Elementary Matrix Lie Theory

Samuel Harshe
Yale University
Professor Igor Frenkel
MATH 480: *Mathematical Topics*
May 8, 2024

Contents

1 Introduction.

Before we go nose-to-grindstone and deal carefully with the details of Lie groups and Lie algebras, we sketch the essential definitions of this paper, motivating the preliminary results and allowing readers to try to anticipate significant connections as they go. We therefore open with the star of the show: a **Lie group** is a group with smooth composition and inversion maps that is also a smooth manifold. All of this structure makes Lie groups very nice to work with. We construct a tangent space at the identity of the group, endow it with a multiplication, and call it the **Lie algebra**. We then define a map that recovers much of the group structure from the algebra. This means we can do much of our work on the algebra, which, conveniently, is linear, instead of on the group, partially reducing group operation to a much simpler linear algebra problem. Let's pause briefly to see this in action.

Example 1.1. The group of 2×2 real orthogonal matrices with determinant 1, denoted $SO(2, \mathbb{R})$, is a Lie group under multiplication. This group corresponds to the rotations of 2-space. Unlike most Lie groups, it can be visualized as a manifold without a headache: it is (isomorphic to) the unit circle. Its Lie algebra is spanned by the matrix

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

which we'll call L for now to avoid spoiling the surprise. (Take my word for it, or use Corollary 6.19 to prove this yourself.) Why should this matrix be familiar? Observe that

$$\begin{aligned} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} &= \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \\ &= -I \end{aligned}$$

which is to say that L is i in matrix form. Exactly as we use $e^{i\theta}$ to map the span of i onto the unit circle, we use exp on matrices to map the span of L onto $SO(2, \mathbb{R})$. Thus, in this case, we recover the whole Lie group from the Lie algebra via the matrix exponential.

This toy example is both abelian and simply connected, which makes it so nice that it's deceiving. We will deal more carefully with trickier Lie groups below. Still, this should give some inutition as to why going from the group to the algebra and vice versa is natural and useful.

2 Smooth manifolds.

Perhaps surprisingly, much Lie theory can be covered without explicitly using the theory of manifolds. This is done through matrix Lie groups. Since there is more than enough matrix Lie theory to fill two lectures, and I believe it is the quickest and most intuitive way to cover ground, matrices are the idiom of most of this paper. However, it would be wrong to leave manifolds completely out of the picture. We will therefore begin with smooth manifolds and prove a connection to matrix Lie groups before leaving them in the background for the rest of the paper. At the end of the paper, in Section 8, we briefly discuss the limitations of this approach, including the Peter-Weyl Theorem (which you may have heard of).

2.1 Topological definitions.

First, we collect all the preliminary definitions we will need in our treatment of smooth manifolds. (This can be skipped if you are already familiar with the notion.) All of these are found in §5.1 of [1].

- A set X equipped with a function $\mathcal{N}(x)$ that assigns to each $x \in X$ a nonempty collection of subsets $\{N\}$ (called **neighborhoods** of x) is a **topological space** if it satisfies the following four axioms.
  1. Each point $x \in X$ belongs to each of its neighborhoods.
  2. Each superset of a neighborhood of x is also a neighborhood of x.
  3. The intersection of any two neighborhoods of x is a neighborhood of x.
  4. Any neighborhood N contains a subneighborhood M such that N is a neighborhood of each point in M.
- A set is **open** if it contains a neighborhood of each point.
- A **basis** of a topological space M is some family $\mathcal{B}$ of open subsets such that every open set in M can be generated by taking the union of some some sub-family of $\mathcal{B}$.
- A topological space is **second countable** if it has a countable basis.
- A **Hausdorff space** is a topological space M where for all $x, y \in M$, there exist U_x, U_y, neighborhoods of x and y, such that $U_x \cap U_y = \emptyset$.
- A **homeomorphism** is a continuous bijection between topological spaces with a continuous inverse.
- A **open cover** of M is a collection of open sets in U_a whose union $\bigcup U_a = M$.
- A space is **locally Euclidean of dimension** n if every point has a neighborhood homeomorphic to an open subset of $\mathbb{R}^n$.
- A **chart** is a neighborhood U of a point p together with a homeomorphism $\phi : U \to V$ for an open subset V of $\mathbb{R}^n$. The chart is written as the pair (U, ϕ).
- A function is **smooth** if it is C^∞, that is, if all its partial derivatives exist, and all their partial derivatives exist, and so on *ad infinitum*.

2.2 Topological manifolds.

All definitions are from §5.1 of [1].

Definition 2.1. We simply concatenate several definitions to define a **topological manifold of dimension *n***: a Hausdorff, second countable, locally Euclidean space of dimension n. Intuitively, this means that if you are standing on a topological manifold with terrible eyesight, as far as you can tell, you are standing in $\mathbb{R}^n$. We also require a few topological properties that make the manifold nicer to deal with.

Remark 2.2. Note that we require each neighborhood on a manifold to be homeomorphic to an open subset of $\mathbb{R}^n$. An object with some neighborhoods homeomorphic to open subsets of the upper half-plane $\mathbb{H}^n$ but not $\mathbb{R}^n$ is called a **manifold with boundary**, which we will not consider a manifold because Lie groups are never manifolds with boundaries.

Example 2.3. Two nonintersecting open line segments $M = \{(x, 0) : x \in (-1, 0) \cup (0, 1)\}$ form a 1-manifold in $\mathbb{R}^2$. Note that no part of the definition requires the manifold to be connected. Here, we have two components, each locally homeomorphic to an open neighborhood of $\mathbb{R}^1$.

To continue building our intuition, we will prove an non-example of a manifold, but first, we need a lemma.

Lemma 2.4. Homeomorphisms preserve connected components.

Proof. Let X and Y be topological manifolds and $X = \bigcup X_i$ and $Y = \bigcup Y_j$ be their decompositions into their connected components. Let $f : X \to Y$ be a homeomorphism. Because f is a homeomorphism, f is continuous, so $f(X_i)$ is connected. Therefore, $f(X_i) \subseteq Y_j$ for some Y_j. Now because f is a homeomorphism, f^{-1} is continuous, so $f^{-1}(Y_j)$ is connected. Also, as long as X_i is nonempty, $f^{-1}(Y_j) \cap X_i \neq \emptyset$. Therefore, $f^{-1}(Y_j) \subseteq X_i$, since X_i is connected. Applying f to both sides yields $Y_j \subseteq f(X_i)$, completing the proof that $f(X_i) = Y_j$. □

Non-example 2.5. The unit cross in $\mathbb{R}^2$, $M = \{(x, y) : \big(x \in (-1, 1) \text{ and } y = 0\big) \text{ or } \big(y \in (-1, 1) \text{ and } x = 0\big)\}$, is not a topological manifold.

Proof. Assume toward contradiction that there exists a local homeomorphism $f : M \to \mathbb{R}^n$ such that for all $x \in M$, there exists a neighborhood $U \ni x$ such that $f(U)$ is open in $\mathbb{R}^n$ and $f\big|_U$ is a homeomorphism. Let $x = 0$. Then there exists a neighborhood V containing x such that $f\big|_V$ is a homeomorphism. Put $f(V) = Y$. Define f' as the restriction of f to $V \setminus \{0\}$. This is a homeomorphism $V \setminus \{0\} \to Y \setminus \{f(0)\}$ because restrictions of homeomorphisms are still homeomorphisms. However, $V \setminus \{0\}$ has 4 connected components, while $Y \setminus \{f(0)\}$ has 2 components for $n = 1$ and 1 component for $f > 1$. Therefore, there exists no homeomorphism $f : M \to \mathbb{R}^n$, so M is not a topological manifold. □

I would like to reiterate here that most of this paper does not use the theory of manifolds explicitly, so you should not get caught up on the details of these results. They are meant only to hone our instincts regarding manifolds.

2.3 Smooth manifolds.

We collect all the notions needed to define smooth manifolds, all from §5.2 and §5.3 of [1].

Definition 2.6. Two charts (U, ϕ) and (V, ψ) are **compatible** if $\phi \circ \psi^{-1}$ and $\psi \circ \phi^{-1}$ are both C^∞.

In other words, we can start in Euclidean space, go up to the manifold via the inverse of one map, and come back down via the other map, all without trouble.
Observe that if $U \cap V = \emptyset$, then the functions $\phi \circ \psi^{-1}$ and $\psi \circ \phi^{-1}$ are trivially C^∞, so this definition is only restrictive for charts on nondisjoint neighborhoods.

Definition 2.7. An **atlas** on a topological manifold M is a collection $\mathcal{U} = \{(U_a, \phi_a)\}$ of pairwise compatible charts such that $\bigcup\{U_a\}$ forms an open cover of M.

Definition 2.8. An atlas (U_a, ϕ_a) is **maximal** if it is not the proper subset of any atlas on M.

Finally, our main definition:

Definition 2.9. A **smooth manifold** is a topological manifold M together with a maximal atlas.

Example 2.10. Any open subset of the Euclidean space $S \subseteq \mathbb{R}^n$ is a topological manifold with chart (S, id).

Example 2.11. The graph of $y = |x|$ in $\mathbb{R}^2$ is also a smooth manifold of dimension 1 with the coordinate map $(x, |x|) \mapsto x$. The cusp at $x = 0$ does not prevent the graph from being a smooth 1-manifold: the projection down to the x-axis remains smooth.

