THE BAKER-CAMPBELL-HAUSDORFF FORMULA AND SIMPLY CONNECTED LIE GROUP HOMOMORPHISMS

AASHIR MEHROTRA

Abstract. In this paper, we explore matrix Lie groups, which are groups in the space of real or complex matrices. The Lie group structure can be used to prove properties of many important groups such as the unitary and orthogonal groups, among others. We explore matrix Lie groups along with Lie algebras, a non-associative skew-symmetric algebra that satisfies a property known as the Jacobi identity. Lie groups and Lie algebras go hand in hand, as locally they can be bijectively mapped to one another by using the exponential and logarithm functions on a matrix, which we define and prove properties of in this paper. We prove that every Lie algebra homomorphism gives rise to a unique Lie group homomorphism, which is equal to the composition of the said Lie algebra homomorphism along with the exponential function. In order to prove this result, we first prove the Baker-Campbell-Hausdorff formula, which shows that $\log(e^X e^Y)$ can be expressed as an infinite series of nested brackets in X and Y, provided both the matrices are sufficiently small in magnitude.

1. Introduction

Lie groups are groups that are also differentiable manifolds, such that the product and inversion operations are smooth. In this paper, we focus on a special case of Lie group, namely matrix Lie groups.

We define a topology on the space of $n \times n$ complex matrices, which allows us to define topological properties such as connectedness and simple-connectedness on groups of matrices.

By using a matrix norm and by absolute continuity, it is possible to define the exponential of a matrix to be the convergent infinite sum that is identical to the complex Taylor series expansion of e^z, with z being replaced by X. While the scalar and matrix exponential satisfy common properties, it is, in general, not true that $e^{X+Y} = e^X e^Y$ for complex matrices X and Y. The matrix logarithm can also be defined for matrices, though just like in the complex case, we must restricted the domain to the open ball $\|A - I\| < 1$ in order to avoid the logarithm to be a multi-valued function. Just like in the complex case, it is possible to prove that the exponential and logarithm function are inverses one another in local neighbourhood of I and 0.

A (real) Lie algebra is a (real) vector space $\mathfrak{g}$ along with a product map $[\cdot] : V \times V \to V$ that is billinear, symmetric, and satisfies the Jacobi product:

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

The Lie bracket $[\cdot]$ needn't be associative.

Every matrix Lie group has an associated Lie algebra, which is the set of matrices X such that $e^{tX} \in G$ for all real numbers t. We show that a Lie algebra defined from this way is indeed a real Lie algebra as defined earlier. We also prove that for any matrix Lie group homomorphism (which is a continuous group homomorphism) $\Phi : G \to H$ between

Date: July 15, 2023.

two matrix Lie groups G and H, there exists a unique Lie algebra homomorphism (which is a linear map that preserves that Lie bracket $[\cdot]$) $\phi : \mathfrak{g} \to \mathfrak{h}$, where $\mathfrak{g}$ and $\mathfrak{h}$ are the Lie algebra of G and H respectively. Moreover, the functions Φ and ϕ satisfy:

$$e^{\phi(X)} = \Phi(e^X)$$

for all $X \in \mathfrak{g}$.

Our main theorem for this paper is a partial converse to this result, which provides a canonical matrix Lie group homomorphism given a Lie algebra homomorphism, such that the domain matrix Lie group G is simply-connected.

Theorem 1.1. *Let G and H be matrix Lie groups with associated Lie algebra $\mathfrak{g}$ and $\mathfrak{h}$ respectively. If $\phi : \mathfrak{g} \to \mathfrak{h}$ is a Lie algebra homomorphism, and G is simply connected, then there exists a unique Lie group homomorphism $\Phi : G \to H$ such that*

$$\Phi(e^X) = e^{\phi(x)}$$

for all $X \in \mathfrak{g}$.

In order to prove this result, we first prove the Baker-Campbell-Hausdorff Formula (or BCH formula for short), which has consequence other key results in Lie Theory. Suppose

$$g(A) = \frac{\log A}{1 - A^{-1}},$$

which is defined for $\|A - 1\| < 1$. Then the BCH formula states if X, Y are $n \times n$ complex matrices with $\|X\|$ and $\|Y\|$ sufficiently small, then

$$\log(e^X e^Y) = X + \int_0^1 g(e^{\mathrm{ad}_Y} e^{t\mathrm{ad}_Y})(Y)dt,$$

where ad_Y is the function on $\mathfrak{g}$ that sends X to $[X, Y]$. The above integral can be expressed in the form of a series of nested brackets in X and Y.

This series formulation of the BCH formula is useful in the proof of Theorem 1.1, as it is used to prove that a "local homomorphism" with the desired properties exists if given a Lie algebra homomorphism. To extend this local homomorphism into a global one, the condition that G is simply-connected is required.

2. Lie Groups

We first start with definitions concerning matrix Lie groups. In what follows we denote $M_n(\mathbb{F})$ as the ring of $n \times n$ matrices over the field $\mathbb{F}$, and $GL_n(\mathbb{F})$ as the group of $n \times n$ invertible matrices over $\mathbb{F}$. Also, log will mean a logarithm in base e.

Definition 2.1. If $A \in M_n(\mathbb{C})$, then we define the **norm** of A to be

$$\|A\| = \sup_{x \in \mathbb{C}^n \setminus \{0\}} \frac{\|Ax\|}{\|x\|}, \tag{2.1}$$

where $\|v\|$ for $v \in \mathbb{C}^n$ is $\sqrt{\sum_{i=1}^n |v_i|^2}$.

Note that other norms can be given to matrices such as the largest absolute magnitude of its entries, or the square root of all sums of the squares of the magnitudes of each entry, similar to the vector norm. Nonetheless, all such other norms can be proven to be equivalent to the norm defined above, so that there is no ambiguity when using $\| \cdot \|$.

The expression 2.1 satisfies the usual properties of a norm (such as the triangle inequality). However, the norm is not multiplicative, but rather sub-multiplicative.

Claim 2.2. *For all* $A, B \in M_n(\mathbb{C})$*,* $\|AB\| \leq \|A\|\|B\|$*.*

Proof. We have

$$\|AB\| = \sup_{v \neq 0} \frac{\|ABv\|}{\|v\|} = \sup_{Bv \neq 0} \frac{\|ABv\|}{\|v\|} = \sup_{Bv \neq 0} \frac{\|ABv\|}{\|Bv\|} \frac{\|Bv\|}{\|v\|} \leq \sup_{w \neq 0} \frac{\|Aw\|}{\|w\|} \sup_{v \neq 0} \frac{\|Bv\|}{\|v\|} = \|A\|\|B\|.$$

■

The norm gives rises to a metric space, and hence a topology over $M_n(\mathbb{C})$. We can thus create a subspace topology over any subset of $M_n(\mathbb{C})$, including $GL_n(\mathbb{C})$.

Definition 2.3. If G is a subgroup of $GL_n(\mathbb{C})$, then G is said to be a **Lie group** if it is a closed set with respect to the subspace topology of $GL_n(\mathbb{C})$. In other words, given any convergent sequence of matrices in G, its limit must either not be invertible or remain in G.

For example, the set of all matrices $n \times n$ complex matrices with determinant 1 is a Matrix Lie Group. This is because, along with being a group, the set can be represented as the pre-image of $\{1\}$ with respect to the determinant, and hence is closed. This set is denoted by $SL_n(\mathbb{C})$.

Of course $GL_n(\mathbb{C})$ is also a Matrix Lie group, along with $GL_n(\mathbb{R})$ and $SL_n(\mathbb{R})$, since matrices in $\mathbb{R}$ still follow the group axioms, and the fact that $\mathbb{R}$ is a closed subset of $\mathbb{C}$.

Suppose we have a vector space (either $\mathbb{R}^n$ or $\mathbb{C}^n$) with the following inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n.$$

The group of operators that preserve the above inner product is called the $n \times n$ orthogonal group of $\mathbb{R}$ (or $\mathbb{C}$), denoted by $O_n(\mathbb{R})$ and $O_n(\mathbb{C})$.

If we impose the additional condition that the determinant must be 1, we get the special orthogonal groups $SO_n(\mathbb{R})$ and $SO_n(\mathbb{C})$.

Another inner product, this time being only applicable to $\mathbb{C}^n$, is as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 \bar{y}_1 + x_2 \bar{y}_2 + \ldots + x_n \bar{y}_n.$$

The group of matrices that preserve this inner product is known as the unitary group, or $U(n)$. Imposing the determinant to be 1, we get the special unitary group $SU(n)$.

Topological properties such as compactness, connectedness, and simple connectedness apply to matrix Lie groups. Note that since we the norm for matrices we use is identical to that over C^n, and hence all connected sets are path-connected.

To conclude this section, we define matrix Lie group homomorphisms.

Definition 2.4. Let G and H be matrix Lie groups. A **matrix Lie group homomorphism** is a map $\Phi : G \to H$ such that

- Φ is a group homomorphism
- Φ is continuous.

A **matrix Lie group isomorphism** is a matrix Lie group homomorphism that is a group isomorphism and a homeomorphism.

3. The Matrix Exponential and Logarithm

Recall that for any complex number z, the exponential is defined as:

$$e^z = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \cdots .$$

The coefficients of z^n, according to the root test

$$\frac{1}{R} = \limsup_{n\to\infty} \sqrt{|a_n|}$$

infer that the radius of convergence of the exponential series is infinite, i.e. the series convergences for all $z \in \mathbb{C}$. By absolute convergence, the infinite series above, with z substituted with a complex matrix A, will also converge (and that too for all $A \in M)_n(\mathbb{C})$. Hence, we write

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots .$$

One of the earliest observations one can make is that e^A commutes with A, as e^A is a power series in A, and multiplication with A results in the exponents of the power series increasing by 1. Similarly, e^{sA} commutes with tA for all $s, t \in \mathbb{C}$.

Claim 3.1. *The function* exp *is continuous over* $M_n(\mathbb{C})$*.*

Proof. The sum

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

is uniformly convergent over $R = \|A\|$ by the Weierstrass M-test, as each summand can be bounded by a convergent infinite geometric series. Since each partial sum is continuous, it must follow that e^A is continuous in the disk whose center is at the origin and radius is R. Since R can get arbitrarily large, e^A is continuous over all of $\mathbb{C}$. ■

We now prove some basic properties of the matrix exponential.

