, Y_R]$ are the same in both cases.

5 Riemannian metrics

A Riemannian metric on a Lie group is a smooth assignment of an inner product defined on each tangent space. For a matrix Lie group, G, the tangent space at any point is a vector space of matrices. At the point $A \in G$, this is the set of matrices $A\mathfrak{g}$, as shown above. As with any Riemannian manifold, there are a multitude of Riemannian metrics, since the inner product can be chosen on each tangent space independently, as long as they vary smoothly. A natural requirement, however, in considering Riemannian metrics on Lie groups is that the choice of metric have something to do with the group operation on the Lie group.

A natural and common way to define an inner product on a vector space of matrices is the Frobenius inner product, given my

$$\begin{aligned}\langle X,\, Y\rangle_F &= \mathrm{Tr}(X^\top Y)\\ &= \sum_{i,j} X_{ij}Y_{ij}\end{aligned}$$

where i, j range over all entries to the matrices.

Since the tangent space at each point in a matrix Lie group is a vector space of matrices, one can choose the Frobenius inner product at each point to define a Riemannian metric on G. This is a possible way to define a Riemannian metric, but it will be seen that it may not always be the best choice.

Invariant metrics. A metric $\langle\cdot,\cdot\rangle$ on M is said to be left-invariant if $\langle X,\, Y\rangle_{\mathbf{m}_0} = \left\langle \lambda_g^*(X),\, \lambda_g^*(Y)\right\rangle_{g\mathbf{m}_0}$. In other words, the linear transform λ_g^* preserves the inner product.

Given an inner product defined on $T_{\mathbf{m}_0}(M)$, one can define an inner product at any other point $\mathbf{x} = g\mathbf{m}_0$ by

$$\langle X,\, Y\rangle_{\mathbf{x}} = \left\langle \lambda_g^{*1}(X),\, \lambda_g^{*-1}(Y)\right\rangle_{\mathbf{x}} \ . \tag{10}$$

The metric so defined is left-invariant.

Exercise 5.17. The discussion just above assumes that the action of the group G on M is transitive and fixed-point free, so that the stabilizer of a point in M is trivial. In this case, the mapping λ_g^* is independent of possible different choices of g. However, one can make do with a slightly weaker condition, namely that if $g \in \mathrm{stab}(\mathbf{m}_0)$ for some point $\mathbf{m}_0$, then λ_g is an isometry with respect to the inner product on $T_{\mathbf{m}_0}(M)$.

In the case of a matrix Lie group G with Lie algebra $\mathfrak{g}$, the tangent space at a point A is equal to $A\mathfrak{g}$. For an element $A \in G$, the left action of G on itself is a mapping $\lambda_A : G \to G$, with $\lambda_A(B) = AB$. The corresponding map $\lambda_A^* : T_I(G) \to T_A(G)$ is given by $X \mapsto AX$. Thus, the condition of left-invariance of the metric is given by

$$\langle X,\, Y\rangle_I = \langle AX,\, AY\rangle_A \ .$$

or equivalently

$$\langle X,\, Y\rangle_A = \left\langle A^{-1}X,\, A^{-1}Y\right\rangle_I \ . \tag{11}$$

for any matrices X and Y in the tangent space at A (which is $A\mathfrak{g}$).

First, note that if inner product at any point $A \in G$ is defined by the Frobenius inner product, then this metric is not left-invariant, for such a condition would require that

$$\begin{aligned}\mathrm{Tr}(X^\top Y) &= \mathrm{Tr}(A^{-1}XY^\top A^{-\top})\\ &= \mathrm{Tr}((AA^\top)^{-1}XY^\top)\end{aligned}$$

for any $A \in G$. This condition clearly does not hold, unless G is a subgroup of the orthogonal matrices.

Defining a left-invariant metric. A left-invariant metric on a Lie group (or a differential manifold with transitive smooth group action) can be constructed by choosing an arbitrary inner product on the tangent space at one point (that is, on $T_{\mathbf{m}_0}(M)$ or $T_I(G) = \mathfrak{g}$, and then transporting it to other points, according to (11).

For instance, in the case of matrix Lie groups, one may choose the Frobenius inner product on $\mathfrak{g}$, and transport it elsewhere using (11).

Right-invariant metrics. A metric on a Lie group is said to be *right invariant* if it satisfies an analogous condition to (11), namely that

$$\langle X,\, Y\rangle_A = \left\langle XA^{-1},\, YA^{-1}\right\rangle_I \ . \tag{12}$$

A metric that is left-invariant is not necessarily right-invariant. In fact, if a metric $\langle X,\, Y\rangle_A$ is both left and right invariant, then this means that

$$\left\langle XA^{-1},\, YA^{-1}\right\rangle_I = \left\langle A^{-1}X,\, A^{-1}Y\right\rangle_I \tag{13}$$

in other words,

$$\begin{aligned} \langle X,\, Y\rangle_I &= \langle XA,\, YA\rangle_A \\ &= \left\langle A^{-1}XA,\, A^{-1}YA\right\rangle_I \end{aligned} \tag{14}$$

for all $X, Y \in \mathfrak{g}$ and $A \in G$. The first equality holds, because of right invariance, and the second because of left-invariance. Thus, a metric is bi-invariant, if and only if conjugation by an element of G is an isometry.

For instance if $\langle\cdot,\cdot\rangle_I$ is the Frobenius norm, then this condition does not normally hold, unless G is a subgroup of the orthogonal group.

The condition above is expressed in terms of conjugation of elements in the Lie algebra $\mathfrak{g}$ by elements in the group G. As such, it makes sense in the context of matrix Lie groups.

A condition for bi-invariance of a metric can be expressed entirely in terms of the Lie algebra, as follows.

Lemma 5.18. *A left-invariant metric $\langle\cdot,\cdot\rangle$ is bi-invariant if and only if the condition*

$$\langle [Z, X],\, Y\rangle + \langle X,\, [Z, Y]\rangle = 0 \ . \tag{15}$$

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

Proof. The proof is given for matrix Lie groups. Suppose that the metric is bi-invariant, so $\langle X,\, Y\rangle = \left\langle A^{-1}XA,\, A^{-1}YA\right\rangle$. If we write $A = \exp(tZ)$, this becomes

$$\langle X,\, Y\rangle = \langle \exp(-tZ)X\exp(tZ),\, \exp(-tZ)Y\exp(tZ)\rangle$$

Now, differentiating with respect to t and setting $t = 0$ leads to

$$0 = \langle XZ - ZX,\, Y\rangle + \langle X,\, YZ - ZY\rangle \ .$$

Writing $XZ - ZX$ using the Lie bracket notation $[X, Z]$ gives the required result.

The inverse implication can be obtained by integration.

This expression is sometimes expressed in terms of the *adjoint map*, $\mathrm{ad}_Z : X \mapsto [Z, X]$, by saying that the adjoint map ad_Z is skew-adjoint for all Z. This means that

$$\langle X,\, \mathrm{ad}_Z Y\rangle = -\langle \mathrm{ad}_Z X,\, Y\rangle \ ,$$

which is easily seen to be the same as (15).