Example 2.12. The sphere S^1 is a smooth 1-manifold in $\mathbb{R}^2$. Define four open neighborhoods $U_+ = \{(x, y) \in S^1 : x > 0\}$, $U_- = \{(x, y) \in S^1 : x < 0\}$, $V_+ = \{(x, y) \in S^1 : y > 0\}$, and $V_- = \{(x, y) \in S^1 : y < 0\}$, with maps projecting onto the axis of the variable not restricted. Clearly, these neighborhoods cover S^1, and each is homeomorphic to an open subset of $\mathbb{R}^1$.

Again, if they don't help, please don't get stuck on the examples. If it's easier, just keep in mind the intuitive notion of a smooth manifold: some object that is locally like Euclidean space that can be covered with patches that fit together nicely.

Now, we define two basic sets of matrices, which we will soon show to be smooth manifolds.

Definition 2.13. $M(n, \mathbb{K})$ is the set of all $n \times n$ matrices with entries in $\mathbb{K}$.

Remark 2.14. From now on, we let $\mathbb{K} = \mathbb{C}$ and elide the field in our definition of matrix groups: we write $M(n)$ instead of $M(n, \mathbb{C})$, $GL(n)$ instead of $GL(n, \mathbb{C})$, and so on, flagging each case when the field is not $\mathbb{C}$. All of the results below except for Proposition 5.7 would hold also if we replaced $\mathbb{C}$ with $\mathbb{R}$ everywhere.

Definition 2.15. $GL(n)$ is the set of all $n \times n$ matrices with nonzero determinant.

To prove the following proposition, we need a lemma:

Lemma 2.16. Any open subset of a smooth manifold is a smooth manifold.

Proof. Let M' be an open subset of M and M a smooth manifold with atlas $\{(U_a, \phi_a)\}$. Then $\{(U_a \cap M', \phi_a)\}$ is an atlas of M', so M' is itself a smooth manifold. $\square$

Proposition 2.17. $GL(n)$ is a smooth $2n^2$ manifold.

Proof. To begin, we identify $M(n)$ with $\mathbb{R}^{2n^2}$ via $\mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$. It is trivial that $\mathbb{R}^{2n^2}$ is a smooth manifold: take atlas $\{(\mathbb{R}^{2n^2}, id)\}$. Thus, $M(n)$ is a smooth manifold. The map det : $GL(n) \to \mathbb{C} \setminus \{0\} \cong \mathbb{R}^2 \setminus \{0\}$ is polynomial in the entries of the matrix of which it is taken, so it is continuous. Its image is an open subset of $\mathbb{R}^2$. The preimage of any open set under a continuous map is open, so the preimage of det, which is $GL(n)$, is an open subset of $M(n)$. Therefore, by Lemma 2.16, $GL(n)$ is a smooth manifold of dimension $2n^2$. $\square$

3 Matrix groups.

Before proceding to matrix Lie theory, we prove some crucial results of a few classical matrix groups.

3.1 Maximality of the general linear group.

We first prove that $GL(n)$ is the maximal matrix group, taking here and always multiplication as our group operation. We then go on to prove that several other classical matrix groups are subgroups.

Proposition 3.1. $GL(n)$ is the maximal matrix group.

Proof. **Maximality**: Recall from linear algebra that nonsquare matrices lack inverses in the sense that there is no matrix that is both the left and the right inverse of a nonsquare matrix. Matrices with determinants of 0 also lack inverses. Then any set with such a matrix fails to provide an inverse for each of its elements, so it is not a group. Thus, no group larger than $GL(n)$ could be a matrix group. It remains to be proven directly that $GL(n)$ is a matrix group.
Associativity: $GL(n)$ inherits associativity from the definition of matrix multiplication.
Composition: Recall from linear algebra that $\det(MN) = \det(M)\det(N)$. Let $M, N \in GL(n)$. By the definition of $GL(n)$, $\det(M) \neq 0$ and $\det(N) \neq 0$. Then $\det(MN) = \det(M)\det(N) \neq 0$, so $MN \in GL(n)$.
Identity: Also, $\det(I) \neq 0$, so $GL(n) \ni I$.

Inversion: Finally, recall from linear algebra that a square matrix with nonzero determinant possesses an inverse. Then for all $M \in GL(n)$, there exists M^{-1} such that $MM^{-1} = M^{-1}M = I$. This implies $\det(M^{-1}) = 1/\det(M)$, so $\det(M^{-1}) \neq 0$, so $M^{-1} \in GL(n)$. Therefore, $GL(n)$ is a matrix group.
Therefore, $GL(n)$ is a matrix group and any set of matrices with an element not in $GL(n)$ is not a matrix group, so $GL(n)$ is the maximal matrix group. □

3.2 Classical matrix group definitions.

First, we define these classical matrix groups.

Definition 3.2. The special linear group, $SL(n)$, is defined $\{M \in GL(n) : \det(M) = 1\}$.

Definition 3.3. The special orthogonal group, $SO(n, \mathbb{R})$, is defined $\{M \in GL(n, \mathbb{R}) : M^T M = MM^T = I$ and $\det(M) = 1\}$. This is a very important matrix group because it corresponds to the rotations of 3-space. Thus, we define it and prove that it is a subgroup of $GL(n, \mathbb{R})$, even though the rest of the paper deals with complex matrices.

Definition 3.4. The special unitary group, $SU(n)$, is defined $\{M \in GL(n) : MM^* = M^*M = I$ and $\det(M) = 1\}$.

Definition 3.5. There are several ways of defining the symplectic group, $SP(2n)$, but we will take as definitional $SP(2n) = \{M \in GL(2n) : M^T \Omega M = \Omega\}$, where

$$\Omega = \begin{bmatrix} 0 & -I_n \\ I_n & 0 \end{bmatrix}.$$

It can also be defined as the group of matrices that preserve a non-degenerate skew-symmetric bilinear form. Note the $2n$: this definition breaks if we tried $SP(3)$, for example, since we couldn't fill Ω's upper-righthand and lower-lefthand quandrants with equal-sized blocks of I_n for any n.

3.3 Classical matrix groups are matrix groups.

This is fairly tedious work, but it must be done: we establish that the so-called classical matrix groups are groups under multiplication.

Proposition 3.6. $SL(n)$ is a matrix group.

Proof. **Associativity**: $SL(n)$ inherits associativity from the definition of matrix multiplication.
Composition: For $M, N \in SL(n)$, $\det(MN) = \det(M)\det(N) = 1 \cdot 1 = 1$, so $MN \in SL(n)$, so $SL(n)$ is closed under composition.

Inversion: For $M \in SL(n)$, $\det(M^{-1}) = 1/\det(M) = 1/1 = 1$, so $M^{-1} \in SL(n)$.
Identity: $\det(I) = 1$, so $SL(n) \ni I$.
Therefore, $SL(n)$ is associative, closed under inversion and composition, and contains the identity, so $SL(n)$ is a matrix group. □

Proposition 3.7. $SO(n, \mathbb{R})$ is a matrix group.

Proof. **Associativity**: $SO(n, \mathbb{R})$ inherits associativity from the definition of matrix multiplication.
Inversion: For $M \in SO(n, \mathbb{R})$, $MM^T = I$ by definition, so $(MM^T)^{-1} = I^{-1} = I$. Also

$$\begin{aligned}(MM^T)^{-1} &= (M^T)^{-1} M^{-1} \\ &= (M^{-1})^T M^{-1} \\ &= M^{-1} (M^{-1})^T.\end{aligned}$$

Each equality is given by a straightforward property of matrices from linear algebra. The final expression implies that $M^{-1}(M^{-1})^T = I$. The condition $\det(M^{-1}) = 1$ is provided by $SO(n, \mathbb{R}) \subseteq SL(n)$. The condition $M_{ij} \in \mathbb{R}$ is provided by the closure of $\mathbb{R}$ under negation and division, which are the operations on the entries that produce inversion. $M^{-1} \in SO(n, \mathbb{R})$. Therefore, $SO(n, \mathbb{R})$ is closed under inversion.
Composition: For $M, N \in SO(n, \mathbb{R})$:

$$\begin{aligned}(MN)(MN)^T &= MN(N^T M^T) \\ &= M(NN^T)M^T \\ &= MIM^T \\ &= MM^T \\ &= I.\end{aligned}$$

Again each equality is provided by a straightforward property of matrices from linear algebra. $\det(MN) = 1$ is provided by $SO(n, \mathbb{R}) \subseteq SL(n)$. The $M_{ij} \in \mathbb{R}$ condition is provided by the closure of $\mathbb{R}$ under addition and multiplication, the operations on the entries that produce composition. Therefore, $SO(n, \mathbb{R})$ is closed under composition.
Identity: $II^T = II = I$, so $SO(n, \mathbb{R}) \ni I$.
Therefore, $SO(n, \mathbb{R})$ is associative, closed under inversion and composition, and contains the identity, so $SO(n, \mathbb{R})$ is a matrix group. □

Proposition 3.8. $SU(n)$ is a matrix group.

Proof. This proof is identical to the preceding one, *mutatis mutandis*. □

Proposition 3.9. $SP(2n)$ is a matrix group.