Proposition 3.2. *Let* $X, Y \in M_n(\mathbb{C})$*. Then*

(1) $e^0 = I$
(2) e^X *is invertible for all* X*, and* $(e^X)^{-1} = e^{-X}$
(3) *if* X *and* Y *commute,* $e^{X+Y} = e^X e^Y = e^Y e^X$
(4) *For* $S \in GL_n(\mathbb{C})$*,* $e^{S^{-1}XS} = S^{-1}e^X S$*.*

Proof. Point (1) follows by plugging in $A = 0$ into the power series.

For point (3), note that since X and Y commute, we can mimic the proof of $e^{z_1+z_2} = e^{z_1}e^{z_2}$ in the case of $\mathbb{C}$, by using the Cauchy product formula. The reason that X and Y must commute is that when expanding the power series of e^{X+Y}, we would require the multiplicands of X and Y in one term to collect in a single exponent, which is not possible if X and Y don't commute since, for example, $XYX \neq X^2Y$. Hence, the Cauchy product formula proof cannot apply in the general case.

Point (2) follows from point (3) as

$$I = e^0 = e^{X+-X} = e^X e^{-X}.$$

In order to prove point (4), note that

$$(S^{-1}XS)^n = (S^{-1}XS)(S^{-1}XS)\cdots(S^{-1}XS) = S^{-1}X^nS,$$

for any $n \in \mathbb{N}$. Actually, we may conclude it to be true for $n \in \mathbb{Z}$ by using point (2) of this proposition. Thus

$$e^{S^{-1}XS} = \sum_{n=0}^{\infty} \frac{(S^{-1}XS)^n}{n!} = \sum_{n=0}^{\infty} \frac{S^{-1}X^nS}{n!} = S^{-1}[\sum_{n=0}^{\infty} \frac{X^n}{n!}]S = S^{-1}e^X S.$$

∎

Suppose $A \in M_n(\mathbb{C})$ is diagonalizable, meaning there exists an invertible matrix S and a diagonal matrix Λ such that $A = S^{-1}\Lambda S$. This set of diagonalizable matrices can be proven to be dense in $M_n(\mathbb{C})$.

As shown above, $e^A = S^{-1}e^\Lambda S$. This makes computing exp for diagonalizable matrices very convenient, as e^Λ is just the exponential of all its diagonal entries. Also, this representation implies that the eigenvalues of e^A are the exponential of the eigenvalues of A.

We now define the logarithm for matrices. Recall for complex numbers, we define $\log z$ by the Taylor series centered at $z = 1$:

$$\log z = \sum_{n=1}^{\infty} (-1)^{m-1} \frac{(z-1)^n}{n}.$$

By the root test, its radius of convergence is 1. Let

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

for a matrix A whenever it's convergent. Again by absolute convergence, the series above converges for all A such that $\|A - I\| < 1$.

Note that some A might still converge even when they don't fall into the aforementioned disk. For example, when $A - I$ is nilpotent, $\log A$ converges.

By a similar argument to the exponential case, $\log A$ is continuous over $\|A - I\| < 1$, as each summand can be bounded by an $\|A - I\|^n/n$, where $\|A - I\| < 1$. Since such a series converges, the Taylor series of $\log A$ will be continuous over the aforementioned disk.

We now prove that exp and log are local inverses of one another.

Lemma 3.3. (1) *For $z \in \mathbb{C}$ such that $|z - 1| < 1$,*

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

(2) *For $u \in \mathbb{C}$ such that $|u| < \log 2$, $|e^u - 1| < 1$, and*

$$\log(e^u) = u.$$

Proof. We have $\exp(\log z) = z$ for all positive reals z, specifically $z \in (0, 2)$. Since $1 \in (0, 2)$ is an accumulation point, the Identity theorem from complex analysis applies, and $\exp(\log z) = z$ for all z such that $|z - 1| < 1$.

Likewise, if $|u| < \log 2$, then

$$|e^u - 1| = |u + \frac{u^2}{2!} + \frac{u^3}{3!} + \cdots | \leq |u| + \frac{|u^2|}{2} + \cdots = e^{|u|} - 1 < 1.$$

Hence, $\log(\exp(u))$ is well-defined for $|u| < \log 2$. We have $\log(\exp(u)) = u$ for all $u \in (-\log 2, \log 2)$. Since $0 \in (-\log 2, \log 2)$ is an accumulation point, $\log(\exp(u)) = u$ for all $|u| < \log 2$. ∎

Proposition 3.4. (1) *For $A \in M_n(\mathbb{C})$ such that $\|A - I\| < 1$,*

$$e^{\log A} = A.$$

(2) *For $X \in M_n(\mathbb{C})$ such that $\|X\| < \log 2$, $\|e^X - I\| < 1$, and*

$$\log(e^X) = X.$$

Lemma 3.5. *If λ is an eigenvalue of a complex matrix X, then $|\lambda| \leq \|X\|$. In particular, If $\|A - 1\| < 1$ and $|X| < \log 2$, then*

$$|z - 1| < 1$$

and

$$|u| < \log 2$$

for all eigenvalues z of A and all eigenvalues u of X.

Proof. Let v be a unit eigenvector for the eigenvalue λ. We have,

$$|\lambda| = \|\lambda v\| = \|Xv\| = \frac{\|Xv\|}{\|v\|} \leq \|X\|.$$

■

Proof of Proposition 3.4. Suppose $A \in M_n(\mathbb{C})$ such that $\|A - I\| < 1$. This means $\log A$ is permissible. By the Lemma 3.3, $|z - 1| < 1$ for all eigenvalues z of A. If $A = S^{-1}\Lambda S$, then $A - I = S^{-1}\Lambda' S$, where Λ' is Λ with the main diagonal decreased by 1. This is because the eigenvalues of $A - I$ are one less than the eigenvalues of A, and the eigenvectors of A and $A - I$ are the same.

By all of this, we must have, similar to the exponential case,

$$\log(A) = S^{-1} \log(\Lambda) S,$$

where $\log(\Lambda)$ is the logarithm of all its non-zero entries. By Lemma 3.3, $e^{\log z} = z$ for $|z - 1| < 1$, and thus

$$e^{\log A} = S^{-1} e^{\log(\Lambda)} S = S^{-1}\Lambda S = A.$$

This proves the claim for all matrices as the set of diagonalizable matrices is dense in $M_n(\mathbb{C})$.

For $\|X\| < \log 2$, a similar argument from Lemma 3.3 shows that $\|e^X - I\| < 1$. By a similar argument as above, since we have $|u| < \log 2$, by Lemma 3.3, for all eigenvalues u of X, we must have $|e^u - 1| < 1$. If $X = S^{-1}\Lambda S$, then

$$e^X = S^{-1} e^{\Lambda} S,$$

and since $\|e^X - I\|$ and $|e^u - 1|$ are less than 1, log is well-defined for both the LHS and the RHS, and thus

$$\log(e^X) = S^{-1} \log(e^{\Lambda}) S = S^{-1}\Lambda S = X.$$

■

Note that a key ingredient used in the proof was how the logarithm and exponential functions acted on diagonal matrices.

Recall the asymptotic notation $f(x) = O(g(x))$ for $f, g : \mathbb{R} \to \mathbb{R}$, which is that there exists $M, x_0 \in \mathbb{R}$ such that $M > 0$ and

$$|f(x)| \leq M|g(x)|$$

for all $x \geq x_0$.

If $f : M_n(\mathbb{C}) \to M_n(\mathbb{C})$, we define $f(A) = O(g(\|A\|))$, where $g : \mathbb{R} \to \mathbb{R}$, if $\|f(A)\| = O(g(\|A\|))$. We now give an asymptotic identity of $\log A$.

Claim 3.6. *For all $A \in M_n(\mathbb{C})$ such that $\|A\| < \frac{1}{2}$,*

$$\|\log(I+A) - A\| = O(\|A\|^2).$$

Equivalently,

$$\log(I+A) = A + O(\|A\|^2) \tag{3.1}$$

Proof. We have

$$\log(I+A) - A = \sum_{n=2}^{\infty}(-1)^{n+1}\frac{A^n}{n} = A^2\sum_{n=2}^{\infty}(-1)^{n+1}\frac{A^{n-2}}{n}$$
$$\implies \|\log(I+A) - A\| \leq \|A\|^2\sum_{n=2}^{\infty}\frac{(\frac{1}{2})^{m-2}}{m},$$

and we're done as the geometric series converges, and hence $\|\log(I+A)-A\|$ is asymptotically bounded by $\|A\|^2$. ■

Note that the choice $\|A\| < \frac{1}{2}$ is arbitrary when we apply the above claim to to the next result.

We now state another important theorem regarding exp.

Theorem 3.7 (*Lie Product Formula*)**.** *For all $X, Y \in M_n(\mathbb{C})$, we have*

$$e^{X+Y} = \lim_{m\to\infty}(e^{\frac{X}{m}}e^{\frac{Y}{m}})^m.$$

Proof. By multiplying the power series of $e^{X/m}$ and $e^{Y/m}$, we see that all except three terms will be asymptotically bounded by $O(\frac{1}{m^2})$. Specifically,

$$e^{\frac{X}{m}}e^{\frac{Y}{m}} = I + \frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}).$$

As $m \to \infty$, $e^{X/m}e^{Y/m}$ gets sufficiently close to I, hence falling into the domain of the logarithm. Also, $\|\frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2})\| < \frac{1}{2}$ if m is sufficiently large. Thus we get (using Claim 3.6),

$$\log(e^{\frac{X}{m}}e^{\frac{Y}{m}}) = \log(I + \frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}))$$
$$= \frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}) + O(\|\frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2})\|^2)$$
$$= \frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}) + O(\frac{1}{m^2}) = \frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}).$$

Exponentiating the logarithm and tending m to ∞, we get

$$e^{\frac{X}{m}}e^{\frac{Y}{m}} = \exp(\frac{X}{m} + \frac{Y}{m} + O(\frac{1}{m^2}))$$
$$\implies \lim_{m\to\infty}(e^{\frac{X}{m}}e^{\frac{Y}{m}})^m = \lim_{m\to\infty}\exp(X + Y + O(\frac{1}{m})) = \exp(X+Y),$$

which is what was desired. ■

We now consider the differentiation and integration of matrix-valued functions. The derivative of a function $A : \mathbb{R} : M_n(\mathbb{C})$ is defined as

$$\left(\frac{dA}{dt}\right)_{ij} = \frac{A_{ij}}{dt}.$$

The linearity and product rule follow the usual proofs from the scalar case.

