Baker-Campbell-Hausdorff Formula

Charles Daly

Summary

These notes are dedicated to a summary of the Baker-Campbell-Hausdorff Formula. They are take directly from Brian Hall's proof from 'Lie Groups, Lie Algebras, and Representations' [3] written with my thoughts. I wrote these notes up to provide to the class when covering for Dr. Goldman. I strongly encourage the reader to check out Brian Hall's proof, as it is outstandingly clear and probably the most accessible version of the proof that I've seen. Though the proof is specialized specifically to the case of matrix groups, it conveys all the main ideas of the general proof in an admirably clear fashion.

1 Motivation

One of the main applications I have seen of the Baker-Campbell-Hausdorff formula, henceforth in these notes referred to as the BCH formula, is to prove Lie's Third Theorem. This theorem loosely states that every lie algebra homomorphism is as good as a lie group homomorphism, provided the domain lie group is simply connected.

That such a formula exists is somewhat remarkable, and is the topic of these notes. Once the formula is introduced, we will go on to show how to construct the corresponding lie group homomorphism. After the proof of Lie's Third Theorem, we return to the proof of the BCH formula in both its series and integral formulation. As stated in the summary, these notes are taken directly from Brian Hall's text 'Lie Groups, Lie Algebras, and Representations' [3]. I have shuffled around some of the ideas, omitted certain parts, added my own, but nevertheless all the ideas here are largely contained in his text. It seems a good a place as any to state the main formula here.

Theorem 1. (BCH Formula) *Let G be a lie group and denote the exponential map* $\exp : \mathfrak{g} \longrightarrow G$ *where $\mathfrak{g}$ is the lie algebra of G. There exists a sufficiently small open subset $0 \in U \subset \mathfrak{g}$ so that the exponential map admits a local inverse, denoted* $\log : \exp(U) \longrightarrow U$ *so that for all* $\exp(X)\exp(Y) \in \exp(U)$ *one has*

$$\log(\exp X \exp Y) = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \dots \tag{1}$$

It is not obvious what the remaining terms or pattern is for the rest of the above formula, but that is fine, as the take away here is that Equation 1 may be written entirely in terms of X, Y and iterated brackets of X and Y. What this is saying is that, at least locally, the lie algebra of a lie group determines the group structure. Perhaps this provides some intuition as to why the following theorem holds.

Theorem 2. (Lie's Third Theorem) *Let G and H be lie groups where G is simply connected, and let $\phi : \mathfrak{g} \longrightarrow \mathfrak{h}$ be a lie algebra homomorphism. There exists a lie group homomorphism $\Phi : G \longrightarrow H$ so that the following diagram commutes*

$$\begin{tikzcd} G \arrow[r, "\Phi"] & H \\ \mathfrak{g} \arrow[u, "\exp"] \arrow[r, "\phi"'] & \mathfrak{h} \arrow[u, "\exp"'] \end{tikzcd} \tag{2}$$

The hypothesis that G be simply connected is indeed necessity, as one can easily construct lie algebra homomorphisms which do not have corresponding lie group homomorphisms. A simple example would be take the circle group as G and the vector space $\mathbb{R}$ as H. Certainly these lie algebras are isomorphic, but there is no non-trivial lie algebra homomorphism from G to H, for if $\Phi(e^{i\theta}) = x \in \mathbb{R}$ for some non-zero x, then $\Phi(e^{in\theta}) = nx$ which is arbitrarily large, thus violating compactness of $\Phi(S^1) \subset \mathbb{R}$.

Before actually moving onto the next section, I find it edifying to take a baby step to prove the following from John Lee's 'Smooth Manifolds' [4]. This is Proposition 20.10 in the second edition.

Theorem 3. (John Lee's Smooth Manifolds Prop 20.10) *Let G be a lie group and denote the exponential map* $\exp : \mathfrak{g} \longrightarrow G$ *where $\mathfrak{g}$ is the lie algebra of G. For each $X, Y \in \mathfrak{g}$, there's an $\epsilon > 0$ and a smooth $Z : (-\epsilon, \epsilon) \longrightarrow \mathfrak{g}$ so that*

$$\exp(tX)\exp(tY) = \exp\big(t(X+Y) + t^2 Z(t)\big) \tag{3}$$

Note that Equation 3 is essentially a baby version of Equation 1. In fact, it's the statement that

$$\log\left(\exp(tX)\exp(tY)\right) = t(X+Y) + t^2 Z(t) \text{ for sufficiently small X,Y}$$