Proof. **Associativity**: $SP(2n)$ inherits associativity from the definition of matrix multiplication.
Inversion: Let $M \in SP(2n)$, so $M^T \Omega M = \Omega$. Observe that $\Omega^{-1} = -\Omega$. Then

$$\begin{aligned}(M^T \Omega M)^{-1} &= \Omega^{-1} \\ &= -\Omega,\end{aligned}$$

so $-(M^T \Omega M)^{-1} = \Omega$, which allows finally

$$\begin{aligned}-(M^T \Omega M)^{-1} &= -M^{-1}\Omega^{-1}(M^T)^{-1} \\ &= M^{-1}\Omega(M^T)^{-1} \\ &= \Omega.\end{aligned}$$

Multiplying the penultimate and ultimate terms by M on the left and by M^T on the right gives

$$\Omega = M\Omega M^T,$$

showing that $M^T \in SP(2n)$. Then we may replace M with M^T in any of the above equations without jeopardizing the equality. Take $\Omega = M^{-1}\Omega(M^T)^{-1}$ and replace M with M^T:

$$\begin{aligned}(M^T)^{-1}\Omega((M^T)^T)^{-1} &= (M^T)^{-1}\Omega M^{-1} \\ &= (M^{-1})^T \Omega M^{-1},\end{aligned}$$

which equals Ω by the original construction, showing that $M^{-1} \in SP(2n)$, which is (finally) the desired result. Therefore, $SP(2n)$ is closed under inversion.
Composition: For $M, N \in SP(2n)$,

$$\begin{aligned}(MN)^T \Omega (MN) &= N^T M^T \Omega M N \\ &= N^T \Omega N \\ &= \Omega,\end{aligned}$$

so $MN \in SP(2n)$, so $SP(2n)$ is closed under composition.
Identity: Trivially, $I^T \Omega I = I \Omega I = \Omega$, so $SP(2n) \ni I$.
Therefore, $SP(2n)$ is associative, closed under inversion and composition, and contains the identity, so $SP(2n)$ is a matrix group. □

Before moving on to matrix Lie groups, we emphasize that this section has established that $SL(n)$,

$SO(n,\mathbb{R})$, $SU(n)$, and $SP(n)$ are subgroups of $GL(n)$ (stipulating that n is even in the case of $SP(n)$, respecting our previous notation). With Cartan's Theorem (Corollary 6.18), to be established later, this gets us further than we might expect.

4 Matrix Lie groups.

4.1 Definitions and discussion.

There are at least three definitions of "matrix Lie group" that could come into effect here. They are intuitively very different but we will show them to be equivalent.

Definition 4.1. A matrix Lie group is a matrix group that is also a Lie group.

In particular, this means that the group is realized as a set of matrices satisfying the group axioms, that the group has smooth inversion and composition maps, and also that the group has smooth manifold structure. This is the most honest and least insightful definition. To use this with no other machinery, for example, we might need to build a new atlas for each matrix Lie group by hand.

Definition 4.2. A matrix Lie group is a subgroup of a matrix group and a submanifold of a manifold.

This definition is slightly smarter. It does not require us to establish all the structure for our matrix Lie group *de novo*. Instead, all it asks is that we show that the set in question can inherit all the relevant properties from a superset. If we allow a matrix group to be a subgroup of itself and a manifold to be a submanifold of itself, this definition is also perfectly broad.

Definition 4.3. A matrix Lie group is a subgroup of $GL(n)$ closed under nonsingular limits.

We have already established that $GL(n)$ is the maximal matrix group and that $GL(n)$ is a smooth manifold, so this definition should not come completely out of the blue. However, it should still be surprising. For any subgroup G of $GL(n)$, as long as any convergent sequence of matrices $M_n \in G$ converges to a matrix in G or leaves $GL(n)$ altogether, G is a matrix Lie group. No direct proof of its manifold structure is necessary. This is the route that allows one to study Lie theory in a surprising amount of depth without ever touching manifolds. (With Corollary 6.18 below, we prove that this is equivalent to Definition 4.1.)

4.2 Direct proof: the general linear group is a Lie group.

It would be cheating to go through this whole paper without a single complete and direct proof that some object is a Lie group. It has already been shown that $GL(n)$ is a matrix group and a smooth manifold, so all that remains is to show that the inversion and composition maps are smooth.

Theorem 4.4. *$GL(n)$ is a Lie group.*

Proof. Let $\mu : GL(n) \times GL(n) \to GL(n)$ be the matrix multiplication map (which is our group composition map). Then $\mu(A, B) = AB$, with entries calculated

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}.$$

This is a polynomial in the entries of A and B, so it is C^∞.
Let $\iota : GL(n) \to GL(n)$ be the matrix inversion map (which is our group inversion map). Recall from linear algebra that the (i, j)-minor, denoted $M_{i,j}$, of a matrix is the determinant of the submatrix formed by deleting row i and column j of the matrix. The formula for entries of the inverse of a matrix A is

$$\begin{aligned}(\iota(A))_{ij} &= (A^{-1})_{ij} \\ &= \frac{1}{\det(A)} (-1)^{i+j} M_{j,i}\end{aligned}$$

by Cramer's rule. $M_{j,i}$ is a polynomial in the entries of A, so it is C^∞. Also, $A \in GL(n)$, so $\det(A) \neq 0$, so $1/\det(A)$ is C^∞. Therefore this formula is C^∞ in the entries of A, so the inversion map is C^∞.
This completes the proof that $GL(n)$ is a Lie group. □

5 Matrix exponential and logarithm functions.

We now pivot to build the machinery that will be used to connect Lie groups to Lie algebras. (Definitions taken from Chapter 2 of [2].)

5.1 Matrix exponential function.

Recall the power series of e^x for $x \in \mathbb{C}$:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}.$$

This is a theorem due to Euler in the scalar case, but we will take it as our definition in the matrix case.

Definition 5.1.

$$\begin{aligned}e^X = \exp X &= \sum_{n=0}^{\infty} \frac{X^n}{n!} \\ &= I + \sum_{n=1}^{\infty} \frac{X^n}{n!}.\end{aligned}$$

Some texts use e^X rather than $\exp X$. The latter is preferred here because it can be confusing to conflate the scalar- and matrix-valued functions, and also because it reads better in-line. (We will elide the parentheses to improve readability whenever it does not detract from clarity.)

It is not yet clear that this is a sensible notion, for we do not know whether the power series converges.

Definition 5.2. To help prove the convergence of this power series, we define the **Frobenius norm** (sometimes also called the **Hilbert-Schmidt norm**) of a matrix:

$$\|X\| = \left[\sum_{i,j=1}^{n} (x_{ij})^2 \right]^{1/2}.$$

The following two inequalities follow quickly from properties of the metric on the field over which $M(n, \mathbb{K})$ is defined, and are not proven here.

Proposition 5.3.

$$\|X + Y\| \leq \|X\| + \|Y\|$$
$$\|XY\| \leq \|X\|\|Y\|.$$

By induction on the second inequality, setting $X = Y$, we also have $\|X^n\| \leq \|X\|^n$.

These properties are crucial in establishing key results regarding the exponential map.

Theorem 5.4. *The exponential map,* exp*, converges absolutely and is continuous for all* $X \in M(n)$.

Proof. We define the convergence of a sequence of matrices entrywise, which means that a sequence of matrices $\{X_m\}$ converges to X if and only if $(X_m)_{ij} \to (X)_{ij}$ as $m \to \infty$. It follows that X_m converges to X if and only if $\|X_m - X\| \to 0$ as $m \to \infty$. Then

$$\sum_{n=0}^{\infty} \frac{\|X^n\|}{n!} \leq \sum_{n=0}^{\infty} \frac{\|X\|^n}{n!}.$$

The right-hand side is the power series for $e^{\|X\|}$ with $\|X\| \in \mathbb{R}$, which converges absolutely. Therefore, exp converges absolutely.
Note also that each entry in X^n is a product of the entries of X, so X^n is a continuous function of X for all $n \in \mathbb{N}$, so the partial sums are continuous. By the Weierstrass M-test, exp converges uniformly for matrices with norms in $(0, \|X\|)$. For any matrix $M \in M(n)$, we can choose $M + \varepsilon$ so that exp converges uniformly on an open set containing M. Therefore, by the uniform convergence theorem, exp is continuous on all of $M(n)$. □

Proposition 5.5. We now prove several important properties of the matrix exponential function.

1. $\exp 0 = I$.

2. If $XY = YX$, $\exp(X + Y) = \exp(X)\exp(Y) = \exp(Y)\exp(X)$.

3. $\exp(X)^{-1} = \exp(-X)$.

4. $\exp((\alpha + \beta)X) = \exp(\alpha X) \cdot \exp(\beta X)$ for $\alpha, \beta \in \mathbb{C}$.