Theorem 3.8. *For $X \in M_n(\mathbb{C})$ e^{tX} is a smooth function, and*

$$\frac{d}{dt}e^{tX} = Xe^{tX} = e^{tX}X.$$

In particular,

$$\left.\frac{d}{dt}e^{tX}\right|_{t=0} = X.$$

Proof. If Λ is diagonal, then

$$\left(\frac{d}{dt}e^{t\Lambda}\right)_{ij} = \frac{d(e^{t\Lambda_{ij}})}{dt} = \Lambda_{ij}e^{t\Lambda_{ij}}.$$

Thus, the identity holds for diagonal matrices.

It suffices to prove the result for diagonalizable matrices X as they are dense in $M_n(\mathbb{C})$. If $X = S^{-1}\Lambda S$, then

$$\begin{aligned}\frac{d}{dt}e^{tX} &= S^{-1}\frac{d}{dt}e^{t\Lambda}S = S^{-1}\Lambda e^{t\Lambda}S \\ &= (S^{-1}\Lambda S)(S^{-1}e^{t\Lambda}S) = Xe^{tX}.\end{aligned}$$

■

The integral of a matrix-valued function is defined similarly as the derivative, by taking the integral of each entry. This will be used in the Baker-Campbell-Hausdorff formula.

To conclude this section, we define and prove a result on one-parameter subgroups, which will be useful later.

Definition 3.9. A continuous function $A : \mathbb{R} \to GL_n(\mathbb{C}$ is called a **one-parameter subgroup** of $GL_n(\mathbb{C})$ if

- $A(0) = I$
- $A(t + s) = A(t)A(s)$ for all $s, t \in \mathbb{R}$

Theorem 3.10 (Characterisation of One-Parameter Subgroups)**.** *If A is a one-parameter subgroup of $GL_n(\mathbb{C})$, then there exists a unique $X \in M_n(\mathbb{C})$ such that*

$$A(t) = e^{tX}$$

for all $t \in \mathbb{R}$.

Lemma 3.11. *Fix $\epsilon < \log 2$. Let $B_{\epsilon/2}$ be the open ball of radius $\epsilon/2$ centered at the origin, and let $U = \exp(B_{\epsilon/2})$. Then for every $B \in U$, there exists a unique $C \in U$ such that $C^2 = B$, and is given by $C = \exp(\frac{1}{2}\log B)$.*

Proof. Since $\|B - I\| < 1$, it's clear that $\exp(\frac{1}{2}\log B)$, and by Proposition 3.4, $C^2 = \exp(2 * \frac{1}{2}\log B) = \exp(\log B) = B$.

In order to establish uniqueness, suppose, for the sake of contradiction, $C' \in U$ such that $(C')^2 = B$. Let $Y = \log C'$, so that by Proposition 3.4 $Y \in B_{\epsilon/2}$, which in turn implies $2Y \in B_\epsilon$. We also have $\exp Y = (C')$ and hence

$$\exp(2Y) = (C')^2 = B = \exp(\log B)).$$

■

Note that $\log B \in B_{\epsilon/2} \subset B_\epsilon$. By point (2) of Proposition 3.4, exp is injective over B_ϵ, and because $\exp(2Y) = \exp(\log B)$, hence $2Y = B$. Hence

$$C' = \exp(Y) = \exp(\frac{1}{2}\log B) = C.$$

Proof of Theorem 3.10. The fact that X is unique, if it exists, is clear, as

$$X = \frac{d}{dt}A(t)\bigg|_{t=0}.$$

In order to prove existence, let $B_{\epsilon/2}$ and U be as described in the previous lemma. Note that U is an open neighbourhood of I, as it is the pre-image of B_ϵ on the function log. Hence, by A's continuity, there exists $t_0 > 0$, such that $A(t) \in U$ for all $t \in \mathbb{R}$ such that $|t| \leq t_0$. Let

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

so that $t_0 X = \log(A(t_0))$. This means that $t_0 X \in B_{\epsilon/2}$, and thus

$$e^{t_0 X} = A(t_0).$$

Now by the definition of t_0 $A(t_0/2) \in U$, with $A(t_0/2)^2 = A(t_0$ by the axioms of a one-parameter subgroup. By Lemma 3.11, $A(t_0)$ has a unique square root, given by $\exp(t_0 X/2)$. Thus, we have

$$A(\frac{t_0}{2}) = e^{\frac{t_0 X}{2}}.$$

By inductively applying this argument, we have that for all positive positive integers k, we get

$$A(\frac{t_0}{2^k}) = e^{\frac{t_0 X}{2^k}}.$$

Also, for all integers n,

$$A(\frac{t_0 X n}{2^k}) = (A(\frac{t_0 X}{n}))^n = e^{\frac{t_0 X n}{2^k}}.$$

Hence $A(t) = \exp(tX)$ for all $t = \frac{n}{2^k}t_0$, with $n \in \mathbb{Z}$ and $k \in \mathbb{N}$. Since all such t are dense in $\mathbb{R}$, and A along with exp are continuous, we can conclude that $A(t) = \exp(tX)$ for all $t \in \mathbb{R}$. ■

4. Directional Derivatives

The total derivative of a function $f : U \to \mathbb{C}^n$ (where U is an open subset of some $\mathbb{C}^m$) at a point $\mathbf{x} \in U$ is the unique complex linear transformation $J_\mathbf{x} : \mathbb{C}^m \to \mathbb{C}^n$ such that

$$\lim_{\mathbf{y}\to\mathbf{x}} \frac{\|f(\mathbf{y}) - f(\mathbf{x}) - J_\mathbf{x}(\mathbf{y}-\mathbf{x})\|}{\|\mathbf{y}-\mathbf{x}\|} = 0.$$

The above should hold regardless of how $\mathbf{y}$ approaches $\mathbf{x}$. The directional derivative of f with respect to a vector $\mathbf{v} \in \mathbb{C}^m$ is defined by the limit

$$\nabla_{\mathbf{v}} f(\mathbf{x}) = \lim_{h \to 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h}.$$

for $h \in \mathbb{R}$. If we set $\mathbf{y}(t) = \mathbf{x} + t\mathbf{v}$ $(t \in \mathbb{R})$, then we have

$$\begin{aligned} 0 &= \lim_{\mathbf{y} \to \mathbf{x}} \frac{\|f(\mathbf{y}) - f(\mathbf{x}) - J_{\mathbf{x}}(\mathbf{y} - \mathbf{x})\|}{\|\mathbf{y} - \mathbf{x}\|} \\ &= \lim_{t \to 0} \frac{\|f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x}) - J_{\mathbf{x}}(t\mathbf{v})\|}{\|tv\|} \\ &= \lim_{t \to 0} \frac{\|\frac{f(\mathbf{x}+t\mathbf{v}) - f(\mathbf{x})}{t} - J_{\mathbf{x}}(\mathbf{v})\|}{\|v\|} \\ &= \frac{\nabla_{\mathbf{v}} f(\mathbf{x}) - J_{\mathbf{x}}\mathbf{v}}{\|v\|}, \end{aligned}$$

hence $\nabla_{\mathbf{v}} f(\mathbf{x}) = J_{\mathbf{x}}\mathbf{v}$ for all $\mathbf{x} \in U$ and $\mathbf{v} \in \mathbb{C}^m$.

We use the directional derivative in the proofs for a couple of crucial theorems, including the Baker-Campbell-Hausdorff Formula.

5. Lie Algebra

Definition 5.1. A **Lie algebra** $\mathfrak{g}$ is a real (or complex) vector space together with a billinear symmetric map $[\cdot] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ that satisfies the *Jacob identity*, meaning

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

for all $X, Y, Z \in \mathfrak{g}$.
A **subalgebra** $\mathfrak{h} \subseteq g$ is a subspace such that $[h_1.h_2] \in H$ for all $h_1, h_2 \in H$.'

For this paper, we only consider real Lie algebras.

One example of a Lie algebra is $\mathbb{R}^3$ with the cross product. The validity of the axioms can be verified, but won't be relevant for this paper.

The more important example of a Lie algebra is that of $M_n(\mathbb{C})$ with an operation known as the **commutator**:

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

The reason for the name commutator is because $[\cdot]$ equals 0 if and only if $XY = YX$, i.e. X and Y commute.

We shall denote the above Lie algebra as $\mathfrak{gl}_n(\mathbb{C})$, whose naming shall become clear in due course.

Note that $\mathfrak{gl}_n(\mathbb{C})$ can be interpreted as either a real or complex Lie algebra, which are distinct. In this paper, we'll consider the real Lie algebra.

Suppose $\mathfrak{sl}_n(\mathbb{C})$ (whose naming will also be explained) is the set of $n \times n$ complex matrices with trace zero. Then it can be checked that $\mathfrak{sl}_n(\mathbb{C})$ along with the commutator as a Lie bracket is a real Lie algebra.

Definition 5.2. If $\mathfrak{g}$ and $\mathfrak{h}$ are real (or complex) Lie algebras, then a linear map $\phi : \mathfrak{g} \to \mathfrak{h}$ is called a **Lie algebra homomorphism** if $[\phi(X), \phi(Y)] = \phi([X, Y])$ for all $X, Y \in \mathfrak{g}$. If the linear map is also invertible, then ϕ is a **Lie algebra isomorphism**.

Definition 5.3. If $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are real (or complex) Lie algebras, then define the **direct sum Lie algebra** $\mathfrak{g}$ of $\mathfrak{g}_1$ and $\mathfrak{g}_2$ to be the vector direct sum of $\mathfrak{g}_1$ and $\mathfrak{g}_2$, along with the bracket given by:

$$[(X_1, X_2), (Y_1, Y_2)] = ([X_1, Y_1], [X_2, Y_2]), \tag{5.1}$$

for $X_1, Y_1 \in \mathfrak{g}_1$ and $X_2, Y_2 \in \mathfrak{g}_2$. We denote the direct sum as $\mathfrak{g}_1 \oplus \mathfrak{g}_2$.

It can be shown that 5.1 adheres to the Lie algebra axioms.