Bi-invariant metrics. One may make an arbitrary choice of inner product at the identity and create a left-invariant metric. It seems plausible that at least one such choice leads to a metric that is both left and right invariant, called bi-invariant. In fact, it can be shown that

Theorem 5.19. *If G is a compact Lie group, then there exists a bi-invariant metric on G.*

The proof goes something like this, for matrix Lie groups.

Proof. It is sufficient to define an inner product at the identity. We start with the Frobenius inner product (or any other inner product) $\langle X, Y\rangle$ for $X, Y \in \mathfrak{g}$. Now, we define a different inner product at the identity, by defining

$$\langle X,\, Y\rangle_I' = \int_{A\in G} \left\langle AXA^{-1},\, AYA^{-1}\right\rangle dA$$

where integration is carried out over compact G with respect to a left-invariant measure (Haar measure). This can be extended to a left-invariant inner product according to (11). This metric will be bi-invariant, according to (13) if

$$\langle X,\, Y\rangle_I' = \left\langle BXB^{-1},\, BYB^{-1}\right\rangle_I' \ .$$

In terms of the definition of the inner product $\langle\cdot,\cdot\rangle$ this is

$$\int_{A\in G} \left\langle ABXB^{-1}A^{-1},\, ABYB^{-1}A^{-1}\right\rangle dA = \int_{A\in G} \left\langle AXA^{-1},\, AYA^{-1}\right\rangle dA \ .$$

Taking the left hand side, and substituting $U = AB$, leads to

$$\int_{A\in G} \left\langle ABXB^{-1}A^{-1},\, ABYB^{-1}A^{-1}\right\rangle dA = \int_{U\in G} \left\langle UXU^{-1},\, UYU^{-1}\right\rangle dU \ .$$

where we have used the fact that the measure is left invariant, namely that the mapping $A \mapsto BA = U$ preserves the measure.

The following theorem extends the known result above.

Theorem 5.20. *There is a bi-invariant metric on a Lie group G if and only if G is isomorphic to $H_1 \times H_2$, and H_1 is a compact Lie group, and H_2 is a commutative Lie group.*

See [**?**] (Milnor: Curvature of left-invariant metrics on Lie groups.)

5.1 Geodesics

Once a Riemannian metric is defined on a Lie group, we may talk about the Riemannian exponential and logarithm maps. These may or may not be the same as the Lie (or matrix) exponential map. In particular, the Riemannian exponential and logarithm maps depend on the particular metric defined on the manifold, which of course the matrix exponential map does not.

The matrix exponential is related to an integration curve on the manifold, as follows. Given a vector field X, a curve $\gamma(t)$ is an integration curve of the vector field X if for all t, $\gamma'(t) = X(\gamma(t))$. So, the derivative of the curve at some point t, which is a vector in $T_{\gamma(t)}(M)$, is equal to the value of the vector field in the same tangent space. It is a basic fact that integration curves of smooth vector fields always exist, are uniquely defined by the initial point and velocity, $(\gamma(0), \gamma'(0))$ and are defined for all time t (See Lee, Differential Manifolds).

Theorem 5.21. *Given a smooth left-invariant vector field, X, the integration curve $\gamma(t)$ defined such that $\gamma(0) = I$ and $\gamma'(0) = X_0$ (a vector in $T_I(G)$) is the curve $\gamma(t) = \exp(tX_0)$, where $\exp$ is the matrix (or Lie) exponential.*

The proof is simple. One computes $\gamma'(t) = \exp(tX)X_0 = \gamma(t)X_0$, so $\gamma'(0) = X_0$ and $\gamma(0) = I$. Since the field is left-invariant, by definition $X(\gamma(t)) = \gamma(t)X_0$, which is equal to $\gamma'(t)$. So $\gamma(t)$ is an integration curve of the vector field.

Since there is nothing special about left-invariant vector fields, compared with right-invariant vector fields, the proof may be easily modified to show that the exponential curves are integration curves of right-invariant vector fields as well.

In the case of a bi-invariant metric, the matrix (or Lie) exponential and the Riemannian exponential maps are the same.

Theorem 5.22. *If the group G is equipped with a bi-invariant metric, then $\exp_M(X) = \exp_R(X)$, for all $X \in \mathfrak{g}$, where $\exp_M$ represents the matrix exponential, and $\exp_R$ denotes the Riemannian exponential map belonging to the metric. Hence, the geodesics are equal to the one-parameter curves $\exp_M(tX)$.*

Note also that according to Theorem 5.21, the curves $\exp_M(tX)$, the one-parameter subgroups, are the integral curves of left-invariant vector fields.

Proof. We consider a Lie group with a left-invariant inner product, and a Riemannian (Levi-Civita) connection $\nabla_X Y$, where X and Y are two left-invariant vector fields.

Lemma 5.23. *The connection $2\nabla_Z X = [Z, X]$ for all left-invariant vector fields X and Z if and only if the inner product $\langle \cdot, \cdot \rangle$ is bi-invariant.*

Proof. The proof in the forward direction is easy. Let X and Y be left-invariant vector fields. Then $\langle X, Y \rangle$ is constant. Therefore $\nabla_Z(\langle X, Y \rangle) = Z \langle X, Y \rangle = 0$. Expanding the expression gives

$$\begin{aligned} 0 = \nabla_Z(\langle X, Y \rangle) &= \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle \\ &= \frac{1}{2} \langle [Z, X], Y \rangle + \langle X, [Z, Y] \rangle \ . \end{aligned}$$

This is the condition for the metric to be bi-invariant, according to lemma 5.18.

For the reverse implication, we use the Koszul formula for the Levi-Civita conection ∇, namely

$$\begin{aligned} 2\langle \nabla_X Y, Z\rangle = & X\langle Y, Z\rangle + Y\langle X, Z\rangle - Z\langle X, Y\rangle \\ & + \langle [X,Y], Z\rangle - \langle [X,Z], Y\rangle - \langle X, [Y,Z]\rangle \ . \end{aligned} \tag{16}$$

Now, for left-invariant vector fields, this simplifies, so that the directional derivatives vanish (they are applied to constant quantities). Therefore, we obtain

$$2\langle \nabla_X Y, Z\rangle = \langle [X,Y], Z\rangle - \langle [X,Z], Y\rangle - \langle [Y,Z], X\rangle \tag{17}$$

which holds for all left-invariant vector fields. The last two terms in this expression are the same as (15), namely zero for a bi-invariant metric, which shows that

$$\nabla_X Y = \frac{1}{2}[X,Y] \ .$$

Lemma 5.24. *The one-parameter subgroup $\exp_M(tX_e)$ is a geodesic for all $X_e \in \mathfrak{g}$, if and only if $\nabla_X Y = [X,Y]/2$ for all left-invariant vector fields X and Y.*

(In the above notation, e is the identity of the group G and X_e is an element of $\mathfrak{g}$, which is the tangent space at the identity.)