5. For all $C \in GL(n)$, $\exp(CXC^{-1}) = C\exp(X)C^{-1}$.

Proof. (1) follows straightforwardly from the definition of the power series beginning at $n = 1$. To see (2), consider $\exp(X)\exp(Y)$, multiplying term-by-term, which is permitted because both series converge absolutely. In particular, let us multiply term-by-term in such a way as to collect all terms where powers add to m. This means:

$$\begin{aligned}\exp(X)\exp(Y) &= \sum_{m=0}^{\infty}\sum_{n=0}^{m}\frac{X^n}{n!}\frac{Y^{m-n}}{(m-n)!}\\ &= \sum_{m=0}^{\infty}\frac{1}{m!}\sum_{n=0}^{m}\frac{m!}{n!(m-n)!}X^nY^{m-n},\end{aligned}$$

after multiplying by $m!/m!$ and rearranging. Now, note that exponentiation of matrices does not behave exactly like exponentiation of reals. For example

$$(X + Y)^2 = X^2 + XY + YX + Y^2,$$

which does not equal the familiar form

$$\sum_{n=0}^{2}\binom{m}{n}X^nY^{m-n}$$

if X and Y fail to commute. In this case, however, since X and Y commute,

$$(X + Y)^m = \sum_{n=0}^{m}\frac{m!}{n!(m-n)!}X^nY^{m-n}.$$

Plugging in to the previous equation

$$\begin{aligned}\exp(X)\exp(Y) &= \sum_{m=0}^{\infty}\frac{(X+Y)^m}{m!}\\ &= \exp(X + Y).\end{aligned}$$

To prove (3), let $Y = -X$. We know that $-XX = X(-X)$, so $-X$ and X commute, so (2) applies. Then

$$\begin{aligned}\exp(-X+X) &= \exp(-X)\exp(X)\\ &= \exp 0\\ &= I,\end{aligned}$$

so $\exp(-X) = \exp(X)^{-1}$. This implies that $\exp(X) \in GL(n)$ for all $X \in M(n)$.
(4) is also a special case of (2), since αX is equivalent to $(\alpha I)X$, and αI commutes with every $X \in M(n)$.
To prove (5), note that $(CXC^{-1})^n = CX^nC^{-1}$. This implies that the power series of $\exp(CXC^{-1})$ and $C\exp(X)C^{-1}$ are term-by-term equivalent. □

Proposition 5.6. Let $X \in M(n)$. Then for $t \in \mathbb{R}$, $\exp(tX)$ is a smooth curve in $M(n)$, and

$$\begin{aligned}\frac{d}{dt}\exp(tX) &= X\exp(tX)\\ &= \exp(tX)X.\end{aligned}$$

This implies

$$\frac{d}{dt}\exp(tX)\bigg|_{t=0} = X,$$

since $\exp 0 = I$.

Proof. This follows simply from differentiating the power series term by term. This is permitted because each entry $(\exp(tX))_{jk}$ of $\exp(tX)$ is given by a convergent power series in t, and one can differentiate a power series term by term within its radius of convergence. This holds for all entries of the matrix, so it holds for the matrix as a whole. □

Here and throughout, keep in mind that the notion of smoothness is the standard analytic one, where convergence on our matrices is defined entrywise.

The following proposition is powerful, but we will not have the space to do much with it in this paper. Still, it is quite striking, so we prove it anyway.

Proposition 5.7. For all X in $M(n)$,

$$\det(\exp X) = e^{tr(X)}.$$

Proof. Suppose X is diagonalizable with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $X = CDC^{-1}$ for D diagonal, so $\exp(X) = \exp(CDC^{-1}) = C\exp(D)C^{-1}$ by Property (5) of Proposition 5.5. We can see that $\exp(D)$ is a diagonal matrix with eigenvalues $e^{\lambda_1}, \ldots, e^{\lambda_n}$. Then $\det(X) = \det(D) = e^{\lambda_1} \cdot \ldots \cdot e^{\lambda_n}$.

On the other hand, $tr(X) = tr(D) = \lambda_1 + \ldots + \lambda_n$. Therefore, $e^{tr(X)} = e^{tr(D)} = e^{\lambda_1+\ldots+\lambda_n} = e^{\lambda_1} \cdot \ldots \cdot e^{\lambda_n}$.

If $X \in M(n, \mathbb{C})$ is not diagonalizable, then there is a sequence $\{D_n\}$ with $\lim_{n\to\infty} D_n = D$ such that $X = CDC^{-1}$ for some invertible matrix C. Then also $tr(X) = tr(D)$ and $\det(X) = \det(D)$, so the proof goes through identically.

Note that this does not hold for $X \in M(n, \mathbb{R})$ because an arbitrary real matrix cannot be approximated by a diagonal matrix. □

5.2 Matrix logarithm function.

It is also necessary to define the matrix logarithm function, which, as we will see, is the inverse of the matrix exponential in its radius of convergence.

For $z \in \mathbb{C}$, recall that $\log(z)$ is defined and holomorphic in a circle of radius 1 about $z = 1$. (Proof in pages 36–7 of [2].) This function has the following two crucial properties:

$$e^{\log z} = z$$

for z with $|z - 1| < 1$, and

$$\log e^u = u$$

for u with $|u| < \log 2$. (This condition means $|e^u - 1| < 1$.)

Definition 5.8. Analogously, we define for $X \in M(n)$

$$\log X = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{(X - I)^n}{n}$$

whenever the series converges. Since the series has radius of convergence 1 for $z \in \mathbb{C}$ and $\|(X - I)^n\| \leq \|X - I\|^n$ for $n \geq 1$, if $\|X - I\| < 1$, the matrix-valued series converges and is continuous.

Remark 5.9. Even outside that radius, the series converges if $X - I$ is nilpotent, that is, if there exists n such that $(X - I)^n = 0$. (We call such an X **unipotent**.)

Proposition 5.10. Within this radius of convergence,

$$\exp(\log X) = X.$$

Also, for all $X \in M(n)$ with $\|X\| < \log 2$, we have $\|\exp X\| < 1$, and

$$\log(\exp X) = X.$$

A proof of this can be found in pages 38–9 of [2].

Proposition 5.11. For all $X, Y \in M(n)$, we have

$$\exp(X+Y) = \lim_{m\to\infty} \big(\exp(X/m)\exp(Y/m)\big)^m.$$

Proof. This proposition is very useful later on. Originally, its proof was omitted with a reference to pages 39–40 of [2], but Jorge very helpfully pointed out that this can be proven without too much difficulty using the Taylor expansion of log (which we have taken as its definition in Definition 5.8) as follows.

As $m \to \infty$, for any X and Y in $M(n)$, X/m and $Y/m \to 0$, so $\exp(X/m)$ and $\exp(Y/m) \to I$, so $\exp(X/m)\exp(Y/m) \to I$. Therefore, $\exp(X/m)\exp(Y/m)$ is in the domain of log for m large enough.

For X with $\|X\| < \frac{1}{2}$, we can calculate

$$\log(X+I) = \sum_{n=1}^{\infty}(-1)^{n+1}\frac{(X+I-I)^n}{n!},$$

which gives an approximation $X + O(X^2)$ after cancelling $I - I$.

Employing the Taylor expansion of exp (as in Definition 5.1), we can also calculate that

$$\exp(X/m)\exp(Y/m) = \frac{X}{m} + \frac{Y}{m} + O\Big(\frac{1}{m^2}\Big).$$

Thus,

$$\log\big(\exp(X/m)\exp(Y/m)\big) = \log\Big(I + \frac{X}{m} + \frac{Y}{m} + O\Big(\frac{1}{m^2}\Big)\Big)$$

for m sufficiently large. By the previous approximation, this can be simplified as

$$\begin{aligned}\log\big(\exp(X/m)\exp(Y/m)\big) &= \frac{X}{m} + \frac{Y}{m} + O\Big(\Big[\frac{X}{m} + \frac{Y}{m} + O\Big(\frac{1}{m^2}\Big)\Big]^2\Big)\\ &= \frac{X}{m} + \frac{Y}{m} + O\Big(\frac{1}{m^2}\Big).\end{aligned}$$

Then, exponentiating each side yields

$$\exp(X/m)\exp(Y/m) = \exp\Big[\frac{X}{m} + \frac{Y}{m} + O\Big(\frac{1}{m^2}\Big)\Big],$$

so, raising each side to the power of m,

$$\big(\exp(X/m)\exp(Y/m)\big)^m = \exp\Big[X + Y + O\Big(\frac{1}{m}\Big)\Big].$$

By Theorem 5.4, exp is continuous, so

$$\lim_{m\to\infty}\Big(\exp(X/m)\exp(Y/m)\Big)^m = \lim_{m\to\infty}\exp\left[X+Y+O\Big(\frac{1}{m}\Big)\right]$$
$$= \exp(X+Y).$$

□

5.3 One-parameter subgroups.

With the matrix exponential and logarithm functions defined, we can now define one-parameter subgroups, which are used in generating the Lie algebra of a Lie group.

Definition 5.12. A one-parameter subgroup of $GL(n)$ is a group homomorphism $A : (\mathbb{R},+) \to GL(n)$ from the reals under addition to $GL(n)$. This implies the following:

- A is continuous.
- $A(t+s) = A(t)A(s)$.
- $A(0) = I$.

We state but do not prove the following lemma, of which a proof be found on pages 41–2 of [2].

Lemma 5.13. Fix some ε with $\varepsilon < \log 2$. Let $B_{\varepsilon/2}$ be the ball of radius $\varepsilon/2$ around the 0 in $M(n)$, and let $U = \exp(B_{\varepsilon/2})$. Then every $X \in U$ has a unique square root Y in U, defined by $Y^2 = X$. This square root Y is given by by $Y = \exp(\frac{1}{2}\log X)$.