Definition 5.4. If $\mathfrak{g}$ is a Lie algebra, and $\mathfrak{g}_1$ and $\mathfrak{g}_2$ are Lie subalgebra of $\mathfrak{g}$, then $\mathfrak{g}$ **decomposes as the Lie algebra direct sum** of $\mathfrak{g}_1$ and $\mathfrak{g}_2$ if $\mathfrak{g}$ is the vector space direct sum of $\mathfrak{g}_1$ and $\mathfrak{g}_2$, and $[X_1, X_2] = 0$ for all $X_1 \in \mathfrak{g}_1$ and $X_2 \in \mathfrak{g}_2$.

Claim 5.5. *If $\mathfrak{g}$ decomposes as a Lie algebra direct sum of $\mathfrak{g}$, then the direct sum $\mathfrak{g}_1$ and $\mathfrak{g}_2$ is isomorphic to $\mathfrak{g}$.*

Proof. Since $\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_2$, every $A \in \mathfrak{g}$ can be expressed as $A = X + Y$, where $X \in \mathfrak{g}_1$ and $Y \in \mathfrak{g}_2$. Since $X \in \mathfrak{g}_1$ and $Y \in \mathfrak{g}_2$ commute with one another,

$$\begin{aligned}[A, B] &= [X_1 + Y_1, X_2 + Y_2] \\ &= [X_1, X_2] + [X_1, Y_2] + [Y_1, X_2] + [Y_1, Y_2] = [X_1, X_2] + [Y_1, Y_2]\end{aligned} \tag{5.2}$$

for all $A, B \in \mathfrak{g}$. The equation 5.2 is essentially identical to 5.1, and hence $\phi([X, Y]) = X + Y$ provides the required Lie algebra isomorphism. ∎

We now explain the connection between matrix Lie groups and Lie algebras.

Definition 5.6. Let G be a matrix Lie group. The **Lie algebra** of G, denoted by $\mathfrak{g}$, the set of all complex matrices (not necessarily invertible) X such that $e^{tX} \in G$ for all $t \in \mathbb{R}$.

Another way of formulating the above definition is that the entire the one-parameter subgroup generated by X lies in $\mathfrak{g}$.

This is the explanation behind the notation $\mathfrak{gl}_n(\mathbb{C})$ and $\mathfrak{sl}_n(\mathbb{C})$. Note that in physics, the definition demands e^{itX} to be in G, rather than e^{tX}, causing the formulations of some Lie groups to be off by a factor of i.

It is possible to define the Lie algebra for a general Lie group (as the tangent space of G at the identity), though that won't be relevant here.

Proposition 5.7. *Let G be a matrix Lie group, with Lie algebra $\mathfrak{g}$. For all $X, Y \in \mathfrak{g}$, $A \in G$, and $s \in \mathbb{R}$, we have*

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

Thus, $\mathfrak{g}$ is a real Lie algebra in the way defined earlier.

Proof. (1) Recalling point (4) from Proposition 3.2,

$$e^{t(A^{-1}XA)} = A^{-1}e^{tX}A \in G$$

for all $t \in \mathbb{R}$, as all three of A^{-1}, e^{tX}, and A are in G. Hence $A^{-1}XA \in \mathfrak{g}$.

(2) Observe that $e^{t(sX)} = e^{tsX} \in G$ for all t, thus showing $sX \in \mathfrak{g}$.
(3) In order to prove this point, we use the Lie product formula:

$$e^{t(X+Y)} = \lim_{n\to\infty} (e^{tX/n} e^{tY/n})^n.$$

Now $(e^{tX/n} e^{tY/n})^n \in G$ for all n, and since G is closed in $GL_n(\mathbb{C})$, the limit $e^{t(X+Y)}$ is either in G or isn't invertible. Due to point (2) in Proposition 3.2, we know $e^{t(X+Y)}$ can't be invertible, and hence $e^{t(X+Y)} \in G$.
(4) By the product rule and Theorem 3.8,

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

Now

$$XY - YX = \lim_{h\to 0} \frac{e^{hX} Y e^{-hX} - Y}{h}.$$

for $h \in \mathbb{R}$. By point (1) of this current result, $e^{hX} Y e^{-hX} \in \mathfrak{g}$. Since we have already proven that $\mathfrak{g}$ is a real vector space, the LHS of the above limit is in $\mathfrak{g}$ for all h. Since $\mathfrak{g}$, being a vector space, is a closed set, the limit $XY - YX$ must be in $\mathfrak{g}$.

■

We now prove a converse to our main result (Theorem 1.1), though this result is more generally applicable.

Theorem 5.8. *Let G and H be matrix Lie groups, with corresponding Lie algebra $\mathfrak{g}$ and $\mathfrak{h}$ respectively. If $\Phi : G \to H$ if a matrix Lie group homomorphism, then there exists a Lie algebra homomorphism $\phi : \mathfrak{g} \to \mathfrak{h}$ such that*

$$\Phi(e^X) = e^{\phi(X)}$$

for all $X \in \mathfrak{g}$. Additionally for all $X, Y \in \mathfrak{G}$ and $A \in G$,
(1) $\phi(A^{-1}XA) = \Phi(A^{-1})\phi(X)\Phi(A)$
(2) $\phi([X,Y]) = [\phi(X), \phi(Y)]$
(3) $\phi(X) = \frac{d}{dt}\Phi(e^{tX})\Big|_{t=0}$.

Proof. Since Φ is a continuous function and a homomorphism, $\Phi(e^{tX})$ is a one-parameter subgroup of $GL_n(\mathbb{C})$. Hence by Theorem 3.10, there exists a unique matrix Z such that

$$\Phi(e^{tX}) = e^{tZ} \tag{5.3}$$

for all $t \in \mathbb{R}$. We define $\phi(X) = Z$. If we substitute $t = 1$ in equation , we easily see that $\Phi(e^X) = e^Z = e^{\phi(X)}$.

If another map $\phi' : \mathfrak{g} \to \mathfrak{h}$ existed satisfying $\Phi(e^X) = e^{\phi'(X)}$ for all $X \in \mathfrak{g}$, then

$$e^{t\phi(X)} = e^{t\phi(X')} = \Phi(e^{tX}).$$

Differentiating at $t = 0$ shows that $\phi(X) = \phi'(X)$. We now prove linearity. Since $\Phi(e^{tX}) = e^{t\phi(X)}$ for all $t \in \mathbb{R}$, we get

$$e^{s[t\phi(X)]} = \Phi(e^{tsX}) = e^{t\phi(sX)}$$

for all $s, t \in \mathbb{R}$. Differentiating at $t = 0$ yields $s\phi(X) = \phi(sX)$.

$$e^{t\phi(X+Y)} = \Phi(\lim_{n\to\infty}(e^{\frac{tX}{n}} e^{\frac{tY}{n}})^n)$$
$$= \lim_{n\to\infty}(\Phi(e^{\frac{tX}{n}} e^{\frac{tY}{n}}))^n = \lim_{n\to\infty}(\Phi(e^{\frac{tX}{n}})\Phi(e^{\frac{tY}{n}}))^n$$
$$= \lim_{n\to\infty}(e^{\frac{t\phi(X)}{n}} e^{\frac{t\phi(Y)}{n}})^n = e^{t(\phi(X)+\phi(Y))}.$$

Differentiating the above result at $t = 0$ proves $\phi(X+Y) = \phi(X) + \phi(Y)$. We now prove the remaining properties (1), (2), and (3) of ϕ.

(1) If $A \in G$, then
$$e^{t\phi(A^{-1}XA)} = e^{\phi(tA^{-1}XA)}$$
$$= \Phi(e^{tA^{-1}XA}) = \Phi(A^{-1})\Phi(e^{tX})\Phi(A) = (\Phi(A)))^{-1}e^{\phi(X)}\Phi(A).$$

(2) As in Proposition 5.7, we have (using the fact that a linear transformation commutes with the derivative, along with (1))
$$\phi(X, Y) = \phi\left(\frac{d}{dt}e^{tX}Ye^{-tX}\bigg|_{t=0}\right)$$
$$= \frac{d}{dt}\phi(e^{tX}Ye^{-tX})\bigg|_{t=0} = \frac{d}{dt}\Phi(e^{tX}\phi(Y)\Phi(e^{tX})\bigg|_{t=0}$$
$$= \frac{d}{dt}e^{t\phi(X)}\phi(Y)e^{-t\phi(X)}\bigg|_{t=0} = [\phi(X), \phi(Y)].$$

(3) Since
$$\Phi(e^{tX}) = e^{\phi(tX)} = e^{t\phi(X)}$$
and
$$\frac{d}{dt}e^{t\phi(X)}\bigg|_{t=0} = \phi(X),$$
we get (3).

■

This theorem shows that a unique Lie group homomorphism gives rise to a unique Lie algebra homomorphism. The converse, in general, not true, however we later use the Baker-Campbell-Hausdorff formula to prove that the converse is true if we assume G is simply-connected.

Next, we define some important maps in Lie theory and for the proof of the BCH formula.

Definition 5.9 (*The Adjoint Map*)**.** Let G be a matrix Lie group and $\mathfrak{g}$ its Lie algebra. For $A \in G$, define the **adjoint map** $\mathrm{Ad}_A : \mathfrak{g} \to \mathfrak{g}$
$$\mathrm{Ad}_A(X) = AXA^{-1}.$$

In what follows, $GL(\mathfrak{g})$ be the space of invertible linear operators of $\mathfrak{g}$, with $\mathfrak{gl}(\mathfrak{g})$ as its associated Lie algebra. Equivalently, $GL(\mathfrak{g})$ can be identified with $GL_m(\mathbb{R})$, and $\mathfrak{gl}(\mathfrak{g})$ can be identified as $M_m(\mathbb{R})$, where m is the vector space dimension of $\mathfrak{g}$.