Consider the left-invariant vector field that extends $X_e \in \mathfrak{g}$, and call it X. Suppose that an integration curve of this vector field is a geodesic, and let $\gamma(t)$ be such a curve. Then $\gamma(t) = \exp(t, X_e)$ and $\gamma'(t)$ is equal to X for all t. Then $\gamma'(0) = X_e$, and X is the vector field that extends the vector field $\gamma'(t)$ defined along $\gamma(t)$. If $\gamma(t)$ is a geodesic, then $\nabla_X(X) = 0$, by definition of geodesic. This shows that $\nabla_X(X) = 0$ for any left-invariant vector field. Now, applying this to $\nabla_{X+Y}(X+Y) = 0$ and expanding yields $\nabla_X Y + \nabla_Y X = 0$.

The torsion-free property of the Riemannian connection is $\nabla_X Y - \nabla_Y X = [X,Y]$. Together with the condition $\nabla_X Y + \nabla_Y X = 0$, this yields $2\,\nabla_X Y = [X,Y]$.

Conversely, if $2\,\nabla_X Y = [X,Y]$ then $\nabla_X X = 0$ for all left-invariant vector fields X, then this is true for the vector field extending $\gamma'(t)$ for the integration curve $\gamma(t) = \exp(tX_e)$, which is therefore a geodesic.

From this we have the following corollary.

Corollary 5.25. *Let G be a connected Lie group, for which there exists a bi-invariant Riemannian metric. Then the Lie exponential map: $\exp_M : \mathfrak{g} \to G$ is* surjective*.*

Proof. If there exists a bi-invariant metric on G, then by Theorem 5.22 the Lie exponential and Riemannian exponential maps related to this metric are the same. The Lie exponential map is defined on the whole of the tangent space at any point (that is, G is geodesically complete); the same must therefore be true of the Riemannian exponential map. By the Hopf-Rinow theorem[4] (see Wikipedia), the Riemannian exponential map is surjective, and therefore, so is the Lie exponential map.

[4]The Hopf Rinow theorem states (among other things) that the exponential map is surjective for a geodesically complete (or, equivalently, metrically complete) connected Riemannian manifold.

Corollary 5.26. *If G is a connected compact Lie group then the Lie exponential map is surjective.*

Proof. If G is compact, then a bi-invariant metric exists, according to Theorem 5.19 In addition, G is complete, because it is compact. Thus, the conditions of Corollary 5.25 hold, and the matrix exponential map is surjective.

Example 5.27. It was shown in example 3.15 that in the Lie Group $SL(2, \mathbb{R})$, the matrix exponential map does not map $\mathfrak{g}$ onto G. This is in agreement with Corollary 5.25, since $SL(2, \mathbb{R})$ is not compact, containing as it does elements of the form $\mathrm{diag}(n, 1/n)$ which are unbounded in norm. It is, however, complete, since a Cauchy sequence of matrices with determinant 1 must converge to a matrix with determinant 1. It follows that there is no bi-invariant metric on $SL(2, \mathbb{R})$, since otherwise, by Corollary 5.25 the corresponding exponential map would be surjective.

6 Curvature

The *sectional curvature* of a Riemannian manifold is given by the formula

$$K(X, Y) = \frac{\langle R(X, Y)Y,\, X\rangle}{\langle X, X\rangle \langle Y, Y\rangle - \langle X, Y\rangle^2} \ , \tag{18}$$

where $R(X, Y)Z$ is defined by

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \ . \tag{19}$$

This is equal to the Gaussian curvature of a surface formed by mapping the plane spanned by X, Y via the exponential map.

Note that the denominator in (18) is just a normalization constant, designed to ensure that the sectional curvature depends only on the subspace of the tangent space spanned by X and Y. It is positive (as long as X and Y are independent) because of the Schwarz inequality. In fact, it is equal to the square of the area of the parallelogram spanned by X and Y (exercise).

Exercise 6.28. The value of $K(X, Y)$ depends only on the span of the vectors X and Y, and not on the particular choice of vectors X and Y.

Consequently, one may define an unscaled version of sectional curvature, denoted by $\kappa(X, Y)$, to equal the numerator of (18). If X and Y are orthogonal unit vectors, or more generally, if the denominator of (18) is equal to 1, then K and κ are the same.

Note that for a Lie group with left-invariant metric, the sectional curvature at any point is determined by the sectional curvatures at the identity, by transporting the vector field at the identity via the group action.

The main reference for this section is [**?**]. (Milnor: Curvature of left-invariant metrics on Lie groups.) We note just the following theorems.

Lemma 6.29. *Given a Lie group G with a bi-invariant metric, then*

$$\kappa(X,Y) = \frac{1}{4} \langle [X,Y], [X,Y] \rangle \ , \tag{20}$$

which is non-negative for all X and Y.

Proof. Since the metric is bi-invariant, it follows from lemma 5.23 that $2\,\nabla_X Y = [X,Y]$ for all left-invariant vector fields X and Y.

Now, we compute

$$\begin{aligned} 4\,R(X,Y)Z &= \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z \\ &= [X,[Y,Z]] - [Y,[X,Z]] - 2\,[[X,Y],Z] \\ &= [X,[Y,Z]] + [Y,[Z,X]] + 2\,[Z,[X,Y]] \\ &= [Z,[X,Y]] \ , \end{aligned} \tag{21}$$

where the last line follows from the Jacobi identity.

Consequently,

$$\begin{aligned} 4\,\kappa(X,Y) &= 4\,\langle R(X,Y)Y,\, X \rangle \\ &= \langle [Y,[X,Y]],\, X \rangle \\ &= \langle [X,Y], [X,Y] \rangle \ , \end{aligned}$$

where the last line follows from a further application of the skew-adjoint property of ad_Y. This shows that $K(X,Y) \geq 0$.

From this argument, one obtains also a simple formula for the Riemannian curvature tensor defined as $R(X,Y,Z,W) = \langle R(X,Y)Z,\, W \rangle$. Supposing that G is a Lie group with bi-invariant metric, then

$$\begin{aligned} R(W,X,Y,Z) &= \langle [Y,[W,X]],\, Z \rangle /4 \\ &= -\frac{1}{4}\,\langle [W,X],\, [Y,Z] \rangle \\ &= -\,\langle \nabla_W X,\, \nabla_Y Z \rangle \ . \end{aligned} \tag{22}$$

Since compact Lie groups admit a bi-invariant metric, we have the following result.

Theorem 6.30. *Every compact Lie group admits a bi-invariant metric. For such a metric, all sectional curvatures are non-negative.*

7 Tracking with Lie groups

We consider the case where a Lie group acts on some set S (non-transitively). For two elements $\mathbf{x}, \mathbf{y} \in S$, it is desired to find the element $g \in G$ so that $g\mathbf{x}$ best approximates $\mathbf{y}$.