Intuitively, this says that if we take some sufficiently small ball around the origin as the preimage of exp, every matrix X in the image has a unique square root Y also in the image, which is found simply by $Y = \exp(\frac{1}{2}\log X)$. It is easy to check that $Y^2 = X$, so the only real work is to prove uniqueness.

With this, we can prove the crucial result relating one-parameter subgroups to the matrix exponential function.

Proposition 5.14. If A is a one-parameter subgroup of $GL(n)$, there exists a unique $n\times n$ matrix X such that $A(t) = \exp(tX)$.

Proof. Uniqueness is immediate: if there exists X such that $\exp(tX) = A(t)$, then $X = \frac{d}{dt}A(t)\big|_{t=0}$. Now to prove existence. Let $U = \exp(B_{\varepsilon/2})$ as in Lemma 5.13, so U is an open set in $GL(n)$. By the continuity of A, there exists $t_0 > 0$ such that $A(t) \in U$ for all t where $|t| \le t_0$. In other words,

A is continuous, and its output is in an open set U at $t = 0$, so if we keep t within some t_0 of 0, the output will stay within U. Now define

$$X = \frac{1}{t_0} \log(A(t_0)),$$

which means that

$$t_0 X = \log(A(t_0)).$$

Because $\log(A(t_0)) \in B_{\varepsilon/2}$, we also have $t_0 X \in B_{\varepsilon/2}$ and $\exp(t_0 X) = \exp(\log(A(t_0))) = A(t_0)$. *A fortiori*, $A(t_0/2)$ is also in U, and $A(t_0/2)^2 = A(t_0)$ by a property of group homomorphisms (Definition 5.12). By Lemma 5.13, $A(t_0)$ has a unique square root in U, which is $\exp(t_0 X/2)$. This implies that $A(t_0/2) = \exp(t_0 X/2)$. By induction, $A(t_0/2^k) = \exp(t_0 X/2^k)$ for all $k \in \mathbb{N}$. We have

$$\begin{aligned} A(mt_0/2k^2) &= A(t_0/2^k)^m \\ &= \exp(t_0 X/2^k)^m, \\ &= \exp(mt_0 X/2^k) \end{aligned}$$

with the first equality by a property of group homomorphisms (Definition 5.12), the second by the inductive result that $A(t_0/2^k) = \exp(t_0 X/2^k)$, and the third by the property of the exponential function $\exp(M)^m = \exp(mM)$.
Therefore, $A(t) = \exp(tX)$ for all $t = mt_0/2^k$. There exist m, t_0, k to recover any arbitrary t because the set of numbers of the form $t = mt_0/2^k$ is dense in $\mathbb{R}$. Moreover, $\exp(tX)$ and $A(t)$ are both continuous. Therefore, $A(t) = \exp(tX)$ for $t \in \mathbb{R}$. □

This is exciting because it means that exp is all we need to define an arbitrary group homomorphism from $(\mathbb{R}, +)$ to $GL(n)$.

Proposition 5.15. The exponential map, exp, is smooth.

Proof. We have already proven that $\exp(tX)$ is smooth, but the present proposition is different: we are proving that we can take the derivative in the direction of an arbitrary matrix, which is stronger than taking the derivative with respect to the parameter t. Nonetheless, this proof proceeds similarly. Note that each entry $(X^m)_{jk}$ of X^m is a homogeneous polynomial of degree m in the entries of X. Thus, the series for the function $(X^m)_{jk}$ has the form of a multivariable power series. Since the series converges on all of $M(n)$, it is permissible to differentiate the power series term-wise as many times as desired, which means that the function $(X^m)_{jk}$ is smooth. The smoothness of the exponential map follows immediately. □

The following proposition is noteworthy because it further establishes the power of exp. It will not be used anywhere in this paper, so its proof is omitted with a reference to a sketch on page 48 of [2].

Proposition 5.16. For $X \in GL(n)$, there exists $A \in GL(n) : \exp(A) = X$.

6 Lie algebras.

We now have the tools to make good use of the notion of a **Lie algebra**, which we will define then apply.

6.1 Definitions.

Definition 6.1. A Lie algebra $\mathfrak{g}$ over $\mathbb{K}$ is a $\mathbb{K}$-vector space with a bracket operation $[\cdot, \cdot]$ that satisfies the following properties:

- bilinearity: $[aX + bY, Z] = a[X, Z] + b[Y, Z]$ and $[Z, aX + bY] = a[Z, X] + b[Z, Y]$
- antisymmetry: $[X, Y] = -[Y, X]$
- Jacobi identity: $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$

for $X, Y, Z \in \mathfrak{g}$ and $a, b \in \mathbb{K}$.

Definition 6.2. The vector space of matrices in $M(n)$ with the bracket defined by the commutator

$$[X, Y] = XY - YX$$

is denoted $\mathfrak{gl}(n)$, or $\mathfrak{gl}$. (We use lowercase Gothic characters to denote Lie algebras, with the Lie algebra corresponding to a Lie group G as $\mathfrak{g}$.)

Proposition 6.3. The vector space of square matrices with the bracket defined by the commutator, $\mathfrak{gl}$, is a Lie algebra.

Proof. We check that the commutator satisfies bilinearity, antisymmetry, and the Jacobi identity. $[aX + bY, Z] = (aX + bY)Z - Z(aX + bY)$ by the definition of the commutator, which equals

$$\begin{aligned}
&= aXZ + bYZ - ZaX - ZbY \\
&= aXZ + bYZ - aZX - bZY \\
&= aXZ - aZX + bYZ - bZY \\
&= a(XZ - ZX) + b(YZ - ZY) \\
&= a[X, Z] + b[Y, Z].
\end{aligned}$$

The proof for linearity in the second coordinate is identical, *mutatis mutandis*.

Straightforwardly from the definition of the commutator, it follows that

$$\begin{aligned}[X,Y] &= XY - YX \\ &= -(YX - XY) \\ &= -[Y,X].\end{aligned}$$

Expanding by the definition of the commutator

$$\begin{aligned}&[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] \\ &= [X, YZ - ZY] + [Y, ZX - XZ] + [Z, XY - YX] \\ &= X(YZ - ZY) - (YZ - ZY)X + Y(ZX - XZ) - (ZX - XZ)Y + Z(XY - YX) - (XY - YX)Z \\ &= XYZ - XZY - YZX + ZYX + YZX - YXZ - ZXY + XZY + ZXY - ZYX - XYZ + YXZ \\ &= 0,\end{aligned}$$

cancelling like terms.
Therefore, the commutator is a valid bracket, so $\mathfrak{gl}$ is a Lie algebra. □

The following theorem is significant but not at all easy to prove, so we merely state it, with a reference to [3], whose proof is concise but hardly trivial.

Theorem 6.4. ***Ado's Theorem****. Every finite-dimensional Lie algebra $\mathfrak{g}$ over a field $\mathbb{K}$ of characteristic zero is isomorphic to a Lie algebra of square matrices under the commutator bracket.*

This means that as long as we are working over non-pathological fields, $\mathfrak{gl}$ and its subalgebras are the only ones we need. Our $\mathfrak{gl}$, a very simple matrix Lie algebra, covers a lot of ground. This theorem is one reason that we get further than we might expect using matrix Lie groups.

Example 6.5. The most familiar example of a Lie algebra is $\mathbb{R}^3$ equipped with the traditional cross-product. To prove that the cross product is a valid Lie bracket operation, it suffices to demonstrate that it is antisymmetric and satisfies the Jacobi identity on the basis vectors. Antisymmetry is definitional:

$$\hat{\mathbf{i}} \times \hat{\mathbf{j}} = -\hat{\mathbf{j}} \times \hat{\mathbf{i}},$$

$$\hat{\mathbf{j}} \times \hat{\mathbf{k}} = -\hat{\mathbf{k}} \times \hat{\mathbf{j}},$$

$$\hat{\mathbf{k}} \times \hat{\mathbf{i}} = -\hat{\mathbf{i}} \times \hat{\mathbf{k}}.$$

The Jacobi identity follows from computation:

$$\begin{aligned}\hat{\mathbf{i}} \times (\hat{\mathbf{j}} \times \hat{\mathbf{k}}) + \hat{\mathbf{j}} \times (\hat{\mathbf{k}} \times \hat{\mathbf{i}}) + \hat{\mathbf{k}} \times (\hat{\mathbf{j}} \times \hat{\mathbf{j}}) &= \hat{\mathbf{i}} \times \hat{\mathbf{i}} + \hat{\mathbf{j}} \times \hat{\mathbf{j}} + \hat{\mathbf{k}} \times \hat{\mathbf{k}} \\ &= 0.\end{aligned}$$

Remark 6.6. Any commutative algebra is also trivially a Lie algebra, where $[\cdot,\cdot] \equiv 0$ because $XY = YX$.

6.2 The Lie algebra of a matrix Lie group.

Definition 6.7. Let G be a matrix Lie group. The **"Lie algebra"** of G, denoted $\mathfrak{g}$, is the set of all matrices X such that $\exp(tX) \in G$ for all $t \in \mathbb{R}$.