Proposition 5.10. (1) *For all $A \in G$, $Ad_A \in GL(\mathfrak{g})$.*

(2) *The map $Ad : G \to GL(\mathfrak{g})$ is a Lie group homomorphism.*

(3) *The map Ad_A satisfies $Ad_A([X, Y]) = [Ad_A(X), Ad_A(Y)]$ for all $X, Y \in \mathfrak{g}$.*

Proof. (1) We have $\mathrm{Ad}_A(tX) = A(tX)A^{-1} = t(AXA^{-1}) = t\mathrm{Ad}_A(X)$.
Also, $\mathrm{Ad}_A(X+Y) = A(X+Y)A^{-1} = AXA^{-1} + AYA^{-1} = \mathrm{Ad}_A(X) + \mathrm{Ad}_A(Y)$. This shows that Ad_A is linear.
To show that it's invertible, consider $\mathrm{Ad}_{A^{-1}}$.
Then $(\mathrm{Ad}_{A^{-1}} \circ \mathrm{Ad}_A)(X) = A^{-1}(AXA^{-1})A = X$ and by symmetry composition the other way around gives X. Hence $\mathrm{Ad}_A \in GL(\mathfrak{g})$

(2) Note that $\mathrm{Ad}_A(X)$ sends X to a vector of n^2 polynomials in x_i, with coefficients being a rational function in A. Thus, if Ad_A and Ad_B are sufficiently close, then the coefficients of the polynomial entries is sufficiently close, which means the entries of A and B themselves must be sufficiently close, imply Ad is continuous.
We also have $\mathrm{Ad}_{AB}(X) = ABX(AB)^{-1} = A(BXB^1)A^{-1} = (\mathrm{Ad}_A \circ \mathrm{Ad}_B)$, hence Φ is a homomorphism.
Hence, Φ, being both continuous and a homomorphism, is a Lie group homomorphism.

(3) We have

$$\begin{aligned}\mathrm{Ad}_A([X,Y]) &= A(XY - YX)A^{-1} = AXYA^{-1} - AYXA^{-1} \\ &= AXA^{-1}AYA^{-1} - AYA^{-1}AXA^{-1} = [\mathrm{Ad}_A(X), \mathrm{Ad}_A(Y)].\end{aligned}$$

■

By Theorem 5.8, there exists a Lie algebra homomorphism $\mathrm{ad} : \mathfrak{g} \to \mathfrak{gl}(\mathfrak{g}$ such that

$$e^{\mathrm{ad}_X} = \mathrm{Ad}_{e^X}$$

for all $X \in \mathfrak{g}$. We shall come to what e^{ad_X} means (see 5.4). For now, we see ad_X as an element of $M_m(\mathbb{R})$.

Proposition 5.11. *For all $X, Y \in \mathfrak{g}$,*

$$ad_X(Y) = [X, Y].$$

Proof. By point (3) of Theorem 5.8,

$$\mathrm{ad}_X = \frac{d}{dt}\mathrm{Ad}_{e^{tX}}\bigg|_{t=0}.$$

Hence,

$$\mathrm{ad}_X(Y) = \frac{d}{dt}\mathrm{Ad}_{e^{tX}}(Y)\bigg|_{t=0} = \frac{d}{dt}e^{tX}Ye^{-tX}\bigg|_{t=0} = [X, Y].$$

■

We now revisit trying to understand the meaning of e^{ad_X}. Observe that

$$e^{\mathrm{ad}_X} = I + \mathrm{ad}_X + \frac{\mathrm{ad}_X^2}{2!} + \cdots$$

where ad_X^n is ad composed with itself n times. We now have

$$\begin{aligned}e^{\mathrm{ad}_X}(Y) &= I(Y) + \mathrm{ad}_X(Y) + \frac{\mathrm{ad}_X^2}{2!}(Y) + \cdots \\ &= Y + [X,Y] + \frac{[X[X,Y]]}{2!} + \cdots.\end{aligned} \tag{5.4}$$

We will now interpret e^{ad_X} as a linear operator on $\mathfrak{g}$. Actually $e^{\mathrm{ad}_X} \in GL(\mathfrak{g})$, since e^X is invertible for any linear operator X.

We now show that exp and log are local inverses in the context of Lie theory. If $X \in \mathfrak{g}$, then $\exp(X)$ is in G by definition. Hence we can consider and analyse the map $\exp : \mathfrak{g} \to G$.

Lemma 5.12. *Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Suppose B_n ($n \in \mathbb{N}$) are elements of G such that $B_n \to I$. Let $Y_n = \log B_n$, which is defined for sufficiently large n. If Y_n is non-zero for all n where it's defined, and $Y_n/\|Y_n\| \to Y$, then $Y \in \mathfrak{g}$.*

Proof. For all $t \in \mathbb{R}$, $\frac{tY_n}{\|Y_n\|} \to tY$. Since $B_n \to I$, we must also have $Y_n \to 0$ and hence $\|Y_n\| \to 0$. Hence for a fixed t,

$$\left\{\frac{t}{\|Y_n\|}\right\}\|Y_n\| \to 0$$

where $\{\cdot\}$ is the fractional part of a real number. This is because the expression inside the $\{\cdot\}$ above is bounded in the interval $[0, 1)$. We now have

$$\begin{aligned} \frac{t}{\|Y_n\|}\|Y_n\| - \left\{\frac{t}{\|Y_n\|}\right\}\|Y_n\| &\to t \\ \implies \quad (\frac{t}{\|Y_n\|} - \left\{\frac{t}{\|Y_n\|}\right\})\|Y_n\| &\to t \\ \implies \quad \left\lfloor\frac{t}{\|Y_n\|}\right\rfloor\|Y_n\| &\to t. \end{aligned}$$

where $\lfloor\cdot\rfloor$ is the floor (or greatest integer) function, which is always an integer. Thus we define the integers $a_n(t)$ as

$$a_n(t) = \left\lfloor\frac{t}{\|Y_n\|}\right\rfloor$$

for sufficiently large n (since Y_n is defined for sufficiently large n). Thus,

$$e^{a_n(t)Y_n} = \exp\left[(a_n(t)\|Y_n\|)\frac{Y_n}{\|Y_n\|}\right] \to e^{tY}.$$

On the other hand,

$$e^{a_n(t)Y_n} = (e^{Y_n})^{a_n(t)} = (B_n)^{a_n(t)} \in G$$

for all $t \in \mathbb{R}$, as $a_n(t)$ are integers. Since G is closed and e^{tY} is invertible, it must be the case that $e^{tY} \in G$ for all $t \in \mathbb{R}$, which implies $Y \in \mathfrak{g}$. ■

Theorem 5.13. *For $0 < \epsilon < \log 2$, let $U_\epsilon = \{X \in M_n(\mathbb{C}) | \|X\| < \epsilon\}$, and let $V_\epsilon = \exp(U_\epsilon)$. Let G be a matrix Lie group with Lie algebra $\mathfrak{g}$. Then there exists $\epsilon \in (0, \log 2)$ such that for any $A \in V_\epsilon$, we have $A \in G$ if and only if $\log A \in \mathfrak{g}$.*

Note that if $\log A \in \mathfrak{g}$, and A is sufficiently close to I, $e^{\log A} = A \in G$ (by the definition of $\mathfrak{g}$. So it suffices to prove only one direction of the above theorem, namely $A \in V_\epsilon \cap G \implies \log A \in \mathfrak{g}$.

Proof. We currently view $\mathfrak{g}$ as a real vector subspace of $M_n(\mathbb{C})$ (which can also be viewed as the real vector space $\mathbb{R}^{2n^2}$). We can define a symmetric inner product on $\mathbb{R}^{2n^2}$, like the one we defined in order to describe O_n and SO_n. Let S be the orthogonal complement of $\mathfrak{g}$ with respect to this inner product. Thus every $n \times n$ Z matrix can be written as $Z = X + Y$, where $X \in \mathfrak{g}$ and $Y \in S$.

Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be defined as

$$\Phi(Z) = e^X e^Y.$$

We interpret the above map over $\mathbb{R}^{2n^2}$ now. We have $X = \text{proj}_{\mathfrak{g}}(Z)$, and $Y = \text{proj}_S(Z)$, where proj is the projection map, which is linear and hence continuously differentiable. Since

$$\Phi(Z) = e^{\text{proj}_{\mathfrak{g}}(Z)} e^{\text{proj}_S(Z)},$$

we can conclude that Φ is continuously differentiable. We now calculate directional derivatives of Φ at $Z = 0$ in the direction of X and Y (which can be viewed as input vectors). We have

$$\nabla_X \Phi(0) = \lim_{h \to 0} \frac{e^{hX} - I}{h} = \frac{d}{dt}\Phi(tX, 0)\Big|_{t=0} = X. \tag{5.5}$$

We used the fact that the D component of X is 0 as $X \in \mathfrak{g}$. Similarly, $\nabla_Y \Phi(0) = Y$.

Since $X = \nabla_X \Phi(0) = J_0 X$ and $\nabla_Y \Phi(Y) = J_0 Y$, we have

$$\nabla_Z \Phi(0) = J_0 Z = J_0(X + Y) = J_0 X = J_0 Y = X + Y = Z$$

for all $Z \in M_n(\mathbb{C})$. This shows that J_0, or the total derivative of Φ at $Z = 0$ is the identity. Since the identity is invertible, the inverse function tells us that Φ is a isomorphism from a neighbourhood of 0 to a neighbourhood of I. Hence there exists a local inverse of Φ in a neighbourhood of I.

If there doesn't exists $\epsilon \in (0, \log 2)$ such that for all $A \in V_\epsilon \cap G \implies \log A \in \mathfrak{g}$, then there must exists a sequence of matrices A_n in G such that $A_n \to I$ and $\log A_n \notin \mathfrak{g}$ for all n. By using the local inverse of Φ, we may write A_n for sufficiently large n as

$$A_n = e^{X_n} e^{Y_n}$$

where $X_n \in \mathfrak{g}$ and $Y_n \in S$ for all n. Also, since $A_n \to I$, $X_n, Y_n \to 0$, it must be that $Y_n \neq 0$ for any n, as then

$$A_n = e^{X_n} \implies \log A_n = X_n \in \mathfrak{g},$$

which is a contradiction. Note that X_n, which is the projection of A_n onto $\mathfrak{g}$, will be sufficiently close to 0, and hence $\log(e^{X_n}) = X_n$. Since $X_n \in \mathfrak{g}$ and $A_n \in G$, we get that

$$B_n = e^{-X_n} A_m = e^{Y_n}$$

is in G. Since Y_n and B_n are sufficiently close to 0 and I respectively,

$$\log B_n = Y_n.$$

This implies that $B_n \to I$. Consider the unit sphere in the subspace S. This set is compact, and thus there exists a subsequence of Y_n such that

$$\frac{Y_n}{\|Y_n\|} \to Y$$

for some $Y \in S$ such that Y is in the unit sphere of S, which implies $\|Y\| = 1$. By Lemma 5, $Y \in \mathfrak{g}$. But $\mathfrak{g}$ and S are orthogonal complements, thus $Y = 0$, which contradicts the conclusion we derived above that $\|Y\| = 1$. ∎

This remarkably powerful theorem has some major consequences, particularly for connected matrix Lie groups.