Example 7.31. Point alignment. Suppose that S are ordered sets of points in some Euclidean space $\mathbb{R}^m$ and G acts on $\mathbb{R}^m$. For instance, various Lie groups that act on $\mathbb{R}^m$ are discussed in section 1.1. We include also the group of homographies (see section 1.2), which are not strictly transformations of $\mathbb{R}^m$, but rather of the projective space $\mathcal{P}^m$.

For instance, G may be the Lie group $SE(3)$ of rigid Euclidean transformations, or groups of scaled Euclidean transformations, rotations, affine or projective transformations.

Each element of G acts on a point $\mathbf{x}_i \in \mathbb{R}^m$, producing a transformed point $g\,\mathbf{x}_i$. Applying G to an ordered set of points $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$, one obtains a transformed set of points $g\mathbf{x} = (g\mathbf{x}_1, g\mathbf{x}_2, \ldots, g\mathbf{x}_n)$. Here, we may consider $\mathbf{x}$ as a point in $\mathbb{R}^{n\times m}$.

Given another set of points $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n)$, one may see the group transformation g such that $g\mathbf{x}$ is "closest" to $\mathbf{y}$, in some appropriate sense. The notion of closeness suggests a distance metric, or more generally a *cost function*, in which case one wishes to find g that minimizes $d(g\mathbf{x},\, \mathbf{y})$. An appropriate cost function may be

$$C(g) = d(g\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{n} \|g\mathbf{x}_i - \mathbf{y}_i\|^2 \tag{23}$$

where $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^m$. (Observe that $d(\cdot,\cdot)$ so defined is actually the *square* of a distance metric in $\mathbb{R}^{n\times m}$). The square (exponent) in $\|\cdot\|^2$ in this expression can be replaced by some other exponent (such as 1) to obtain a more *robust* alignment.[5] The task is to find the value of g that minimizes the cost.

In a further variant of this problem, the points $\mathbf{y}$ vary with time and may be denoted by $\mathbf{y}_t$. The task of finding the sequence of transformations (group elements) g_t such that $g_t\mathbf{x} = \mathbf{y}_t$ is the problem of tracking a set of points, moving according to a common motion model.

Example 7.32. Another example we wish to consider is the problem of image alignment. Let $I(\mathbf{x})$ represent an image. Here, $\mathbf{x}$ represents a location in the image (perhaps a pixel) and $I(\mathbf{x})$ represents the intensity of that pixel. Although in practice, digital images are represented by intensities at a fixed finite set of points, or pixels, it is easier to consider $\mathbf{x}$ to be a point of $\mathbb{R}^2$ (or some region, perhaps rectangular) in $\mathbb{R}^2$. Thus, for convenience, an image may be modelled as a smooth function with bounded support.

Given a transformation g, one may apply this to an image to produce a *transformed image* gI, defined by[6]

$$gI(\mathbf{x}) = I(g^{-1}\mathbf{x}) \ . \tag{24}$$

Now, given two images I_1 and I_2 one may wish to align via some sort of specific transformation (rigid transformation, homography, affine transformation). The task is then to minimize some distance function $C(g) = d(g\,I_1,\, I_2)$, where $d(\cdot,\cdot)$ measures some sort of "distance" between the images.

A possible measure of the distance between the two images I_1 and I_2 may be given by

$$d(I_1,\, I_2)^2 = \int \left(I_1(\mathbf{x}) - I_2(\mathbf{x})\right)^2\, d\mathbf{x} \ .$$

[5] This is the topic of L_1 or L_q optimization, which will not be considered in detail. Other possibilities such as m-estimators, such as Huber cost functions are also possible.

[6] Note that the definition of gI is given in this way so that if D is the support of image I, then the support of image gI is gD. Thus applying gI is obtained from I by shifting the whole image according to the translation g.

where the integration is taken over all $\mathbf{x} \in \mathbb{R}^2$. This is the L_2 distance between the two images considered as functions. Since images have finite extent, this integral is really an integral over a bounded domain, and hence is defined. Another possibility is to take the integral only over the intersection of the supports of images I_1 and I_2. [7] The problem of aligning the two images now comes down to a problem of minimizing $C(g) = d(gI_1, I_2)^2$ over choices of g in the given group.

In the case of either of these two examples, it is not likely that the problem being posed can be solved exactly in closed form. This is particularly unlikely with the image alignment problem, since the cost of alignment depends upon the particular intensity profile of the images.[8] The cost function will generally have numerous local minima.[9]

Iteration. Therefore, it is necessary to apply an iterative algorithm to solve the problem. The usual way that this will be carried out is as follows. Given some aligment g_t (initialized with g_0), one carries out the following set of operations to update the estimate to a new value g_{t+1}. (Here, the subscript t denotes iteration number, not time or frame-number as in video alignment).

First, let us consider a standard way of iterative optimization, using Gauss-Newton method. We have in mind the optimization problem is formulated (at least locally) as optimization of a function defined on a manifold of dimension m, taking values in some Euclidean space $\mathbb{R}^n$, where $n > m$. To do optimization of a cost function over some set, which will be called G, though no use is made of the set being a group, one requires a *parametrization* of the set G, at least locally around the current estimate. What this means is that one needs a function, called here $\phi : G \times \mathbb{R}^m \to G$, where $(g, \theta) \mapsto \phi(g, \theta)$. This function provides a local parametrization of G around some point $g \in G$, in particular, around the current estimate g_t. This function should satisfy the condition that $\phi(g, \mathbf{0}) = g$, where $\mathbf{0} \in \mathbb{R}^m$.

The idea is that given a current point $g_t \in G$, and some parameter $\delta \in \mathbb{R}^m$ representing an update, one obtains a new (*updated*) value $g_{t+1} = \phi(g_t, \delta)$.

Also given is some function $f : G \to \mathbb{R}^n$, and the optimization task is to minimize $\|f(g)\|$ over values of $g \in G$. This task relates to the point alignment problem by setting $f(g) = g\mathbf{x} - \mathbf{y}$.

The optimization algorithm is described now. Starting from some initial estimate g_0, through a sequence of iterations a sequence of estimates g_t are generated until convergence, as follows.

---

[7] It is not suggested here that either of these two possibilities represent the ideal cost function for comparing two images. There are difficulties that arise from the finite extent of the images, and consequent overlap of the transformed image and the target image, and also from sampling into pixels. However, these may be thought of as implementation details to be solved when coding an algorithm.

[8] The point alignment problem has a closed-form solution for certain groups G, and depending on the particular cost function being used. For instance alignment of points under translation with an L_2 cost function is as simple as aligning the centroids of the two point sets $\mathbf{x}$ and $\mathbf{y}$. Alignment under Euclidean transformations is possible in closed form [**?**] with an L_2 cost function, and under scaled Euclidean transformations with an L_2 cost, but in this case, minimzing $d(g\mathbf{x}, \mathbf{y})$ (transforming points $\mathbf{x}$ to fit points $\mathbf{y}$) does not give the same result as minimizing $d(\mathbf{x}, g^{-1}\mathbf{y})$ (transforming points $\mathbf{y}$ to fit points $\mathbf{x}$).