This definition states that the "Lie algebra" of a Lie group is the set of all matrices whose corresponding one-parameter subgroup lies entirely in G. We call it a "Lie algebra" for now because we have not shown that it is a Lie algebra according to Definition 6.1.

Note that $\exp X \in G$ does not necessarily imply $X \in \mathfrak{g}$. Our requirement is stronger.

The following theorem establishes that $\mathfrak{g}$, our "Lie algebra," is indeed a Lie algebra in the sense of Definition 6.1.

Theorem 6.8. *Let G be a matrix Lie group with "Lie algebra" $\mathfrak{g}$. Then for all $X, Y \in \mathfrak{g}$, the following results hold.*

1. *For all $A \in G$, we have $AXA^{-1} \in \mathfrak{g}$.*
2. *For all $s \in \mathbb{R}$, we have $sX \in \mathfrak{g}$.*
3. *$X + Y \in \mathfrak{g}$.*
4. *$XY - YX \in \mathfrak{g}$.*

It follows from (2) and (3) that $\mathfrak{g}$ is a vector space, and from (4) that $\mathfrak{g}$ is closed under the bracket $[X,Y] = XY - YX$, so it is a Lie algebra.

Proof. For (1), recall that
$$\exp(t(AXA^{-1})) = A\exp(tX)A^{-1}$$
by Proposition 5.5. This is in G because all three of its terms are in G.
For (2), note that $\exp(t(sX)) = \exp((ts)X)$, which is in G by the definition of $\mathfrak{g}$. This implies that $sX \in \mathfrak{g}$.
For (3), note that
$$\exp(t(X+Y)) = \lim_{m\to\infty} [\exp(tX/m)\exp(tY/m)]^m$$
by Proposition 5.11. We know $\exp(tX/m)$ and $\exp(tX/m)$ are in G. We know that G is closed under composition, so $\exp(tX/m)\exp(tY/m) \in G$. Exponentiation is repeated composition and, again, G is closed under composition, so $[\exp(tX/m)\exp(tY/m)]^m \in G$. By Proposition 5.5

$\exp(t(X+Y))$ in $GL(n)$. G is defined as a matrix Lie group, so by Definition 4.3 it is closed in $GL(n)$. Therefore $\exp(t(X+Y)) \in G$. This finally shows that $\exp(t(X+Y)) \in G$, so $X+Y \in \mathfrak{g}$.
For (4), let $X, Y \in \mathfrak{g}$ and consider

$$\begin{aligned}\frac{d}{dt}(\exp(tX)Y\exp(-tX))\Big|_{t=0} &= (XY)\exp(0) + (\exp(0)Y)(-X)\\ &= XY - YX.\end{aligned}$$

Now by (1), $\exp(tX)Y\exp(-tX) \in G$ for all $t \in \mathbb{R}$. Moreover, by (2) and (3), $\mathfrak{g}$ is a real subspace of $M(n)$, so it is topologically closed, so

$$\lim_{t\to 0}\frac{\exp(tX)Y\exp(-tX) - Y}{t}$$

remains in the subspace. This is the definition of the derivative of $\exp(tX)Y\exp(-tX)$ at $t = 0$, which has just been shown to equal $XY - YX$. Therefore, $\mathfrak{g}$ is closed under the bracket $[X, Y] = XY - YX$.
This completes the proof that $\mathfrak{g}$ is a Lie algebra, so our notion of "Lie algebra" is indeed a Lie algebra, and we can remove the scare quotes. □

Remark 6.9. We do not have the space to do much with the bracket of a Lie algebra in this paper, so I would at least like to remark on it here. Our bracket, the commutator, can be interpreted as a measure of non-abelian-ness: if X and Y commute, $[X, Y] = 0$, while if XY and YX differ significantly, then $[X, Y] = XY - YX$ is very large. The purpose of the bracket, then, is to recover some of the non-abelian structure of the group. If we did not endow our Lie algebra with a bracket, our only operation would be vector addition, which is commutative, so the algebra would obliterate all of the non-abelian structure of the group. For example, we saw above that $\exp(X)\exp(Y) = \exp(X+Y)$ if and only if X and Y commute. To reprise non-commutative group operations on our vector space, we use the bracket in the following formula.

Definition 6.10. (**Baker-Campbell-Hausdorff formula**.)

$$\exp(X)\exp(Y) = \exp(X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] + [Y, [Y, X]]) + \ldots),$$

continuing by adding higher-order compositions of the commutator.

To guarantee convergence, though, we must restrict the norms, as usual.

Proposition 6.11. The Baker-Campbell-Hausdorff formula

$$\log\big(\exp(X)\exp(Y)\big) = X + Y + \frac{1}{2}[X, Y] + \frac{1}{12}([X, [X, Y]] + [Y, [Y, X]]) + \ldots$$

converges absolutely for $\|X\| < 1$ and $\|Y\| < 1$.

For discussion of this and several other results related to the convergence and calculation of the Baker-Campbell-Hausdorff Formula, see [4]

For further discussion of the applications of the Baker-Campbell-Hausdorff Formula, an essential application of the bracket, that builds on the approach taken in this paper, see Chapter 5 of [2].

Two straightforward facts about the correspondence between Lie groups and Lie algebras follow. Then, we prove a theorem that gets us the rest of the significant results of this paper.

Definition 6.12. The identity component of G, denoted G_0, is the connected component of G containing the identity. In the context of matrix Lie groups, connectedness is equivalent to path-connectedness, so G_0 is also the path-connected component of identity.

Proposition 6.13. Let G be a matrix Lie group and $X \in \mathfrak{g}$ an element of its Lie algebra. Then $\exp X \in G_0$.

Proof. By the definition of $\mathfrak{g}$, $\exp(tX) \in G$ for all $t \in \mathbb{R}$. Then as t goes from 0 to 1, $\exp(tX)$ goes from I to $\exp X$, so I and $\exp X$ are path-connected, so $\exp X \in G_0$. $\square$

Proposition 6.14. If G is commutative, then $\mathfrak{g}$ is commutative.

Proof. For $X, Y \in M(n)$, we can calculate

$$[X, Y] = \frac{d}{dt}\left(\frac{d}{ds}\exp(tX)\exp(sY)\exp(-tX)\Big|_{s=0}\right)\Big|_{t=0}.$$

If $X, Y \in \mathfrak{g}$ and G is commutative, then $\exp(tX)$ commutes with $\exp(sY)$, giving

$$\begin{aligned}[X, Y] &= \frac{d}{dt}\left(\frac{d}{ds}\exp(tX)\exp(-tX)\exp(sY)\Big|_{s=0}\right)\Big|_{t=0} \\ &= \frac{d}{dt}\left(\frac{d}{ds}\exp(sY)\Big|_{s=0}\right)\Big|_{t=0}.\end{aligned}$$

We are differentiating a function that is independent of t with respect to t, so $[X, Y] \equiv 0$, so $\mathfrak{g}$ is commutative. $\square$

The reverse direction requires additionally that G be connected. It will be shown shortly.

This lemma is a bit tricky, but it helps us get the following theorem from which many interesting results follow.

Lemma 6.15. Let $\{B_m\}$ be a sequence of matrices in G such that $B_m \to I$ as $m \to \infty$. Define $Y_m = \log B_m$, which is defined for all sufficiently large m because log is defined around I. Suppose

$Y_m \neq 0$ for all m: this is equivalent to supposing $B_m \neq I$ for all m. Define further that $Y_m/\|Y_m\| \to Y \in M(n)$ as $m \to \infty$. Then $Y \in \mathfrak{g}$.

Proof. For any $t \in \mathbb{R}, Y_m(t/\|Y_m\|) \to tY$ by construction. $B_m \to I$, so $\|Y_m\| \to 0$. Then we can construct a sequence of integers k_m such that $k_m\|Y_m\| \to t$. Then

$$\exp(k_m Y_m) = \exp\left[(k_m\|Y_m\|)\frac{Y_m}{\|Y_m\|}\right] \to \exp(tY),$$

since the parentheses within the bracket approach t and the fraction within the bracket approaches Y. On the other hand,

$$\exp(k_m Y_m) = (\exp(Y_m))^{k_m}$$

for an integer k_m. This equals ${B_m}^{k_m}$ by the definition of B_m. We know that G is closed under multiplication, so it is closed under exponentiation, so $B_m \in G$. This implies that $\exp(tY) \in G$. Then, by definition, $Y \in \mathfrak{g}$. □

Theorem 6.16. *For* $0 < \varepsilon < \log 2$*, let* $U_\varepsilon = \{X \in M(n) : \|X\| < \varepsilon\}$ *and let* $V_\varepsilon = \exp U_\varepsilon$*. Suppose* $G \subseteq GL(n)$ *is a matrix Lie group with Lie algebra* $\mathfrak{g}$*. Then there exists* $\varepsilon \in (0, \log 2)$ *such that for all* $A \in V_\varepsilon$*,* $A \in G$ *if and only if* $\log A \in \mathfrak{g}$.