Proof. The proof below is really just an exercise in taking derivatives. Let $0 \in U \subset \mathfrak{g}$ be so small that $\exp|_U : U \longrightarrow \exp(U)$ is a diffeomorphism. If we take $X, Y \in \mathfrak{g}$, we can find an ϵ sufficiently small so that $\exp(tX)\exp(tY) \in U$ for all $|t| < \epsilon$. Perhaps it worth noting that here we're taking a norm on $\mathfrak{g}$, and it doesn't really matter which one, as all norms of vector spaces are equivalent in the sense that they all induce the same topology.

Define $f : (-\epsilon, \epsilon) \longrightarrow \mathfrak{g}$ by $\log\left(\exp(tX)\exp(tY)\right)$. The map f is smooth as it is the composition of

$$(-\epsilon, \epsilon) \xrightarrow{\exp_X \times \exp_Y} \exp(U) \times \exp(U) \xrightarrow{m} \exp(U) \xrightarrow{\log} U$$

where $\exp_X(t) = \exp(tX)$ and $\exp_Y(t) = \exp(tY)$, so f is smooth. Taking the differential at zero yields the sequence of linear maps.

$$T_0\mathbb{R} \xrightarrow{(d(\exp_X)_0, d(\exp_Y)_0)} \mathfrak{g} \oplus \mathfrak{g} \xrightarrow{dm_{(e,e)}} \mathfrak{g} \xrightarrow{(d(\exp)_0)^{-1}} \mathfrak{g} \tag{4}$$

It is worth noting there are several identifications at hand. Firstly, note that since $\exp_X$ and $\exp_Y$ are restrictions of one parameter subgroups, there's no harm in making the identification of T_0G with $\mathfrak{g}$ as we have done above. That said, note by construction $d(\exp_X)_0(\partial/\partial t|_{t=0}) = X$, and similarly for Y. Moreover, it is a good exercise to show that the derivative of the multiplication map m at (e, e) takes any pair of tangent vectors $(X, Y) \in \mathfrak{g} \times \mathfrak{g}$ to their sum $X + Y$. Finally, since the derivative of the exponential map $d(\exp)_0 : \mathfrak{g} \longrightarrow \mathfrak{g}$ is the identity map, up to appropriate identification, we have that $f'(0) = df_0(\partial/\partial|_{t=0}) = X + Y$. By Taylor's theorem, there exists a $Z : (-\epsilon, \epsilon) \longrightarrow \mathfrak{g}$ so that $f(t) = f(0) + tf'(0) + t^2 Z(t)$, and thus

$$f(t) = 0 + t(X+Y) + t^2 Z(t) \text{ for all } t \in (-\epsilon, \epsilon)$$

□

As mentioned previously, this is arguably the starting point for the BCH-formula and says loosely that $\exp(tX)\exp(tY) = \exp\left(t(X+Y) + \text{higher order terms}\right)$. The solution for what those higher order terms are is the BCH formula.

2 Proof of Lie's Third Theorem

Before providing the proof of the BCH formula, I wanted to provide a proof of Lie's Third Theorem for the following reason. There are lots of proofs of the BCH formula, which to my knowledge, all in some sense end up taking the derivative of the exponential map $\exp : \mathfrak{g} \longrightarrow G$, and this derivative is gnarly.

That said, the proof of Lie's Third Theorem uses a construction which is arguably quite similar to a method I've seen several times in geometry/topology similar to that of analytic continuation. A little more specifically, one defined a function not on a space M itself, but rather the homotopy classes of paths of M, which consequently induces a map on the universal cover of M. Dr. Goldman's book 'Geometric Structures on Manifolds' [2] provides a treatment of this process in detail with a construction called the developing map. This map plays an important role in the theory of geometric structures, and its construction uses this sort of 'define along a path' argument.

That said I think I want to spend quite a bit of time filling in the details for this argument before approaching the proof of the BCH formula. Again as mentioned, this is taken directly from Brian Hall's text 'Lie Groups, Lie Algebras, and Representations.'

Just as was done in the beginning of the proof of this section, I wanted to provide some motivation before illustrating the actual steps. Let's say one assumes Equation 1 and wants to construct the commutative square as in Equation 2. It seems not entirely unreasonable to begin by 'defining' $\Phi : G \longrightarrow H$ via $\Phi(\exp(X)) = \exp(\phi(X))$.