[9] The standard method of approaching non-convexity in the problem of image alignment is to carry out the alignment at several levels of detail (a “pyramid”), in successively smoothed images, smoothed with a low pass (Gaussian) filter. One starts with a highly smoothed images and works down the pyramid towards a final alignment at the initial resolution of the images.

1. Compute $\epsilon = f(g_t)$, which is an element of $\mathbb{R}^n$.
2. Considering the function $F : \mathbb{R}^m \to \mathbb{R}^n$ defined by $F_t(\theta) = f(\phi(g_t, \theta))$, for $\theta \in \mathbb{R}^m$, define

$$J = \left. \frac{dF_t}{d\theta} \right|_{\theta=0} .$$

Observe that J is an $n \times m$ matrix. Normally, one requires that $m < n$, so the dimension of the parameter space, m, is less than the number of measurements, n.

3. Solve $J\delta_\theta = -\epsilon$ in least-squares. That is,

$$\delta_\theta = -(J^\top J)^{-1} J^\top \epsilon \ .$$

So, δ_θ is an m-vector, representing an update to the parameters.

4. Set $g_{t+1} = \phi(g_t, \delta_\theta)$.

This is the basic algorithm. To ensure convergence, some damping may be required, such as the use of a Levenberg-Marquardt algorithm.

There are two time consuming steps in this algorithm, namely the computation of the Jacobian, and the solution of the set of linear equations. At each time step the Jacobian is being taken of a different function.

Gauss-Newton on a Lie group. In the case of optimization on a Lie group, the obvious choice of function $\phi(g, \theta)$ is to define

$$\phi(g, \theta) = \exp_g(\theta) = g \ \exp(\theta)$$

for θ in the Lie algebra $\mathfrak{g}$.

The maps are the following: the first line shows the domain and range of each map; the second line shows the sequence of values, and the third line gives the sequence of values when $\theta = 0$. Note that the symbol e is used here to represent the identity of the Lie group G.

$$\begin{array}{ccccccc} \mathfrak{g} & \stackrel{\exp}{\to} & G & \stackrel{\lambda_g}{\to} & G & \stackrel{f}{\to} & \mathbb{R}^n \\ \theta & & \exp(\theta) & & g\exp(\theta) & & f(g\exp(\theta)) \\ 0 & & e & & g & & f(g) \end{array} \tag{25}$$

Next, we compute the derivatives of this map

$$\begin{array}{ccccccc} \mathfrak{g} & \stackrel{\exp^*}{\to} & T_e(G) & \stackrel{\lambda_g^*}{\to} & T_g(G) & \stackrel{f_g^*}{\to} & \mathbb{R}^n \\ \mathfrak{g} & \stackrel{I}{\to} & \mathfrak{g} & \stackrel{\lambda_g^*}{\to} & g\mathfrak{g} & \stackrel{f_g^*}{\to} & \mathbb{R}^n \end{array} \tag{26}$$

The first and second line shows where the maps go from and to. The map $\exp^* : \mathfrak{g} \to \mathfrak{g}$ is the identity map.

Evaluating this at the identity of G reveals that the derivative of the combined mapping is given by

$$J = f_g^* \circ \lambda_g^*$$

The derivative in a direction X is given by

$$\begin{aligned} JX &= \frac{d}{dt}\left(f(g\exp(tX))\right)\Big|_{t=0} \\ &= f_g^*(gX) \end{aligned} \tag{27}$$

and JX represents a linear mapping, expressible as a matrix in some choice of basis.

In the notation here, $f^* : T(G) \rightarrow \mathbb{R}^n$ is the derivative of f, and f_g^* is this derivative evaluated as g, and it is a mapping $T_g(G) \rightarrow \mathbb{R}^n$.

For computational purposes, we choose a basis for each tangent space $T_g(G)$. If $X_1, \ldots, X_n$ is a basis for $\mathfrak{g}$, then we can choose $gX_1, \ldots, gX_n$ for the basis of $T_g(G)$. In this case, it is clear that the mapping λ_g^* is represented by the identity matrix of dimension n. Therefore, the mapping J is represented by an $m \times n$ matrix, which is the matrix representing f^* at the point g.

The difficulty is that f^* must be evaluated anew at each iteration, at a different point g_t. This may be a considerable computational burden; in addition, the equations $J\delta = -\epsilon$ must be solved at each iteration – each time a different set of equations.

Inverse composition. The idea behind inverse composition for iteration on a Lie group is to change the goal of the iteration slightly.

In forward composition, one seeks to find the value of $g \in G$ to minimize

$$g^* = \operatorname*{argmin}_{g \in G} d(g\mathbf{x}, \mathbf{y}) \ . \tag{28}$$

This can be otherwise stated as the task of finding g^* such that

$$e = \operatorname*{argmin}_{g \in G} d(g^* g\mathbf{x}, \mathbf{y}) \ . \tag{29}$$

(Here, e is the group identity.) By contrast, the goal in inverse decomposition is to find g^* such that

$$e = \operatorname*{argmin}_{g \in G} d(g\mathbf{x}, {g^*}^{-1}\mathbf{y}) \tag{30}$$

It is easy to see that this is the same thing if $d(g\mathbf{x}, \mathbf{y}) = d(\mathbf{x}, g^{-1}\mathbf{y})$. Of course, this does not usually hold. Nevertheless, suppose that $\mathbf{x}$ and $\mathbf{y}$ can be matched exactly by the action of G. For instance, suppose that in the point matching problem, there exists a transformation $g \in G$ that maps $\mathbf{x}$ to $\mathbf{y}$ exactly. In this case $d(g^*\mathbf{x}, \mathbf{y}) = 0$, or $g^*\mathbf{x} = y$, so, because of the property of the Lie group, $\mathbf{x} = {g^*}^{-1}\mathbf{y}$. So g^* is also the solution to the inverse decomposition problem. Consequently, the goal of the inverse-decomposition method (30) is not unreasonable.

Now, the advantage of inverse decomposition is as follows. The Jacobian has to be evaluated **only once**. At each step, one solves (in least squares) a set of equations of the form $J\delta = -\epsilon$, where only the value of ϵ changes – not the matrix J. The least-squares solution to these equations is obtained by solving the system

$$(J^\top J)\delta = -J^\top \epsilon$$

Since only ϵ changes, one can speed up the solution, either by computing an LU (or Cholesky) factorization of $J^\top J$ in advance, or by computing (just once) the pseudo-inverse $(J^\top J)^{-1}J^\top$, so that $\delta = (J^\top J)^{-1}J^\top \epsilon$ can be computed rapidly.