Proof. Begin by identifying $M(n)$ with $\mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$. Let $\mathfrak{g}^\perp$ denote the orthogonal complement of $\mathfrak{g}$ with respect to the usual inner product on $\mathbb{R}^{2n^2}$. Let $Z = X \oplus Y$ with $X \in \mathfrak{g}$ and $Y \in \mathfrak{g}^\perp$. Consider $\Phi : M(n) \to M(n)$ given by $\Phi(Z) = \Phi(X, Y) = \exp(X)\exp(Y)$. The exponential is smooth, so Φ is also smooth. By $D_Z\Phi(0)$ denote the derivative of Φ at 0 in the direction Z:

$$D_Z\Phi(0) = \lim_{t\to 0}\frac{\Phi(0 + tZ) - \Phi(0)}{t}.$$

Then

$$\begin{aligned} D_{(X,0)}\Phi(0,0) &= \frac{d}{dt}\Phi(tX, 0)\Big|_{t=0} \\ &= (X, 0), \end{aligned}$$

and

$$\begin{aligned} D_{(0,Y)}\Phi(0,0) &= \frac{d}{dt}\Phi(0, tY)\Big|_{t=0} \\ &= (0, Y), \end{aligned}$$

both by direct calculation, employing the definition $\Phi(X, Y) = \exp(X)\exp(Y)$. Then

$$\begin{aligned} D_Z\Phi(0) &= D_{(X,Y)}\Phi(0,0) \\ &= D_{(X,0)+(0,Y)}\Phi(0,0), \end{aligned}$$

first by the definition of Z and second simply by addition. By the linearity of the derivative, this equals

$$\begin{aligned} D_{(X,0)}\Phi(0,0) + D_{(0,Y)}\Phi(0,0) &= (X,0) + (0,Y) \\ &= Z. \end{aligned}$$

This is all to say that $D_Z\Phi(0) = Z$. Recall that the full derivative of a multivariable function is a linear transformation: if we fix a point and define a direction vector, the derivative tells us how to transform the direction vector. The above work shows that this linear transformation $D\Phi$ at 0 is the identity. Thus, it is nonsingular. We have already shown exp to be locally injective (Proposition 5.10). This means that the Inverse Function Theorem, which states that any locally injective function with non-singular derivative has a local inverse, applies. (For more on the IVT, which I had completely wrong in my second lecture, see Chapter 8 of [5].)
By the Inverse Function Theorem, Φ has a continuous inverse in a neighborhood of $\Phi(0) = I$.
Let $A \in V_\varepsilon \cap G$. Assume toward contradiction that $\log A \notin \mathfrak{g}$. By the local inverse of Φ permitted by the Inverse Function Theorem, $A_m = \exp(X_m)\exp(Y_m)$ for sufficiently large m, with $X_m, Y_m \longrightarrow 0$ as $m \longrightarrow \infty$. Then $Y_m \neq 0$, otherwise we would have

$$\begin{aligned} \log A_m &= \log[\exp(X_m)\exp(Y_m)] \\ &= \log[\exp(X_m)\, I] \\ &= \log(\exp X_m), \end{aligned}$$

which is in $\mathfrak{g}$, which violates our supposition that $\log A \notin \mathfrak{g}$. However, $\exp(X_m)$ and $A_m \in G$, so defining

$$\begin{aligned} B_m &= \exp(-X_m)A_m \\ &= \exp(Y_m), \end{aligned}$$

which is in G. The unit sphere in $\mathfrak{g}^\perp$ is compact, so there exists a subsequence $\{Y_m\}$ such that $Y_m/\|Y_m\|$ converges to $Y \in \mathfrak{g}^\perp$, where $\|Y\| = 1$. But by Lemma 6.15 this implies that $Y \in \mathfrak{g}$. Yet $\mathfrak{g}^\perp$ is the orthogonal complement of $\mathfrak{g}$, so the two are only trivially nondisjoint, so $Y \in \mathfrak{g}$ and $Y \in \mathfrak{g}^\perp$ is a contradiction. Therefore, there must be ε such that $\log A \in \mathfrak{g}$ for all $A \in V_\varepsilon \cap G$. □

This theorem is proven because many important results follow quickly as corollaries, although in itself it isn't obviously clarifying.

Corollary 6.17. If G is a matrix Lie group with corresponding Lie algebra $\mathfrak{g}$, then there exist U, a neighborhood of 0 in $\mathfrak{g}$, and V, a neighborhood I in G, such that the exponential map takes U homeomorphically onto V.

Proof. Let ε be small enough that Theorem 6.16 holds. Set $U = U_\varepsilon \cap \mathfrak{g}$ and $V = V_\varepsilon \cap G$.

Then Theorem 6.16 implies that $\exp : U \to V$ is surjective. Moreover, exp is a homeomorphism, since there is a continuous inverse map $\log\big|_V$. ☐

In plain English, this means that the matrix exponential is a bijection in some neighborhood of the origin. This also means that the dimension of a Lie group as a manifold is the dimension of its Lie algebra.

Corollary 6.18. (**Cartan's Theorem**). Any closed subgroup H of a Lie group G is a Lie subgroup (and thus a submanifold) of G.

Proof. By the previous corollary, $\exp^{-1} : U \to V$ is a diffeomorphism from some neighborhood of $I \in G$ to some neighborhood of $0 \in V$. This implies that $\exp^{-1}\big|_H : U \cap H \to V \cap \mathfrak{h}$ is a diffeomorphism from some neighborhood of H at the identity to some neighborhood of $\mathfrak{h}$ at the identity. Then $(H \cap U, \exp^{-1}\big|_H)$ is a chart, and we can left-multiply to get a chart for any other $h \in H$ ([6]). ☐

This corollary completes the line of reasoning that shows our definitions Definition 4.1 and Definition 4.2 equivalent. Given our earlier direct proof showing $GL(n)$ to be a Lie group, this proves that closed subgroups inherit the relevant manifold structure and are therefore Lie subgroups.

Corollary 6.19. Suppose $G \subseteq GL(n)$ is a matrix Lie group with corresponding Lie algebra $\mathfrak{g}$. Then $X \in \mathfrak{g}$ if and only if there exists a smooth curve γ in $M(n)$ with $\gamma(t) \in G$ for all t and such that $\gamma(0) = I$ and $d\gamma/dt|_{t=0} = X$.

Proof. The forward direction is easy. Define $\gamma(t) = \exp(tX)$. Then $\gamma(0) = I$ and $d\gamma/dt|_{t=0} = X$ by properties of exp already established. In the other direction, let $\gamma(t)$ be smooth with $\gamma(0) = I$. For sufficiently small t, we have $\gamma(t) = \exp(\delta(t))$, where δ is a smooth curve in $\mathfrak{g}$. The derivative of $\delta(t)$ at $t = 0$ is the same as the derivative of $f(t) : t \mapsto t\delta'(0)$ at $t = 0$. This trick simplifies the algebra. Then by the chain rule

$$\begin{aligned}\gamma'(0) &= \frac{d}{dt}\exp(\delta(t))\Big|_{t=0}\\ &= \frac{d}{dt}\exp(f(t))\Big|_{t=0}\\ &= \frac{d}{dt}\exp(t\delta'(0))\Big|_{t=0}\\ &= \delta'(0).\end{aligned}$$

Now $\gamma(t) \in G$ by construction, and $\exp(\delta(t)) = \gamma(t)$ for sufficiently small t, so $\delta(t) \in \mathfrak{g}$ for sufficiently small t by the definition of $\mathfrak{g}$. Then also $\delta'(0) \in \mathfrak{g}$, so $\gamma'(0) \in \mathfrak{g}$. ☐

This means that $\mathfrak{g}$ is the tangent space at the identity to G. Many textbooks work the other way, defining the Lie algebra as the tangent space at the identity and recovering other properties we already have.

To prove the following corollary, we require this lemma, whose proof we omit with a reference to page 73 of [2].

Lemma 6.20. Suppose $A : [a, b] \to GL(n)$ is continuous. Then for all $\varepsilon > 0$, there exists $\delta > 0$ such that for s, t where $|s - t| < \delta$, $\|A(s)A(t)^{-1} - I\| < \varepsilon$.

Corollary 6.21. If G is a connected matrix Lie group, then every element A of G can be written in the form $A = \exp(X_1) \dots \exp(X_m)$ for $X_1, \dots, X_m \in \mathfrak{g}$.

Proof. Let $V_\varepsilon = \exp(U_\varepsilon)$ for U_ε a neighborhood of 0, as in Theorem 6.16. For $A \in G$, define a continuous path $\gamma : [0, 1] \to G$ where $\gamma(0) = I$ and $\gamma(1) = A$. By Lemma 6.20, we can pick $\delta > 0$ such that $\gamma(s)\gamma(t)^{-1} \in V_\varepsilon$ for $|s - t| < \delta$.
Next, we partition $[0, 1]$ into m pieces of size $1/m$, choosing m so that $1/m < \delta$. Then for $j \in \{1, \dots, m\}$, we have $\gamma((j - 1)/m)^{-1}\gamma(j/m) \in V_\varepsilon$ because the two arguments are within δ of each other. This implies that

$$\gamma((j - 1)/m)^{-1}\gamma(j/m) = \exp(X_j)$$

for some element $X_j \in \mathfrak{g}$. Then

$$\begin{aligned} A &= \gamma(1) \\ &= \gamma(0)\gamma(0)^{-1}\gamma(1) \\ &= I \cdot \gamma(0)^{-1}\gamma(1) \\ &= \gamma(0)^{-1}\gamma(1) \\ &= \gamma(0)^{-1}\gamma(1/m)\gamma(1/m)^{-1} \dots \gamma((m - 1)/m)\gamma((m - 1)/m)^{-1}\gamma(1) \\ &= \exp(X_1) \dots \exp(X_m) \end{aligned}$$

for $X_1, \dots X_m$ as constructed earlier. □

Thus, if G is connected, we can build up all of its elements out of pieces produced by the matrix exponential.