There are some immediate and obvious challenges to this though. Firstly, it not obvious at all whether every element of G is expressible as $\exp(X)$ for some $X \in \mathfrak{g}$. In fact, it's a good exercise to cook up a lie group where the exponential map is not surjective. As a hint, looking at two by two matrices will suffice, in fact the negative identity almost does the trick. In addition, it is far from obvious that even if such an X exists, it is not clear that $\exp(\phi(X))$ would be independent of the choice of X. To make matters worse, why would Φ defined as such, be a group homomorphism?

The BCH-formula resolves the third issue, whereas the hypothesis that G be simply connected will resolve the first and second. Let us assume for the moment that we have the BCH formula and we pick neighborhoods U and V of $\mathfrak{g}$ and $\mathfrak{h}$ so small where the BCH formula holds on both, and $\phi(U) \subset V$, so that ϕ is a 'local' lie algebra-homomorphism. Now because we have

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

we may apply ϕ to both sides to yield

$$\begin{aligned} \phi\left(\log(\exp X \exp Y)\right) &= \phi(X) + \phi(Y) + \frac{1}{2}[\phi(X),\phi(Y)] + \frac{1}{12}[\phi(X),[\phi(X),\phi(Y)]] \\ &\quad - \frac{1}{12}[\phi(Y),[\phi(X),\phi(Y)]] + \dots = \log\left(\exp(\phi(X))\exp(\phi(Y))\right) \end{aligned} \tag{5}$$

where the last equality holds because we've assumed the BCH on V as well. Since $e^X e^Y \in U$, and the exponential map is invertible on $\phi(U)$, we have

$$e^X e^Y = e^{\log(e^X e^Y)} \tag{6}$$

That said, if we define $\Phi : \exp(U) \longrightarrow \exp(V)$ via $\Phi(\exp(X)) := \exp(\phi(X))$, then by Equation 6, we have that

$$\begin{aligned}\Phi(e^X e^Y) &= \exp\left(\phi\left(\log(e^X e^Y)\right)\right) = \exp\left(\log\left(\exp(\phi(X))\exp(\phi(Y))\right)\right) \\ &= e^{\phi(X)}e^{\phi(Y)} = \Phi(e^X)\Phi(e^Y)\end{aligned}$$

where the first and fourth equalities are by definition, the second by Equation 5, and the third by prospect that log and exp are inverses. Note that in the definition of Φ as above, since we restricted our attention to the case where $\exp : U \longrightarrow \exp(U)$ is a diffeomorphism, we avoid the issues of whether or not each element in $\exp(U)$ is expressible as $\exp(X)$ for some $X \in \mathfrak{g}$, and questions about whether this map well defined.

It is perhaps worth emphasizing once again that the utility of the BCH formula is not to be found in the entire expansion of Equation 1. The utility of the BCH formula is found in the fact that one can express the logarithm of the product of exponentials of X and Y as the summation of X, Y, and corresponding lie brackets involving X and Y. Roughly speaking, such an expression allows us then to show that if under the conditions that $\Phi : G \longrightarrow H$ as defined above is actually a function, it may be promoted to a group homomorphism.

As advertised, we now proceed onto the proof of Theorem 2. This proof will be broken into several steps, and I strongly encourage the reader to understand each step, as this type of argument is used in many areas of geometry and topology.

To provide the context again, let G and H be lie groups, with G simply connected, and $\phi : \mathfrak{g} \longrightarrow \mathfrak{h}$ be a lie algebra homomorphism. Using the BCH formula, we wish to construct a corresponding $\Phi : G \longrightarrow H$ that is a is lie group homomorphism.

Proof. **(i). A local solution:** Much like we did above, let us choose a neighborhood $0 \in U \subset \mathfrak{g}$ so small such that $\log : \exp(U) \longrightarrow U$ is a diffeomorphism. We may also shrink U if necessary to assume that the BCH formula holds for all $e^X e^Y \in \exp(U)$. Define $\Phi : \exp(U) \longrightarrow H$ via

$$\Phi(g) = \exp\left(\phi(\log(g))\right)$$

Note that this definition provides a local solution to the desired homomorphism. We wish to extend it to a global solution. As is often the case in the scenario of simply connected spaces, one defines a global quantity by extending a local quantity along paths.