Corollary 5.14. *If G is a connected matrix Lie group, then for every $A \in G$, there exists $X_1, X_2, \dots, X_n \in \mathfrak{g}$ such that*

$$A = e^{X_1}e^{X_2}\dots e^{X_n}.$$

Lemma 5.15. *Suppose $a \leq b \in \mathbb{R}$, and $f : [a,b]^n \to GL_n(\mathbb{C})$ is continuous. Then for all $\epsilon > 0$ there exists $\delta > 0$ such that for all $\mathbf{s}, \mathbf{t} \in [a,b]^n$ with $\|\mathbf{s} - \mathbf{t}\| < \delta$,*

$$\|f(\mathbf{s})f(\mathbf{t})^{-1} - I\| < \epsilon.$$

Proof. Observe that

$$\|f(\mathbf{s})f(\mathbf{t})^{-1} - I\| = \|(f(\mathbf{s}) - f(\mathbf{t}))f(\mathbf{t})^{-1}\| \leq \|f(\mathbf{s}) - f(\mathbf{t})\|\|f(\mathbf{t})^{-1}\|, \tag{5.6}$$

where we used the sub-multiplicativity of $\|\cdot\|$. Since $[a,b]^n$ is compact, $\|f(\mathbf{t})^{-1}\|$ is bounded above by a constant, say by $C > 0$. Another consequence of $[a,b]^n$ being compact is that f is uniformly continuous, which implies for any $\epsilon > 0$ there exist $\delta > 0$ such that for all $s, t \in [a,b]$ with $\|\mathbf{s} - \mathbf{t}\| < \delta$, $\|f(\mathbf{s}) - f(\mathbf{t})\| < \epsilon/C$. Multiplying this inequality with $\|f(\mathbf{t})^{-1}\| < C$, we get our desired result. ■

Proof of Corollary 5.14. Let V_ϵ be as defined as Theorem 5.13. For any $A \in G$, let $f : [0,1] \to G$ be a continuous path such that $f(0) = I$ and $f(1) = A$. By Lemma 5.15, there exists $\delta < 0$ such that if $|s - t| < \delta$, then $f(s)f(t)^{-1} \in V_\epsilon$. We partition $[0,1]$ into n pieces, such that $\frac{1}{n} < \delta$. Thus, for $i = 1, 2, \dots, n$, we have $f(\frac{i-1}{n})^{-1}f(\frac{i}{n}) \in V_\epsilon$, which implies

$$f(\frac{i-1}{n})^{-1}f(\frac{i}{n}) = e^{X_i}$$

for all i and some $X_1, X_2, \dots X_n$, which follows by Theorem 5.13. Hence

$$\begin{aligned} A &= f(0)^{-1}f(1) \\ &= f(0)^{-1}f(\frac{1}{m})f(\frac{1}{m})^{-1}f(\frac{2}{m})\cdots f(\frac{m-1}{m})^{-1}f(1) \\ &= e^{X_1}e^{X_2}\cdots e^{X_n}. \end{aligned}$$

■

6. The Baker-Campbell-Hausdorff Formula

Consider the complex function

$$g(z) = \frac{\log z}{1 - \frac{1}{z}}.$$

g is analytic centred on $z = 1$, and has a radius of convergence 1 since the closest pole of g to 1 is 0. If

$$g(z) = \sum_{n=0}^{\infty} a_m(z-1)^m$$

for some $a_m \in \mathbb{C}$, then by absolute convergence, the above can series can be defined to any matrix $A \in M_n(\mathbb{C})$, such that $\|A - I\| < 1$. This can be generalized to any linear operator of a finite-dimensional complex vector space V, by identifying V with $\mathbb{C}^n$ and hence defining the norm of linear operator on V. To summarise, for any linear operator A of a finite-dimensional complex vector space V,

$$g(A) = \frac{\log A}{I - A^{-1}} = \sum_{n=0}^{\infty} a_m(A - I)^m \tag{6.1}$$

is well-defined provided $\|A - I\| = 1$.

We shall now state and prove the Baker-Campbell-Hausdorff Formula, which is the biggest tool that will be used to prove Theorem 1.1.

Theorem 6.1 (*BCH Formula*). *For all $X, Y \in M_n(\mathbb{C})$ with $\|X\|$ and $\|Y\|$ sufficiently small, we must have*

$$\log(e^X e^Y) = X + \int_0^1 g(e^{ad_X} e^{t ad_Y})(Y) dt.$$

Consider the derivative

$$\Delta(X, Y) = \frac{d}{dt} e^{X+tY} \Big|_{t=0}$$

where $X, Y \in M_n(\mathbb{C})$ and $t \in \mathbb{R}$.

Also consider the directional derivative of $\exp : \mathbb{C}^{n^2} \to \mathbb{C}^{n^2}$ in the direction a complex $n \times n$ complex matrix Y acting as a n^2-dimensional vector.

We have:

$$\nabla_Y \exp(X) = \lim_{t \to 0} \frac{e^{X+tY} - e^X}{t} = \Delta(X, Y). \tag{6.2}$$

Therefore, we must also have $\Delta(X, Y) = J_X Y$, where J_X is a linear operator on $\mathbb{C}^{n^2}$. Hence

for fixed X, $\Delta(X, Y)$ is a linear function on Y.

Theorem 6.2 (*Derivative of* exp). *For all $X, Y \in M_n(\mathbb{C}^n)$, we have*

$$\begin{aligned}\Delta(X, Y) &= e^X \left\{ \frac{I - e^{ad_X}}{ad_X}(Y) \right\} \\ &= e^X \left\{ Y - \frac{[X, Y]}{2!} + \frac{[X, [X, Y]]}{3! -} - \cdots \right\}.\end{aligned}$$

Note that ad_X need not be invertible, by $\frac{I - e^{\mathrm{ad}_X}}{\mathrm{ad}_X}$ we mean the power series expansion of the expression, which is

$$\frac{I - e^{\mathrm{ad}_X}}{\mathrm{ad}_X} = I + \frac{\mathrm{ad}_X}{2!} + \frac{\mathrm{ad}_X^2}{3!} + \cdots .$$

Lemma 6.3. *If Z is a linear operator on a complex finite-dimensional vector space, then*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n} (e^{-Z/m})^n = \frac{I - e^{-Z}}{Z} \tag{6.3}$$

Similar to above, the RHS is to be interpreted as a power series so that it's well-defined for all operators Z.

Proof. We have

$$\begin{aligned}\frac{1}{n} \lim_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} (e^{-Z/n})^m &= \lim_{n \to \infty} \frac{1}{n} \frac{I - e^{-Z}}{I - e^{-Z/n}} \\ &= \lim_{n \to \infty} \frac{I - e^{-Z}}{n\frac{1}{n} Z - n \frac{1}{2!} \frac{1}{n^2} Z^2 + \cdots} \\ &= \frac{I - e^{-Z}}{Z}\end{aligned}$$

for all operators Z with $\det(Z) \neq 0$ and none of Z's eigenvalues are 0 (the second condition as $I - e^{-Z/n}$ is invertible if and only if Z's eigenvalues are all non-zero). Actually, the first condition is enough as $\det(Z) = 0$ if and only if Z has a 0 eigenvalue.

Since the set of all operators with determinant not equal to 0 is dense in M_n, we see that 6.3 is true for all operators (when viewed as a power series). ∎

Proof of Theorem 6.2. For every positive integer n, we have

$$e^{X+tY} = \left[\exp(\frac{X}{n} + \frac{tY}{n})\right]^n.$$

$\Delta(X,Y)$ is the derivative of the above expression at $t = 0$. By the product rule, we have:

$$\begin{aligned}\Delta(X,Y) &= \sum_{m=0}^{n-1}(e^{X/n})^{n-m-1}[\frac{d}{dt}\exp(\frac{X}{n} + t\frac{Y}{n})](e^{(X/n})^m \\ &= e^{(n-1)X/n}\sum_{m=0}^{n-1}(e^{X/n})^{-m}\Delta(\frac{X}{n},\frac{Y}{n})(e^{(X/n)})^m \\ &= e^{(n-1)X/n}\frac{1}{n}\sum_{m=0}^{n-1}\mathrm{Ad}_{(e^{(X/n})^{-m}}(\Delta(\frac{X}{n},Y)) \\ &= e^{(n-1)X/n}\frac{1}{n}\sum_{m=0}^{n-1}[\exp(\frac{-\mathrm{ad}_X}{n})]^m(\Delta(\frac{X}{n},Y)).\end{aligned}$$

Note that we also used to linearity with respect to Y above.

If we tend n to ∞ in the final expression, we see that $e^{(n-1)X/n}$ tends to e^X, $\Delta(\frac{X}{n},Y)$ tends to $\Delta(0,Y) = Y$. Lastly,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{m=0}^{n-1}[\exp(\frac{-\mathrm{ad}_X}{n})]^m = \frac{1-e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}.$$

Hence we have shown that

$$\left.\frac{d}{dt}e^{X+tY}\right|_{t=0} = e^X\left\{\frac{1-e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}(Y)\right\}.$$

∎

This means that the total derivative (or Jacobian) of exp at X equals

$$J_X = e^X\left\{\frac{1-e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}\right\}.$$

Let U be an open subset of $\mathbb{R}$. Suppose $X(t) : U \to M_n(\mathbb{C})$ is a smooth matrix-valued function. By the chain rule, we must have

$$\frac{d}{dt}e^{X(t)} = J_{X(t)}\left.\left(\frac{dX}{dt}\right)\right|_{t=0} = e^{X(t)}\left\{\frac{1-e^{-\mathrm{ad}_{X(t)}}}{\mathrm{ad}_{X(t)}}\right\}\left(\frac{dX}{dt}\right)$$

for all real scalar complex matrix-valued functions X.