Corollary 6.22. If G is a connected matrix Lie group and the Lie algebra $\mathfrak{g}$ of G is commutative, then G is commutative.

Proof. Since $\mathfrak{g}$ is commutative, any two elements of G, when written as in Corollary 6.21, will commute. □

This further establishes the correspondence between Lie groups and their Lie algebras.

Corollary 6.23. If G is a matrix Lie group, the identity component $G_0 \subseteq G$ is a closed subgroup of $GL(n)$ and thus a matrix Lie group. Moreover, $\mathfrak{g}_0 = \mathfrak{g}$.

Proof. Take $\{A_m\}$, a sequence in G_0 converging to some $A \in GL(n)$. If G is a matrix Lie group, then G is closed under nonsingular limits by Definition 4.3, so $A \in G$. Moreover, $A_m A^{-1} \in G$ for all m because G is a group. Also, $A_m A^{-1} \to I$ as $m \to \infty$ because $A_m \to A$. By Theorem 6.16, $A_m A^{-1} = \exp(X)$ for $X \in \mathfrak{g}$ for m large enough. Left-multiplying by $\exp(-X)$ and right-multiplying by A gives $\exp(-X)A_m = A$. Because $A_m \in G_0$ by construction, there is a path joining I to A_m. Since $\exp(-X)A_m = A$, the path $\exp(-tX)A_m$ connects A_m to A, letting t go from 0 to 1. Combining this path with the path from I to A_m provides a path from I to A, so $A \in G_0$. Therefore, G_0 is a closed subgroup of $GL(n)$, so it is a matrix Lie group.
Now, since $G_0 \subseteq G$, it follows that $\mathfrak{g}_0 \subseteq \mathfrak{g}$. Now, pick an arbitrary element X from $\mathfrak{g}$. By the definition of $\mathfrak{g}$, we have $\exp(tX) \in G$ for all $t \in \mathbb{R}$. Consider an arbitrary element $Y = \exp(t_0 X) \in G$. Then $\exp(tX)$ connects I to Y, letting t go from 0 to t_0. Then $\mathfrak{g} \subseteq \mathfrak{g}_0$. Therefore, $\mathfrak{g}_0 = \mathfrak{g}$. □

This is a limitation on the ability to recover Lie group structure from Lie algebra structure. It means that two groups with the same identity component have the same Lie algebra regardless of whether they correspond globally. For example, $O(n, \mathbb{R})$, the group of $n \times n$ orthogonal matrices, has the same identity component as $SO(n, \mathbb{R})$, which you may recall is the strict subgroup of $O(n, \mathbb{R})$ created by adding the restriction that det = 1. Thus, the two share a Lie algebra.

We state one more essential result of the Lie group-Lie algebra correspondence. The standard proof requires Ado's Theorem (Theorem 6.4), so we will omit it with reference to [7], which has helpful discussion and links, and page 152 of [8], which is classic but as of now beyond my paygrade.

Theorem 6.24. *(**Lie's Third Theorem**, or the **Cartan-Lie Theorem**) Any finite-dimensional Lie algebra $\mathfrak{g}$ is isomorphic to the Lie algebra of some Lie group G.*

In this paper, we have mostly discussed how to generate a Lie algebra $\mathfrak{g}$ from a Lie group G. This result shows that for finite-dimensional Lie algebras, we can go the other way, too.

7 Classical matrix groups are closed.

I hope that the preceding section will give the reader some familiarity with the basics of the correspondence between Lie groups and Lie algebras. Now, we tie up some loose ends with proofs of the closure of the classical matrix groups.

The following proofs all go through very similarly: we prove that each matrix group is the preimage of a closed set under a continuous map.

Proposition 7.1. $SL(n)$ is closed.

Proof. The determinant map $\det : GL(n) \to \mathbb{R}^2$ is a polynomial in the entries of the input. The set $\{1\}$ is closed in $\mathbb{C}$. By definition $SL(n) = \{\det^{-1}(\{1\})\}$.
Therefore, $SL(n)$ is the preimage of a closed set under a continuous map, so it is closed. □

Proposition 7.2. $SO(n, \mathbb{R})$ is closed.

Proof. The determinant map $\det : GL(n, \mathbb{R}) \to \mathbb{R}$ is a polynomial in the entries of the input. Define $L : M(n, \mathbb{R}) \to M(n, \mathbb{R})$ by $L(M) = M^T M$ and $R : M(n, \mathbb{R}) \to M(n, \mathbb{R})$ by $R(M) = MM^T$. The sets $\{1\}$ and $\{I\}$ are closed. By definition $SO(n, \mathbb{R}) = \{L^{-1}(\{I\}) \cap R^{-1}(\{I\}) \cap \det^{-1}(\{1\})\}$.
Therefore, $SO(n, \mathbb{R})$ is the finite intersection of the preimage of closed sets under continuous maps, so it is closed ([9]). □

Proposition 7.3. $SU(n)$ is closed.

Proof. This proof is identical to that for $SO(n, \mathbb{R})$, *mutatis mutandis*. □

Proposition 7.4. $SP(n)$ is closed.

Proof. Define $S : M(2n) \to M(2n)$ by $S(M) = A^T \Omega A - \Omega$. S is a polynomial in the entries of the input, so it is continuous. The set $\{0\}$ is closed in $M(2n)$. By definition $SP(2n) = S^{-1}(\{0\})$.
Therefore, $SP(2n)$ is the preimage of a closed set under a continuous map, so it is closed ([10]). □

Then, by Corollary 6.18, these classical matrix groups are Lie subgroups and therefore also submanifolds of $GL(n)$ (or, in the case of $SO(n, \mathbb{R})$, of $GL(n, \mathbb{R})$.)

8 Limitations of matrix groups.

This ends the main part of our paper, which deals with matrix Lie groups and their Lie algebras. Now, we briefly discuss a significant result regarding the reach of this approach.

The following is a consequence of Peter-Weyl. Its proof is beyond the scope of this paper. For more, see pages 191–5 of [11].

Theorem 8.1. *All compact Lie groups are matrix groups.*

As this suggests, however, there are non-compact Lie groups that do not have faithful representations as matrices. The most common example of such a Lie group is $\overline{SL(2, \mathbb{R})}$, the universal cover of $SL(2, \mathbb{R})$. Another example, slightly easier to show, is that the quotient of the Heisenberg group H, defined as

$$M \in M(n, \mathbb{R}) = \left\{ \begin{pmatrix} 1 & a & c \\ 0 & 1 & b \\ 0 & 0 & 1 \end{pmatrix} \right\}$$

by the discrete normal subgroup N

$$N \in H \cap M(n, \mathbb{Z}) = \left\{ \begin{pmatrix} 1 & 0 & n \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}$$

is not a matrix Lie group. For more on this group, see §4.8 of [2].

The classification of non-compact Lie groups is much more difficult than of compact Lie groups, since the latter can make use of classification of compact Lie algebras. However, all Lie groups, being manifolds, are locally compact. Therefore, all Lie groups are locally isomorphic to a matrix Lie group. And even some non-compact Lie groups have faithful finite-dimensional matrix representations. The general linear group and the Heisenberg group are two such examples! Thus, this paper's matrix-based approach, although far from perfectly broad, goes a long way in elementary Lie theory.

Thank you for taking the time to read! Please still feel free to send me your feedback and questions. □♥

References

[1] Loring C. Tu. *An Introduction to Manifolds*. Springer Universitexts, 2011.

[2] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer Graduate Texts in Mathematics, 2015.

[3] Yurii A. Neretin. A construction of finite-dimensional faithful representation of Lie algebra. In *Proceedings of the 22nd Winter School Geometry and Physics*, pages 159–161. Circolo Matematico di Palermo, 2003.

[4] Sergio Blanes and Fernando Casas. On the convergence and optimization of the Baker–Campbell–Hausdorff formula. *Linear Algebra and its Applications*, 378:135–158, 2004.

[5] James R. Munkres. *Analysis on Manifolds*. Advanced Book Classics, 1991.

[6] Zuoqin Wang. Lecture 11: Cartan's closed subgroup theorem. https://bit.ly/3PX0AhF.

[7] nLab. Lie's three theorems. https://bit.ly/3Uv5HqT.

[8] J.P Serre. *Lie Algebras and Lie Groups*. W. A. Benjamin, Inc., 1992.

[9] Freakish. Showing SO(n) is a compact topological group for every n. https://bit.ly/4cTQ6cV (2023-05-18).

[10] B.Hueber. Show that the symplectic group $\mathrm{Sp}(2n, \mathbb{C})$ is a closed sub group of $\mathrm{Gl}(2n, \mathbb{C})$. https://bit.ly/3Jj2O7n (2020-11-08).

[11] A.W. Knapp. *Lie Groups Beyond an Introduction*. Birkhäuser, 1996.