(ii). A global solution: Since G is simply connected, it is in particular connected, and consequently path-connected. Thus for each $g \in G$, we may find a path $\gamma : [0,1] \longrightarrow G$ so that $\gamma(0) = e$ and $\gamma(1) = g$. By compactness of the unit interval, we may choose a mesh of points $0 = t_0 < t_1 < \ldots < t_m = 1$ so that for all $s, t \in [t_i, t_{i+1}]$ for each i we have

$$g(t)g(s)^{-1} \in \exp(U) \tag{7}$$

Let us denote this condition in Line 7 on a partition by $(*)$. In particular, since $t_0 = 0$ and $\gamma(0) = 1$, we know that $g(t_1) \in V$. Now, we can write $g = \gamma(1)$ as

$$g = \left(\gamma(1)\gamma(t_{m-1})^{-1}\right) \cdot \left(\gamma(t_{m-1})\gamma(t_{m-2})^{-1}\right) \cdot \ldots \cdot \left(\gamma(t_2)\gamma(t_1)^{-1}\right) \cdot \gamma(t_1)$$

Let us now define Φ via

$$\Phi(g) = \Phi\left(\gamma(1)\gamma(t_{m-1})^{-1}\right) \cdot \Phi\left(\gamma(t_{m-1})\gamma(t_{m-2})^{-1}\right) \cdot \ldots \cdot \Phi\left(\gamma(t_2)\gamma(t_1)^{-1}\right) \cdot \Phi\left(\gamma(t_1)\right) \tag{8}$$

Note that each factor is inside of $\exp(U)$ and consequently Φ of each factor is well defined. We claim that this product is well-defined on homotopy classes of paths and is independent of partition. We begin by showing the independence of the partition.

(iii). Independence of Partition: Before getting into the proof of this part, it is worth mentioning that this is the only part of the proof that utilizes the BCH formula. This is also a relatively standard argument about independence of partitions. One shows that the quantity is unchanged under a refinement, and that every two partitions union to a refinement, and consequently have the same common value. This type of argument is somewhat similar to that of the riemann integral construction.

That said let $s \in (t_i, t_{i+1})$. Then the factor $\Phi\left(\gamma(t_{i+1}\gamma(t_i)^{-1})\right)$ as in Equation 8 is replaced by the product

$$\Phi\left(\gamma(t_{i+1})\gamma(s)^{-1}\right) \cdot \Phi\left(\gamma(s)\gamma(t_i)^{-1}\right)$$

Since s is between $t_i \leq s \leq t_{i+1}$, we know that both $\gamma(t_{i+1})\gamma(s)^{-1}$ and $\gamma(s)\gamma(t_i)^{-1}$ in addition to $\gamma(t_{i+1})\gamma(t_i)^{-1}$ are all in $\exp(U)$. By prospect of the BCH formula, Φ is a local group homomorphism, thus we have that

$$\Phi\left(\gamma(t_{i+1})\gamma(t_i)\right) = \Phi\left(\left(\gamma(t_{i+1})\gamma(s)^{-1}\right) \cdot \left(\gamma(s)\gamma(t_i)^{-1}\right)\right) = \Phi\left(\gamma(t_{i+1})\gamma(s)^{-1}\right) \Phi\left(\gamma(s)\gamma(t_i)^{-1}\right)$$

Consequently, the value of $\Phi(g)$ is left unchanged by the addition of an extra partition point. One can repeat this argument adding any finite number of points to the original partition.

Given two partitions, one may take their union, which is refinement of both partitions. Since these all share a common value, namely the value of the union, this shows the independence of partition.

(iv). Well Defined on Homotopy Classes of Paths: Here we use the fact that G is simply connected. Choose two paths γ and β starting at the identity e and ending at $g \in G$. Since G is simply connected, there's a homotopy $h : [0,1]^2 \longrightarrow G$ so that

$$h(0,t) = \beta(t) \text{ and } h(1,t) = \gamma(t) \text{ with fixed endpoints } h(s,0) = e \text{ and } h(s,1) = g$$

The compactness of $[0,1]^2$ guarantees the existence of an integer $N \in \mathbb{N}$ so that for all (s,t) and (s',t') in $[0,1]^2$ where $|s-s'| \leq 2/N$ and $|t-t'| \leq 2/N$, we have $h(s,t)h(s',t')^{-1} \in V$.