The steps of the inverse-composition method are as follows. Recall that one is attempting to find g^* such that

$$e = \operatorname*{argmin}_{g \in G} d(g\mathbf{x},\, {g^*}^{-1}\mathbf{y})$$

At any point one has found an approximation g_t, and one proceeds as follows. Here $\phi(X) : \mathfrak{g} \to G$ is a function such as $\phi(X) = \exp(X)$. The function $f_t(X)$ is given by $f_t(X) = \phi(X)\mathbf{x} - g_t^{-1}\mathbf{y}$.

1. Compute $\epsilon = f_t(0) = \mathbf{x} - g_t^{-1}\mathbf{y}$.
2. Considering the function $F : \mathbb{R}^m \to \mathbb{R}^n$ defined by

   $$F_t(X) = f_t(\phi(X)) = \phi(X)\mathbf{x} - g_t^{-1}\mathbf{y} \ ,$$

   for $X \in \mathfrak{g}$, define

   $$J = \left.\frac{dF_t}{dX}\right|_{X=0} \ .$$

   Here, observe that J is the same at all times. (This is the key point.)
3. Solve $J\delta_X = -\epsilon$ in least-squares. That is,

   $$\delta_X = -(J^\top J)^{-1} J^\top \epsilon \ .$$

4. Set $g_{t+1} = g_t\phi(\delta_X)$, which is the same as $g_{t+1}^{-1} = \phi(\delta_X)^{-1}g_t^{-1}$.

One way to think about this is that one is assuming an approximation

$$d(\phi(\delta_X)\mathbf{x},\, g_t^{-1}\mathbf{y}) \approx d(\mathbf{x},\, \phi(\delta_X)^{-1}g_t^{-1}\mathbf{y}) \ ,$$

which explains the update step $g_t^{-1} \mapsto g_{t+1}^{-1} = \phi(\delta_X)^{-1}g_t^{-1}$.

Although complete justification of this point is perhaps lacking, one observes, writing $\delta_g = \phi(\delta_X)$, that the stationary point of this algorithm is, as desired, when $d(\delta_g\mathbf{x},\, g_t^{-1}\mathbf{y})$ is at a local minimum, with respect to small increments δ_g.

8 Homogeneous spaces

Consider a Lie group G with a Lie subgroup H (not necessarily a normal subgroup). There is a right action of H on G. Denote by G/H the set of *left* cosets of H in G.[10] The left cosets are of the form gH for $g \in G$, and they form the orbits of the action of H on G. There is a mapping $\pi : G \to G/H$ taking $g \in G$ to gH.

The set G/H may be endowed with a topology, the strongest topology such that π is continuous. With this topology, G/H is a manifold, according to the following theorem (and using the fact that G is a manifold).

Theorem 8.33 Quotient Manifold Theorem. *Suppose a Lie group H acts* **smoothly***,* **freely** *and* **properly** *on a smooth manifold G. Then G/H is a topological manifold of dimension* $\dim(G) - \dim(H)$*, and has a unique smooth structure. with the property that $\pi : G \to G/H$ is a smooth submersion.*

[10] Of course the following discussion applies also to the set of right cosets, with obvious modifications. We shall, however, always assume that we are dealing with the set of *left* cosets.

Here, G/H means the set of all orbits of the action of H on G. This may be applied in the case where G is a Lie group G show that G/H is a smooth manifold, which will be called M.

This theorem is true when G is an arbitrary manifold; it is not necessary to assume that G is a Lie group.

Proof. See Theorem 21.10 in [**?**]. (Lee – Introduction to Smooth Manifolds).

To explain the notation:

1. Freely means that the stabilizer of any point is trivial.
2. Properly means that the mapping $(g, \mathbf{m}) \mapsto (g\mathbf{m}, \mathbf{m})$ is proper (inverse images of compact sets are compact). If G is a compact Lie group, then the action is always proper.

In this case, there is also a left action of G on $G/H \equiv M$, given by $g'(gH) = (g'g)H$. This is a transitive action, so G acts transitively on M.

Viewed the other way around. One may look at this from another point of view. Consider a Lie group G acting transitively (by left action) on a manifold M and suppose that H is the stabilizer of some point x_0 in M. In this case, H is a Lie subgroup of G.

Since H is a subgroup of G, there is a *right* action of H on G, and H acts freely on G. The set of orbits of the action of H on G is equal to the set of left-cosets, G/H, that is, sets of the form gH, and these cosets are in one-to-one correspondence with the elements of M. Then, G/H may be given a topology and smooth structure such that the mapping $G/H \rightarrow M$ is a diffeomorphism.

We summarize the following relations between G, H and M.

1. G acts with a left action on M, transitively, with stabilizer H.
2. H acts on G with a right action. The set of orbits is identified as a manifold, M.
3. H also has a left action (multiplication from the left) on G.

The normal metric. We now describe how G/H may be given a Riemannian metric, called the *normal metric* derived from a suitable metric on G.

Let $\langle \cdot, \cdot \rangle_g$ be a smooth inner product on $T_g(G)$, defined at each point $g \in G$. For now, the only assumption that we make about $\langle \cdot, \cdot \rangle$ is that it is **right invariant under the action of** H . Some notation follows. Denote by $\lambda_g : G \rightarrow G$ the mapping corresponding to left multiplication by g. Thus for $g' \in G$, $\lambda_g(g') = gg'$. Similarly, define $\rho_g(g') = g'g$, the mapping corresponding to right multiplication. Note the choices of λ and ρ are meant to suggest *left* and *right* multiplication. Correspondingly, there are mappings $\lambda_g^* : T(G) \rightarrow T(G)$ and $\rho_g^* : T(G) \rightarrow T(G)$, the derivatives of the mappings λ_g and ρ_g. Here $T(G)$ is the tangent bundle of G. The mapping λ_g^* restricted to $T_{g'}(G)$ (the tangent space at g') is a linear mapping $T_{g'}(G) \rightarrow T_{gg'}(G)$ for any g'. For convenience, we shall also denote this mapping by λ_g^*. The assumption that the metric is right-invariant under the action of H is to be interpreted to mean that the mapping ρ_h is an isometry for each $h \in H$. In other words, for $X, Y \in T_g(G)$ we have

$$\langle X, Y \rangle_g = \langle \rho_h^*(X),\ \rho_h^*(Y) \rangle_{gh} \ .$$

It will often be convenient to think of tangent vectors as derivatives of curves in G. We will denote such curves as $\gamma : \mathbb{R} \to G$. When using the notation $\gamma(t)$, it will always be assumed that $\gamma(0) = e$, the identity of G. The notation $\dot\gamma$ will represent the element in $T_e(G)$ represented by the derivative of this curve evaluated as 0. Given $g \in G$ and a curve $\gamma(t)$, one can define a curve $g\gamma$ by $(g\gamma)(t) = g(\gamma(t)) = \lambda_g(\gamma(t))$. The derivative of the curve $g\gamma$ is $g\dot\gamma = \lambda_g^*(\dot\gamma)$. One can similarly define a curve γg, in which case $\dot\gamma g = \rho_g^*(\dot\gamma)$.