Recall that $\mathrm{ad}_X(Y) = [X,Y]$ for a complex matrix Y. Hence, $\mathrm{ad}_X^n(Y)$ equals the nested bracket expression $[X, [X, \ldots [X, Y]] \ldots]$, where there are n Xs. Thus

$$\frac{1 - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}.$$

is a series of nested brackets, as shown in 6.2.

If X is sufficiently small, then $\mathrm{ad}_X(Y) = XY - YZ$ is sufficiently close to the zero operator on $M_n(\mathbb{C})$, and thus

$$\frac{1 - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X} = I - \frac{\mathrm{ad}_x}{2!} + \frac{\mathrm{ad}_X^2}{3!} - \ldots \tag{6.4}$$

approaches I as X approaches 0. Specifically, for sufficiently small X, 6.4 will be so close to I so that its determinant cannot be zero (since $\det I = 1$), and hence will be invertible.

Proof of Theorem 6.1. For sufficiently small $X, Y \in M_n(\mathbb{C})$, let

$$Z(t) = \log(e^X e^{tY}) \tag{6.5}$$

so that $Z(t)$ is defined in $[0, 1]$. By the generalization of Lemma 6.2, we have

$$e^{-Z(t)} \frac{d}{dt} e^{Z(t)} = \left\{ \frac{I - e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}} \right\} \left(\frac{dZ}{dt} \right).$$

Since $e^{Z(t)} = e^X e^{tY}$, we must also have have

$$e^{-Z(t)} \frac{d}{dt} e^{Z(t)} = (e^X e^{tY})^{-1} e^X e^{tY} Y = Y.$$

Hence,

$$\left\{ \frac{I - e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}} \right\} \left(\frac{dZ}{dt} \right) = Y.$$

If X and Y are sufficiently close to 0, then $Z(t)$ would as well be sufficiently close to 0 for $0 \leq t \leq 1$. Hence by arguments above,

$$\frac{I - e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}}$$

is invertible for sufficiently small X and Y.

Hence,

$$\frac{dZ}{dt} = \left\{ \frac{I - e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}} \right\}^{-1} (Y).$$

Since $e^{Z(t)} = e^X e^{tY}$, we may use the properties of Ad and ad to conclude

$$\begin{aligned} \mathrm{Ad}_{e^{Z(t)}} &= \mathrm{Ad}_{e^X} \mathrm{Ad}_{e^{tY}} \\ \implies e^{\mathrm{ad}_{Z(t)}} &= e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} \\ \implies \mathrm{ad}_{Z(t)} &= \log(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}) \end{aligned}$$

The last implication follows as X, Y, and $Z(t)$ are small and hence log is defined for $e^{\mathrm{ad}_{Z(t)}}$ and $e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}$.

This implies that

$$\frac{dZ}{dt} = \left\{ \frac{I - (e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y})^{-1}}{\log(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y})} \right\}^{-1} (Y).$$

Recall that the function g defined in equation 6.1 is

$$g(A) = \left\{\frac{1 - A^{-1}}{\log A}\right\}.$$

Thus,

$$\frac{dZ}{dt} = g(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y})(Y)$$

is a differential equation with initial condition $Z(0) = \log(e^X e^0) = X$. Integrating, we conclude that

$$\log(e^X e^Y) = Z(1) = X + \int_0^1 g(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y})(Y)dt,$$

which proves Theorem 6.1. ■

We conclude this section by proving the series form of the BCH formula, which involves a series of nested commutators. By using the Taylor series expansion of g at $z = 1$, it can be shown that

$$g(A) = 1 + \frac{1}{2}(A - I) - \frac{1}{6}(A - I)^2 + \frac{1}{12}(A - I)^3 + \cdots$$

for all A such that $\|A - I\| < 1$. We also must have

$$\begin{aligned} e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} - I &= \\ = (I + \mathrm{ad}_X + \frac{\mathrm{ad}_X^2}{2!} + \cdots)(I + t\mathrm{ad}_Y + \frac{t^2\mathrm{ad}_Y^2}{2!} + \cdots) - I& \\ = \mathrm{ad}_X + t\mathrm{ad}_Y + t\mathrm{ad}_X\mathrm{ad}_Y + \frac{\mathrm{ad}_X^2}{2} + \frac{t^2\mathrm{ad}_Y^2}{2} + \cdots .& \end{aligned}$$

Since $e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} - I$ has no zeroth-order term with respect to ad_X and ad_Y, $(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} - I)^m$ will have no mth-order term. Thus, computing g upto degree 2 gives

$$\begin{aligned} g(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y})& \\ = I + \frac{1}{2}(\mathrm{ad}_X + t\mathrm{ad}_Y + t\mathrm{ad}_X\mathrm{ad}_Y + \frac{\mathrm{ad}_X^2}{2} + \frac{t^2\mathrm{ad}_Y^2}{2})& \\ -\frac{1}{6}(\mathrm{ad}_X^2 + t^2\mathrm{ad}_Y^2 + t\mathrm{ad}_X\mathrm{ad}_Y + t\mathrm{ad}_Y\mathrm{ad}_X) + \cdots .& \end{aligned}$$

Applying the above expressing to Y and integrating (as well as nothing that $\mathrm{ad}_Y(Y) = 0$, hence all terms with ad_Y at the beginning are zero), we get

$$\begin{aligned} \log(e^X e^Y)& \\ = X + \int_0^1 (Y + \frac{1}{2}[X,Y] + \frac{1}{4}[X,[X,Y]] - \frac{1}{6}[X,[X,Y]] - \frac{t}{6}[Y,[X,Y]] + \cdots)& \\ = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdots ,& \end{aligned}$$

continuing as an infinite series of nested brackets.

7. Simply Connected Lie Group Homomorphisms

Recall that if (X, τ) is a topological space, then it is said to be simply connected if for all continuous maps $f_0, f_1 : [0,1] \to X$ such that $f_0(0) = f_1(0)$ and $f_0(1) = f_1(1)$, there must exist a continuous map $F : [0,1] \times [0,1] \to X$ such that

- $F(0,t) = f_0(t)$
- $F(1,t) = f_1(t)$
- $F(s,0) = f_0(0) = f_1(0)$
- $F(s,1) = f_0(1) = f_1(1)$

for all $s, t \in [0,1]$.

Equivalently, for all closed loops $f(t)$ in X, there exists a map $F(s,t)$ that contracts $f(t)$ into a single point as s ranges from 0 to 1.

We prove a partial converse to 5.8, which gives a unique canonical matrix Lie group homomorphism, when given a Lie algebra homomorphism. This is Theorem 1.1. We first prove this result locally

Definition 7.1. If G and H are matrix Lie groups, then a **local homomorphism** of G to H is a pair (Φ', U), where U is a connected neighbourhood of the identity in G, and $\Phi' : U \to H$ is a continuous map such that $\Phi'(AB) = \Phi(A')\Phi(B')$ for all $A, B \in U$ with $AB \in U$.

Note that U doesn't necessarily need to be a subgroup of G.

Theorem 7.2. *Let G and H be matrix Lie groups, with Lie algebras $\mathfrak{g}$ and $\mathfrak{h}$. Let $\phi : \mathfrak{g} \to \mathfrak{h}$ be a Lie algebra homomorphism. Define $U_\epsilon \in G$ as*

$$U_\epsilon = \{A \in G | \|A - I\| < 1 \text{ and } \|\log A\| < \epsilon\}.$$

Then there exists $\epsilon > 0$ such that the map $\Phi' : U_\epsilon \to H$ given by

$$\Phi'(A) = e^{\phi(\log A)}$$

is a local homomorphism.

Note that for this theorem, G needn't be simply connected.

Proof. Firstly, observe that Φ' must be continuous as it is the composition of the continuous function log, ϕ (which is continuous as it's a linear map), and exp.

Let $\epsilon > 0$ be small enough such that Theorem 5.13 holds, and small enough so that for any $A, B \in U_\epsilon$, the BCH formula holds for $X = \log A$ and $Y = \log B$, as well as $\phi(X))$ along with $\phi(Y)$ (note that $X, Y \in \mathfrak{g}$ and $\phi(X), \phi(Y) \in \mathfrak{h}$) . If $AB \in U_\epsilon$, then

$$\Phi'(AB) = \Phi'(e^X e^Y) = e^{\phi(\log(e^X e^Y))}.$$

We now apply the BCH formula, which can be expressed as an infinite series of nested brackets. We have

$$\begin{aligned}
&\phi(\log(e^X e^Y)) \\
&= \phi(X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdots) \\
&= \phi(X) + \phi(Y) + \frac{1}{2}[\phi(X), \phi(Y)] + \frac{1}{12}[\phi(X), [\phi(X), \phi(Y)]] - \frac{1}{12}[\phi(Y), [\phi(X), \phi(Y)]] + \cdots \\
&= \log(e^{\phi(X)} e^{\phi(Y)}),
\end{aligned}$$

where we've used the fact that $\phi([X,Y]) = [\phi(X),\phi(Y)]$. We now get

$$\Phi'(AB) = e^{\phi(\log(e^X e^Y))} = e^{\log(e^{\phi(X)}e^{\phi(Y)})} = e^{\phi(X)}e^{\phi(Y)} = \Phi'(A)\Phi'(B),$$

which proves the result. Note that $\exp \circ \log$ is the identity in this case as X and Y are sufficiently close to I. ■

We now wish to extend Φ' into a global homomorphism over G, which will be possible thanks to the group's simple-connectedness.

Theorem 7.3. *Suppose G and H are matrix Lie groups with G simply-connected. If (Φ', U) is a unique local homomorphism of G into H, then there exists a unique Lie grop homomorphism $\Phi : G \to H$ such that $\Phi|_U = \Phi'$ (i.e. Φ agrees with Φ' on U).*

Proof. Let $A \in G$, and define a continuous path $f : [0,1] \to G$ such that $f(0) = I$ and $f(1) = A$. We say that a partition $0 = t_0 < t_1 < \cdots < t_n = 1$ is a **good partition** if for each $s,t \in [t_i, t_{i+1}]$ for any $0 \leq 0 \leq n-1$,

$$f(s)f(t)^{-1} \in U.$$

By Lemma 5.15, a good partition exists, (by making the mesh of the partition less than the δ defined in Lemma 5.15). We define

$$\Phi(A) = \Phi'(f(t_n)f(t_{n-1})^{-1})\cdots\Phi'(f(t_2)f(t_1)^{-1})\Phi(f(t_1)f(t_0)^{-1}). \tag{7.1}$$

We now prove that $\Phi(A)$ is independent of the partition or path chosen.