The following argument is one that makes several appearances in topology, whereby one deforms β onto γ bit by bit. Specifically, we define a sequence of paths $\alpha_{k,l}$ where $k = 0, \ldots, N-1$ and $l = 0, \ldots, N$. Define these paths

$$\alpha_{k,l}(t) = \begin{cases} h\left((k+1)/N, t\right) & 0 \leq t \leq (l-1)/N \\ h\left(k/N, t\right) & l/N \leq t \leq 1 \\ h(\text{connecting diagonal}) & (l-1)/N \leq t \leq l/N \end{cases}$$

where connecting diagonal is the line segment in $[0,1]^2$ connecting $((k+1)/N,(l-1)/N)$ and $(k/N,l/N)$. Figure 1 taken from Brian Hall [3] clarifies this definition. Finally, when define $\alpha_{k,0}(t)=\alpha(k/n,t)$ for all $t\in[0,1]$, so in particular, $\alpha_{0,0}=\beta$ and one could say $\alpha_{N,0}=\gamma$.

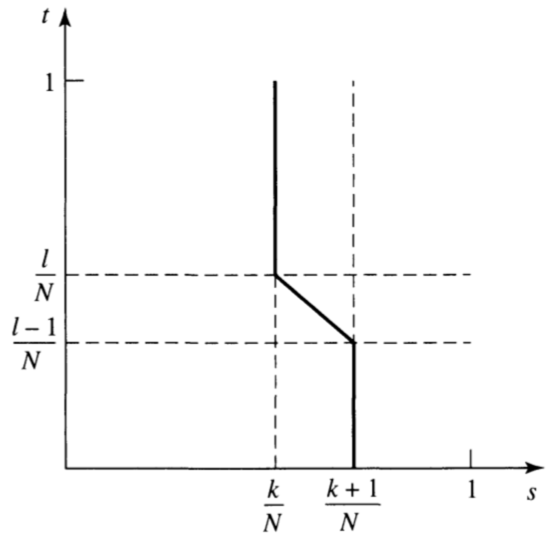

Figure 1: The path in $[0,1]^2$ defining $\alpha_{k,l}$. Note for $t\leq(l-1)/N$ we have $\alpha_{k,l}=h\left((k+1)/N,\cdot\right)$ and for $l/N\leq t$ we have $\alpha_{k,l}=h\left(k/N,\cdot\right)$, both of these are represented by the vertical bold line segments. Finally, in between $(l-1)/N\leq t\leq l/N$ we have $\alpha_{k,l}$ is defined on connecting diagonal attaching the two vertical lines.

Think of deforming the path β onto γ in steps. First, deform $\beta=\alpha_{0,0}$ into $\alpha_{0,1}$ and then into $\alpha_{0,2},\ldots,\alpha_{0,N}$. From $\alpha_{0,N}$, deform this into $\alpha_{1,0}$, then $\alpha_{1,1},\ldots,\alpha_{1,N}$. Continue until you reach $\alpha_{N-1,N}$ which deforms onto $\alpha_{N,0}=\gamma$. Below are several pictures to clarify the nature of this process.

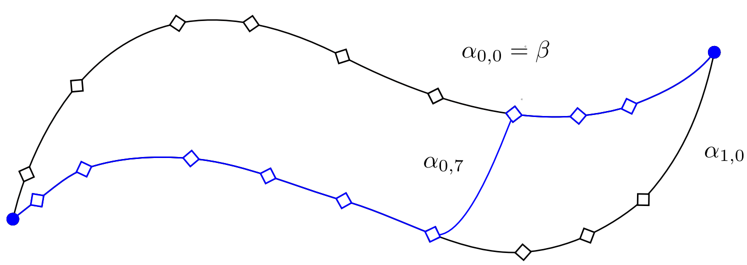

Figure 2: Here the bounding paths represent $\alpha_{0,0}$ and $\alpha_{1,0}$. Imagine they are black, but then covered by the blue path $\alpha_{0,7}$. Imagine each diamond as one of the i/N points in time, here we have $N=10$. The curve $\alpha_{0,7}$ agrees with $\alpha_{1,0}$ up until time $t=(7-1)/10=6/10$ time parameter, then in a single $1/10$ time, diagonally homotopes back to $\alpha_{0,0}$.