The projection $\pi : G \to G/H$ given by $g \mapsto gH$ induces a mapping $\pi^* : T(G) \to T(G/H)$. Restriction to a tangent space at g gives a map $\pi_g^* : T_g(G) \to T_{gH}(G/H)$. This map is a linear epimorphism. The kernel of π_g^* is known as the *vertical subspace* of $T_g(G)$. Under the metric $\langle \cdot, \cdot \rangle_g$, the orthogonal complement of the vertical subspace is known as the *horizontal subspace*. The vertical and horizontal subspaces in $T_g(G)$ will be denoted by $V_g(G)$ and $V_g^\perp(G)$.

There exists a left-action of the group G on the quotient G/H; for $g \in G$ a coset $g'H$ is mapped to $(gg')H$. It is necessary to verify that this mapping is well-defined, in that if $g_1'H = g_2'H$, then $(gg_1')H = (gg_2')H$. The verification is straightforward. It is important to note that one cannot define a right action is a similar way. If one tries to define an action $g'H \mapsto (g'g)H$, then this mapping is not well-defined.

Given a vector $X \in T_{gH}(G/H)$, it is easily seen that there exists a unique vector $\widetilde{X}$ in the horizontal subspace $V_g^\perp(G)$ such that $\pi_g^*(\widetilde{X}) = X$.

Lemma 8.34. *For all $h \in H$ and $g \in G$, $\pi_{gh}^* \circ \lambda_h^* = \pi_g^*$. The mapping $\rho_h^* : T_g(G) \to T_{gh}(G)$ maps the vertical subspace $V_g(G)$ isomorphically onto $V_{gh}(G)$. If ρ_h^* is an isometry (in other words, the metric is right-invariant under H), then it maps the horizontal subspace $V_g^\perp$ isomorphically onto $V_{gh}^\perp$.*

Proof. For all $h \in H$ and $g \in G$, $\pi_{gh}^* \circ \rho_g^* = \pi_g^*$. Let $Z \in V_g(G)$, so $\langle Z, \widetilde{X} \rangle_g = 0$. Since (by assumption) the inner product is right invariant under H, it follows that $\langle \lambda_h^*(Z), \lambda_h^*(\widetilde{X}) \rangle = 0$. If Z is a vertical vector, then we must show that $\lambda_h^*(Z)$ is vertical. If Z is vertical, then $\pi_g^*(Z) = 0$. Let $\gamma(t)g$ be a curve representing Z. Then $\pi_g^*(Z)$ is represented by the curve $g\gamma(t)H$ in G/H, and $\rho_h^*(Z)$ is represented by the curve $g\gamma(t)h$. Finally, $\pi_{gh}^*(\rho_h^*(Z)$ is represented by the curve $g\gamma(t)hH = g\gamma(t)H = \pi_g(g\gamma(t))$ Thus, $\pi_{gh}^*(\rho_h^*(Z) = \pi_g^*(Z) = 0$, as was required. □

Now, with this preparation, one can define the normal metric on G/H as follows. Given two vectors $X, Y \in T_{gH}(G/H)$, one lifts them to two vectors $\widetilde{X}, \widetilde{Y}$ in the horizontal subspace $V_g^\perp(G)$ and defines an inner product $\langle X, Y \rangle$ by

$$\langle X, Y \rangle_{gH} = \langle \widetilde{X}, \widetilde{Y} \rangle_g \ .$$

It is necessary to verify that this inner-product is well defined. In this definition, an element g is chosen and X, Y are lifted to the horizontal space $V_g^\perp(G)$ at g. Here, a choice is made, since the coset gH may also be written in a different way as $g'H$, where

$g' = gh$ for some $h \in H$. It is necessary to show that the same result arises from lifting both X and Y to gh instead.

Assume that $\widetilde{X}_g \in V_g^\perp(G)$, and $\widetilde{X}_{gh} \in V_{gh}^\perp(G)$ both project to X. It follows from lemma 8.34 that $\rho_h^*(\widetilde{X}_g) = \widetilde{X}_{gh}$, and similarly for $\widetilde{Y}_g$. Therefore

$$\left\langle \widetilde{X}_{gh}, \widetilde{Y}_{gh} \right\rangle_{gh} = \left\langle \rho_h^*(\widetilde{X}_g), \rho_h^*(\widetilde{Y}_g) \right\rangle_{gh} = \left\langle \widetilde{X}_g, \widetilde{Y}_g \right\rangle_g \tag{31}$$

so the inner-product is well-defined.

8.1 Transport

Let us now consider the effect of left multiplication on the horizontal and vertical subspaces. Recall that the space G/H affords a left action of G. For $g \in G$, denote the map $g'H \mapsto (gg')H$ by λ_g. (Thus we use the notation λ_g to represent a map from G/H to G/H as well as a map from G to G, allowing the context to determine which map is meant.) Correspondingly, there is a mapping $\lambda_g^* : T(G/H) \to T(G/H)$.

We assume that the group G has a metric $\langle \cdot, \cdot \rangle$ which is **right** invariant under the action of H. This is necessary so that the normal metric on G/H can be defined, as shown above.

Lemma 8.35. *Left-multiplication and projection commute: $\lambda_g \circ \pi = \pi \circ \lambda_g$. The map $\lambda_g^* : T_{g'}(G) \to T_{gg'}(G)$ maps the vertical subspace $V_{g'}(G)$ isomorphically onto $V_{gg'}(G)$. If in addition the metric $\left\langle \cdot, \cdot \right\rangle$ is left-invariant (under the action of the whole of G), then λ_g^* maps the horizontal subspace $V_{g'}^\perp$ onto $V_{gg'}^\perp$ isomorphically.*

Proof. Let $g' \in G$. Applying the map $\lambda_g \circ \pi$ takes $g' \mapsto g'H \mapsto gg'H$. On the other hand, $\pi \circ \lambda_g$ takes $g' \mapsto gg' \mapsto gg'H$, and so the mappings are equal. Consequently $\lambda_g^* \circ \pi^* = \pi^* \circ \lambda_g^*$, so if $X \in \ker(\pi^*)$, then $\lambda_g^* \circ \pi^*(X) = 0 = \pi^* \circ \lambda^*(X)$, and so $\lambda^*(X)$ is in $\ker(\pi^*)$, which is the vertical subspace.