If $t_0, \dots, t_n$ is a good partition, then inserting any s in between some $[t_i, t_{i+1}]$ the partition will still be good, as the mesh would either stay the same or decrease. The effect the addition of s has on 7.1 is the replacing the factor $\Phi'(f(t_{i+1})f(t_i)^{-1})$ with

$$\Phi'(f(t_{i+1})f(s)^{-1})\Phi'(f(s)f(t_i)^{-1}).$$

Since Φ' is a local homomorphism, and $f(t_{i+1})f(s)^{-1}, f(s)f(t_i)^{-1}$, and $f(t_{i+1})f(t_i)^{-1}$ are all in U, we have

$$\Phi'(f(t_{i+1})f(t_i)^{-1}) = \Phi'(f(t_{i+1})f(s)^{-1})\Phi'(f(s)f(t_i)^{-1}),$$

hence not changing the value of $\Phi(A)$. Hence, $\Phi(A)$ will remain the same for any finite number of additions in its defining partition. Now, any two permutations have a common refinement, namely their union. Since $\Phi(A)$ on each partition stays the same in their union, we conclude that both partitions yield the same $\Phi(A)$, and hence Φ partition independent.

We now prove $\Phi(A)$ is path independent. Let $f_0(t)$ and $f_1(t)$ be continuous paths in G joining I to A. Since G is simply-connected, there exists a function F such that for all $s,t \in [0,1]$,

(1) $F(0,t) = f_0(t)$ and $F(1,t) = f_1(t)$
(2) $F(s,0) = I$ and $F(s,1) = A$.

As in Lemma 5.15, there exists $N \in \mathbb{N}$ such that for all $s, s', t,' \in [0,1]$ with $|s-s'|, |t-t'| < 2/N$,

$$f(s,t)f(s',t')^{-1} \in U.$$

Let $g_{j,k}(t)$ (with $j,k = 0,1,\dots,N-1$ and $k = 0,1,\dots,N$) be continuous paths from $[0,1]$ to G that connect I to A. They are defined as:

$$g_{j,k}(t) = \begin{cases} F(\frac{j+1}{N}, t) & 0 \leq t \leq \frac{k-1}{N} \\ F(\frac{j+k}{N} - t, t) & \frac{k-1}{N} \leq t \leq \frac{k}{N} \\ F(\frac{j}{N}, t) & \frac{k}{N} \leq t \leq 1 \end{cases}$$

when $k \neq 0$. If $k = 0$, then define $g_{j,k}(t) = F(\frac{j}{N}, t)$ for all $t \in [0,1]$. In particular, we have $g_{0,0} = f_0(t)$.

We now deform f_0 to $g_{0,1}$, then $g_{0,1}$ to $g_{0,2}$, and so unit we reach $g_{0,N}$, which we then reform to $g_{1,0}$, and continue in a lexicographical manner, until we reach $g_{N-1,N}$, which we deform into f_1. The claim is that $\Phi(A)$ is the same over all such deformations. Note that the pair $g_{j,k}(t)$ and $g_{j,k+1}(t)$ for some $0 \leq j, k \leq N-1$ are identical over $[0,1]$ except for the interval

$$\left(\frac{k-1}{N}, \frac{k+1}{N}\right).$$

We have proven above that $\Phi(A)$ is partition independent, and we are free to choose any good partition. Let that partition be

$$0, \frac{1}{N}, \cdots, \frac{k-1}{N}, \frac{k+1}{N}, \frac{k+2}{N}, \cdots, 1.$$

which is good by definition of N. By 7.1, $\Phi(A)$ only depends on the partition points. Since $g_{j,k}(t)$ and $g_{j,k+1}(t)$ agree on the above partition points, we must have both paths yield the same $\Phi(A)$.

Similarly, the pair $g_{j,N}$ and $g_{j+1,0}$ for some $0 \leq j \leq N-2$, along with the pair $g_{N-1,N}$ and f_1 are identical over all of $[0,1]$ except

$$(\frac{N-1}{N}, 1).$$

Choosing the partition

$$0, \frac{1}{N}, \cdots, \frac{N-1}{N}, N$$

will suffice as $g_{j,N}$ and $g_{j+1,0}$ agree on all of the partition points. Hence both paths yield the same $\Phi(A)$. All of this together implies that by deforming f_1 to f_2 via this process preserves $\Phi(A)$, and hence Φ is path independent.

We observe that Φ is continuous as Φ' is continuous and for two sufficiently close points A and B, they have paths f_A and f_B from I to A and B respectively whose max distance from one another is sufficiently small.

If $f_A(t)$ and $f_B(t)$ are paths connecting I to A and B respectively, define $f_{AB}(t)$ as

$$f_{AB}(t) = \begin{cases} f_B(2t) & 0 \leq t \leq \frac{1}{2} \\ f_A(2t-1)B & \frac{1}{2} \leq t \leq 1. \end{cases}$$

If $t_0, \ldots, t_n$ is a good partition of f_A and $s_0, \ldots, s_m$ is a good partition of f_B, then

$$\frac{s_0}{2}, \cdots, \frac{s_m}{2}, \frac{1+t_0}{2}, \cdots, \frac{1+t_n}{2}$$

is a good partition of f_{AB}, noting that $\frac{s_m}{2} = \frac{1}{2} = \frac{1+t_0}{2}$. Then

$$\begin{aligned}\Phi(AB) = \Phi'(f_{AB}(\frac{1+t_n}{2})f_{AB}(\frac{1+t_{n-1}}{2})^{-1}) \cdots \Phi'(f_{AB}(\frac{1+t_1}{2})f_{AB}(\frac{1+t_0}{2})^{-1})\Phi'(f_{AB}(\frac{t_0+1}{2})f_{AB}(\frac{s_m}{2})^{-1}) \\ \Phi'(f_{AB}(\frac{s_m}{2})f_{AB}(\frac{s_{m-1}}{2})^{-1}) \cdots \Phi'(f_{AB}(\frac{s_2}{2})f_{AB}(\frac{s_1}{2})^{-1})\Phi'(f_{AB}(\frac{s_1}{2})f_{AB}(\frac{s_0}{2})^{-1}) \\ = \Phi(A)\Phi(B)\end{aligned}$$

since $\frac{s_m}{2} = \frac{1+t_0}{2}$, and thus

$$f(\frac{1+t_0}{2})f(\frac{s_m}{2})^{-1} = I.$$

Hence $\Phi(AB) = \Phi(A)\Phi(B)$, and Φ is a Lie group homomorphism.

Lastly, we prove that Φ agrees with Φ' on U. Let $A \in U$, and $t_0, \dots, t_n$ a good partition over a path f in U (f exists since we assume U is connected). Then we use induction to prove $\Phi(A) = \Phi'(A) = \Phi'(f(t_n))$.

For the base case, we note that

$$\Phi(f(t_1)) = \Phi'(f(t_1)f(t_0)^{-1}) = \Phi'(f(t_1))$$

as $t_0 = 0$ and $f(t_0) = I$.

For induction, assume $\Phi(f(t_j)) = \Phi'(f(t_j))$, we have

$$\Phi(f(t_{j+1})) = \Phi'(f(t_{j+1})f(t_j)^{-1})\Phi'(f(t_j)f(t_{j-1})^{-1}) \cdots \Phi'(f(t_1))$$
$$= \Phi'(f(t_{j+1})f(t_j)^{-1})\Phi(f(t_j)) = \Phi'(f(t_{j+1})f(t_j)^{-1})\Phi'(f(t_j)) = \Phi'(f(t_{j+1})),$$

where we used the fact that Φ' is a local homomorphism. Hence $\Phi(A) = \Phi'(A)$ for all $A \in U$. ■

Proof of Theorem 1.1. For existence, let Φ' be the local homomorphism in Theorem 7.2, and Φ the global homomorphism in Theorem 7.1. For $X \in \mathfrak{g}$, the matrix $e^{X/m}$ will be in U for sufficiently large m, and

$$\Phi(e^{X/m}) = \Phi'(e^{X/m}) = e^{\phi(X)/m}.$$

Since Φ is a homomorphism,

$$\Phi(e^X) = \Phi(e^{X/m})^m = e^{\phi(X)}$$

as desired.

Assume for contradiction that Φ_1 and Φ_2 are two homomorphisms such that $\Phi_1(A) \neq \Phi_2(A)$. Since G is connected, by Corollary 5.14, we have $A = e^{X_1} \cdots e^{X_n}$ for $X_i \in \mathfrak{g}$. Hence

$$\Phi_1(A) = \Phi_2(A)$$

which contradicts our assumption. Hence Φ is unique, and our main theorem has been proven ■

8. Conclusion and Acknowledgements

The above exploration into Lie Theory can be applied to the representation theory of Lie groups and algebras, as representations are a special case of homomorphisms.

Using the Baker-Campbell-Hausdorff formula, one can also prove that if G is a matrix Lie group with Lie algebra $\mathfrak{g}$, then for any subalgebra $\mathfrak{h} \subseteq \mathfrak{g}$, there exists a unique "Lie connected subgroup" H such that the Lie algebra of H is $\mathfrak{h}$.

By the above result and Ado's theorem (which state that every Lie algebra is a Lie algebra of matrices with the Lie product being the commutator), it follows that for every real Lie algebra $\mathfrak{g}$, there exists a matrix Lie group G such that $\mathfrak{g}$ is isomorphic to the Lie algebra of G.

I would like to thank my instructor Simon Rubinstein-Salzedo, my peer Atticus Kuhn and my mentor Annika Mauro for their guidance and feedback on this paper, which was done for an independent research writing class of Euler Circle.

References

[1] Lars V Ahlfors. *Complex analysis*. 1979.
[2] Alen Alexanderian. Matrix lie groups and their lie algebras.
[3] Robert Gilmore. *Lie groups, Lie algebras, and some of their applications*. Courier Corporation, 2006.
[4] Brian C Hall and Brian C Hall. *Lie groups, Lie algebras, and representations*. Springer, 2013.
[5] JA1088363 Oteo. The baker–campbell–hausdorff formula and nested commutator identities. *Journal of mathematical physics*, 32(2):419–424, 1991.

[1] [2] [3] [4] [5]