We wish to show that the value of $\Phi(g)$ along each path is the same as the value of $\Phi(g)$ as the next one. By definition, $\alpha_{k,l}(t)$ and $\alpha_{k,l+1}(t)$ are equal, with the exception of t

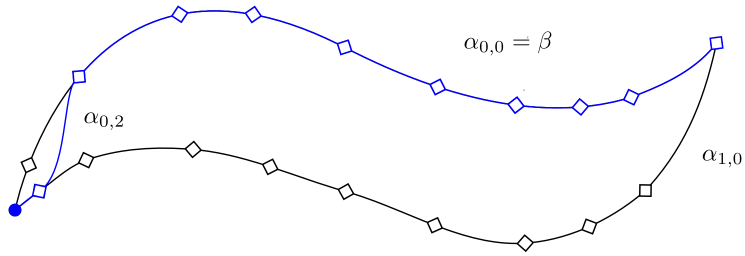

Figure 3: Just as in Figure 2, the black bounding paths are $\alpha_{0,0}$ and $\alpha_{1,0}$. Here this is a case to illustrate how if the second parameter in $\alpha_{k/N,*}$ is small, then $\alpha_{k/N,*}$ mostly agrees with $\alpha_{k/N,0}$. In this case, $\alpha_{0,0}$ and $\alpha_{0,2}$ agree everywhere except the first two $1/N$'s in time.

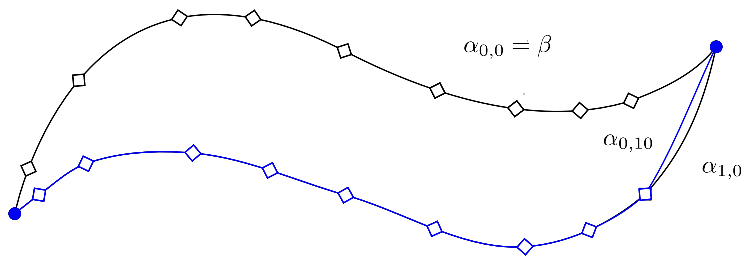

Figure 4: This example is to contrast Figure 3. If the second parameter in $\alpha_{k/N,*}$ is large, then $\alpha_{k/N,*}$ mostly agrees with $\alpha_{(k+1)/N,0}$. In this case, $\alpha_{1,0}$ and $\alpha_{0,10}$ agree everywhere except the last $1/N$ in time.

in $[(l-1)/N, (l+1)/N]$. To evaluate Φ along a path, we can choose any partition we like satisfying condition $(*)$. If we choose the partition below

$$0, \frac{1}{N}, \ldots, \frac{l-1}{N}, \frac{l+1}{N}, \frac{l+2}{N}, \ldots, 1$$

by choice of N, this satisfies $(*)$ on both paths $\alpha_{k,l}$ and $\alpha_{k,l+1}$. Moreover, since this partition has been chosen is in such a way that $\alpha_{k,l}$ and $\alpha_{k,l+1}$ are identical at all point of the partition, by definition of Φ along a path, namely Equation 8, Φ is the same for the paths $\alpha_{k,l}$ and $\alpha_{k,l+1}$. A similar argument shows that Φ is the same along $\alpha_{k,N}$ and $\alpha_{k+1,0}$. This will show that $\Phi(g)$ is the same for each path from $\beta = h_{0,0}$ to $h_{N-1,N}$, and applying the same argument, yields the same value of Φ on γ. Consequently we have independence of homotopy of path. Since G is simply connected, this means Φ is well defined.

(v). The map lifts ϕ: That Φ is a homomorphism is indeed relatively straight forward. One can pick a g and an h and paths α and β connecting them. From that one constructs a new path γ to gh by defining γ as α in double time, then $g\beta$ in shifted double time. The details are left to the reader and can be verified in a straight forward manner using a combination of partitions to satisfy the $(*)$ condition.