If the metric is left-invariant, then λ^* preserves the horizontal subspace, by the same argument as before. $\square$

Corollary 8.36. *Let $X_0 \in V_g^\perp$, and assume that the metric $\left\langle \cdot, \cdot \right\rangle$ is left-invariant. Let $\gamma(t) = \exp(tX)$ be the corresponding curve in G (here* exp *represents the matrix exponential, which may or may not be the same as the Riemannian exponential map). Let X_t in $T_{\gamma(t)}(G)$ be the vector $\gamma'(t)$, the derivative of the curve at time t. Then, $X_t \in V_{\gamma(t)}^\perp$. Thus, the derivative of the curve remains in the horizontal subspace, for all t.*

Proof. The derivative of $\gamma(t) = \exp(tX_0)$ is given by $\exp(tX_0)X_0 = \gamma(t)X_0$. which is equal to $\lambda_{\gamma(t)}^*(X_0)$. However, according to lemma 8.35, λ_g^* preserves the horizontal subspace. $\square$

Question. This proof involves the matrix exponential. Can we devise a proof for arbitrary Lie groups, not just matrix Lie groups? Answer: probably. The exponential map in group G, starting at the identity, is a one-parameter subgroup, satisfying $\gamma(s+t) = \gamma(t)\gamma(s)$. Taking the derivative with respect to s and setting $s = 0$ gives $\gamma'(t) = \gamma(t)\gamma'(0)$. Writing X_0 for $\gamma'(0)$, we get the same result as before.

8.2 Lifting of curves

Theorem 8.37. *Let a metric in G be right-invariant under H, and G/H be equipped with the normal metric. Given a curve $\gamma(t)$ in G/H, and a point $g \in G$ such that $\gamma(0) = gH$, then there exists a unique lifted curve $\tilde{\gamma}(t)$ in G such that $\gamma(t) = \pi(\tilde{\gamma}(t))$ and $\dot{\gamma}(t)$ lies in the horizontal subspace $V^{\perp}_{\tilde{\gamma}(t)}(G)$ for all t. The lengths of the curves $\gamma(t)$ and $\tilde{\gamma}(t)$ are equal, over any range of t.*

Proof. We construct a vector field along $\gamma(t)$, thus, a vector defined at each point in the tangent spaces in a neighbourhood of the curve. Each of these vectors is lifted to a horizontal vector at each point lying above a point in the curve. Then starting at any point g, one takes the integral curve of this vector field. By definition, this curve will be horizontal at each point. Not entirely obvious is that it will cover the original curve γ.

If the curve $\gamma(t)$ crosses itself, then it is necessary to break up the γ into segments and lift each segment in this way independently. □

The proof here does not require that the metric is left-invariant.

The following theorem shows that the horizontal lifts of geodesics are geodesics, and vice versa.

Theorem 8.38. *Suppose that the metric is left-invariant, and the space G/H is equipped with the normal metric. If γ is a geodesic in G/H, and $\tilde{\gamma}(t)$ is a lifted horizontal curve in G, then $\tilde{\gamma}$ is a geodesic. All geodesics in G/H are of this form, and the lengths of the $\gamma(t)$ and $\tilde{\gamma}(t)$ are equal.*

8.3 Curvature

In the case where the metric on G is bi-invariant under the action of H, the lifting of geodesic curves in G/H are geodesic curves in G. In particular, (at least locally) given two vectors X and Y in $T_x(G/H)$, they lift to two vectors in the horizontal subspace of G. Furthermore, since exponentials lift to horizontal exponentials, the mapping through exp of the subspace of the tangent space at x, generated by X and Y may be lifted locally to a two-dimensional surface in G, generated by the horizontal lifts $\widetilde{X}$ and $\widetilde{Y}$. This surface maps locally isometrically onto the surface in G/H. Correspondingly, the Gaussian curvature of the two surfaces are the same.

The Gaussian curvature, denoted $\kappa(X, Y)$ is the same as $\kappa(\widetilde{X}, \widetilde{Y})$. These are the sectional curvatures in the two manifolds. However, by Theorem **??** the sectional curvatures in G, with a bi-invariant metric, are non-negative. Consequently, we have the following theorem.

Theorem 8.39. *If the Lie group G is equipped with a bi-invariant metric, and G/H is given the inherited normal metric, then the sectional curvature of G/H is non-negative.*

Note from Theorem 5.20 that if G is compact, or the product of compact and abelian Lie groups, then there always exists a bi-invariant metric. Hence, the quotient space has a metric with non-negative sectional curvatures.

9 Grassmann Manifolds and their Riemannian Structure

For $0 < p \leq d$, the space of $d \times p$ matrices with orthonormal columns is not a Euclidean space but a Riemannian manifold, the Stiefel manifold $\mathrm{St}(p, d)$. That is,

$$\mathrm{St}(p, d) < \{X \in \mathbb{R}^{d \times p} : X^T X = \mathbf{I}_p\}. \tag{32}$$

By grouping together all points on $\mathrm{St}(p, d)$ that span the same subspace we obtain the Grassmann manifold $\mathcal{G}(d, p)$. More formally, the Stiefel manifold $\mathrm{St}(p, d)$ admits a right action by the orthogonal group $O(p)$ (consisting of $p \times p$ orthogonal matrices); for $X \in \mathrm{St}(p, d)$ and $U \in O(p)$, the matrix XU is also an element of $\mathrm{St}(p, d)$. Furthermore the columns of X and XU span the same subspace of $\mathbb{R}^d$, and are to be thought of representatives of the same element of the Grassmann manifold, $\mathcal{G}(d, p)$. Thus, the orbits of this group action form the elements of the Grassman manifold.

An element $\mathcal{X}$ of $\mathcal{G}(d, p)$ can be specified by a basis, that is, a set of p vectors $\mathbf{x}_1, \cdots, \mathbf{x}_p$ such that $\mathcal{X}$ is the set of all their linear combinations. When the $\mathbf{x}$ vectors are ordered as the columns of a $d \times p$ matrix X, then X is said to span $\mathcal{X}$ and we write $\mathcal{X} = \mathrm{span}(X)$. In what follows, we refer to a subspace $\mathcal{X}$ and hence a point on $\mathcal{G}(d, p)$ by its $d \times p$ basis matrix X. The choice of the basis is not unique but it has no effect in what we develop later.

A Riemannian metric on a manifold is defined formally as a smooth inner product on the tangent bundle. (See [**?**] for the form of Riemannian metric on $\mathcal{G}(d, p)$). However, we shall be concerned only with geodesic distances on the Grassmann manifold, which allows us to avoid many technical points and give a straight-forward definition.

Geodesics. On a Riemannian manifold, points are connected via smooth curves. The geodesic distance between two points is defined as the length of shortest curve in the manifold (called a *geodesic*) connecting them. The Stiefel manifold $\mathrm{St}(p, d)$ is embedded in the set of $d \times p$ matrices, which may be seen as a Euclidean space $\mathbb{R}^{d \times p}$ with distances defined by the Frobenius norm. Consequently the length of a smooth curve (or path) in $\mathrm{St}(p, d)$ is defined as its length as a curve in $\mathbb{R}^{d \times p}$. Now, given two points $\mathcal{X}$ and $\mathcal{Y}$ in $\mathcal{G}(d, p)$, the distance $d_{\mathrm{geo}}(\mathcal{X}, \mathcal{Y})$ is defined as the length of the shortest path in $\mathrm{St}(p, d)$ between any two points X and Y in $\mathrm{St}(p, d)$ that are members of the equivalence classes $\mathcal{X}$ and $\mathcal{Y}$.