The very last part is to show that the differential of Φ is as claimed. Note that

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

where the second equality follows simply from definition of Φ, as for small enough t, $\exp(tX) \in \exp(U)$. This completes the proof of Lie's Third Theorem. $\square$

3 Derivative of Exponential Map

Most proofs of the BCH formula that I've seen endeavor to calculate the derivative of the exponential map $\exp : \mathfrak{g} \longrightarrow G$. Finding the derivative at the identity is a relatively straight forward task, whereas finding the derivative at an arbitrary point $X \in \mathfrak{g}$ becomes quite computationally taxing. While the general case is certainly more desirable, I don't think it is necessarily as enlightening as first working the case of matrix groups. The advantage of working in the case of matrix groups provides grounds in which we can multiply elements of the lie group by elements of the tangent space as per consequence of the fact that if G is a matrix group, namely a closed subgroup of $\mathrm{GL}(n, \mathbb{R})$, then its lie algebra $\mathfrak{g}$ naturally identifies with a lie subalgebra of the $n \times n$ matrices, which we denote by $\mathfrak{gl}(n, \mathbb{R})$. This identification is a consequence of embedding $G \leq \mathrm{GL}(n, \mathbb{R})$ where the topology of $\mathrm{GL}(n, \mathbb{R})$ is inherited from the euclidean space, $\mathbb{R}^{n \times n}$.

For example let $X \in \mathfrak{g}$. For any $g \in G$, consider the map $L_g : G \longrightarrow G$ and its derivative at the identity $d(L_g)_e : \mathfrak{g} \longrightarrow \mathfrak{g}$. As previously mentioned, the lie algebra $\mathfrak{g}$ sits inside the lie algebra of $\mathfrak{gl}(n, \mathbb{R})$. We claim that $d(L_g)_e(X) = gX$.

Let X be represented by the one parameter subgroup $\exp_X : \mathbb{R} \longrightarrow G$ where $\exp_X(t) = \exp(tX)$. Certainly $(L_g \circ \exp_X) : \mathbb{R} \longrightarrow G \subset \mathbb{R}^{n \times n}$, as per consequence

$$d(L_g)_e(X) = d(L_g \circ \exp_X)_0 \left(\frac{d}{dt}\bigg|_{t=0} \right) = \frac{d}{dt}\bigg|_{t=0} (L_g \circ \exp_X)(t) = g \frac{d}{dt}\bigg|_{t=0} \exp(tX) = gX \quad (9)$$

There are some subtitles in Equation 9. In particular, the third equality follows from the fact that that $G \subset \mathbb{R}^{n \times n}$, whereby we may identify tangent space of G at the identity, $T_{I_n}G$, and consequently the lie algebra itself, as a subspace of $T_{I_n}\mathbb{R}^{n \times n}$ where I_n is the $n \times n$-identity matrix. Because $\mathbb{R}^{n \times n}$ is a vector space, its tangent spaces at any point enjoys a canonical isomorphism between the tangent space and the vector space itself. Specifically, if $p \in V$ where V is some vector space, then T_pV is canonically isomorphic as a vector space to V.

This is absurdly false for general manifolds, and the fact that we may make this identification provides us the privilege of being able to take elements in the group G and multiply them by elements of the lie algebra $\mathfrak{g}$. For example, the adjoint representation of a matrix group $G \leq \mathrm{GL}(n, \mathbb{R})$ is given by what one would naively hope to be true, namely

$$\begin{aligned}\mathrm{Ad}_g(X) &= d(L_g \circ R_{g^{-1}})_e(X) = d(L_g \circ R_{g^{-1}} \circ \exp_X)_0 \left(\frac{d}{dt}\bigg|_{t=0} \right) \\ &= \frac{d}{dt}\bigg|_{t=0} g \exp(tX) g^{-1} = gXg^{-1}\end{aligned} \tag{10}$$

As mentioned previously the inability to do this in the general context must be taken into account. From what I have seen, this tends to involve woefully complex derivatives. In fact, I have another bit of notes entirely dedicated to a single formula found in Duistermaat and Kolk [1] which has a proof of the BCH formula in full generality. This particular proof though implements a variational formula for a derivative of a family of vector fields which is far from intuitive, and warrants quite a bit of explanation.

All that said, we now proceed to the proof of the BCH formula in the context of matrix lie groups. As advertised previously, this will involve us taking the derivative of the exponential map at a point $X \in \mathfrak{g}$. In particular we will look at the expression,

$$\frac{d}{dt}\bigg|_{t=0} e^{X+tY}$$

where $X, Y \in \mathfrak{g}$. One would ideally like this to be equal to $e^X Y$, but this is unfortunately far from the case in general, and such a statement assumes that $[X, Y] = 0$.

Before proceeding we need to introduce a function

$$\frac{1 - e^{-z}}{z} = \frac{1 - (1 - z + z^2/2! - \dots)}{z}$$

which is analytic on the complex plane. In fact, its power series is given by

$$\frac{1 - e^{-z}}{z} = \sum_{k=0}^{\infty} (-1)^k \frac{z^k}{(k+1)!} = 1 - z/2! + z^2/3! - \dots$$

Since this series converges everywhere on $\mathbb{C}$, one may define for any linear transformation A on a vector space V the linear transformation

$$\frac{1 - e^{-A}}{A} := \sum_{k=0}^{\infty} (-1)^k \frac{A^k}{(k+1)!} = 1 - A/2! + A^2/3! - \dots \tag{11}$$

where 1 in the above equations denotes the identity map on $V \longrightarrow V$. Since the norm $||A||$ is finite, the series above indeed converges, and thus the definition in Equation 11 makes sense.

It is unfortunate that I have no better explanation of why we would consider this function apart from its implementation in the calculations below. Specifically, I mean that I don't see any reason why intuitively one would even begin to consider the above function in trying to take the derivative of the exponential map. That said, perhaps it will become more clear in the proof below, and consequently, I will try to be detailed and thorough so as to attempt to motivate such considerations.

Theorem 4. (Hall Theorem 3.5) *Let X and Y be $n \times n$ complex matrices, then*

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=0} e^{X+tY} &= e^X \left(\frac{1-e^{-\mathrm{ad}_X}}{\mathrm{ad}_X}(Y)\right) \\ &= e^X \left(Y - \frac{[X,Y]}{2!} + \frac{[X,[X,Y]]}{3!} - \dots\right) \end{aligned} \tag{12}$$

More generally if $X(t)$ is a smooth path of matrices, then

$$\left.\frac{d}{dt}\right|_{t=s} e^{X(t)} = e^{X(s)} \left(\frac{1-e^{-\mathrm{ad}_{X(s)}}}{\mathrm{ad}_{X(s)}} \left(\frac{dX}{dt}(s)\right)\right) \tag{13}$$

It is worth noting that Equation 13 follows immediately from Equation 12. One may verify this by considering the smooth map $Y(t) := X(s+t)$. Because $Y(t)$ is smooth, we may locally write

$$Y(t) = Y(0) + tY'(0) + t^2 Z(t) = X(s) + tX'(s) + t^2 Z(t)$$

where $Z(t)$ is some smooth path of matrices. The equation above for $Y(t)$ holds for some small enough neighborhood about $t = 0$. Now if we consider $\exp(Y(t))$ as a composition of functions $Y : (-\epsilon, \epsilon) \longrightarrow \mathfrak{g}$ and $\exp : \mathfrak{g} \longrightarrow G$, then the chain rules yields that

$$\begin{aligned} \left.\frac{d}{dt}\right|_{t=s} e^{X(t)} = \left.\frac{d}{dt}\right|_{t=0} e^{Y(t)} &= d(\exp)_{X(s)} \left(X'(s)\right) = \left.\frac{d}{dt}\right|_{t=0} e^{X(s)+tX'(s)} \\ &= e^{X(s)} \left(\frac{1-e^{-\mathrm{ad}_{X(s)}}}{\mathrm{ad}_{X(s)}} \left(X'(s)\right)\right) \end{aligned}$$

thus providing Equation 13.

The formula for the derivative of the exponential map as provided below follows Tuynman [5].

For any $n \times n$ matrices, X, Y, define

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

Since the exponential map is smooth, Δ is a linear map in Y for each fixed X, as it is the directional derivative. Now for each positive integer $m \in \mathbb{N}$,

$$\exp(X+tY) = \left(\exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right)^m$$

whereby we may differentiate and apply the product rule. Note we will obtain m-summands, each consisting of m-factors, $(m-1)$ of which are the evaluation of $\exp(X/m + t(Y/m))$ at $t = 0$, whereas one factor is differentiated at $t = 0$. Bear in mind we must preserve the order, as we lack commutativity.

$$\Delta(X,Y) = \sum_{k=0}^{m-1} \exp\left(\frac{X}{m}\right)^{m-k-1} \left(\left.\frac{d}{dt}\right|_{t=0} \exp\left(\frac{X}{m} + t\frac{Y}{m}\right)\right) \exp\left(\frac{X}{m}\right)^k$$

Factor out the $(m-1)$ terms of $\exp(X/m)$ in the expression $\exp(X/m)^{m-k-1}$ and rewrite the derivative of $\exp(X/m + t(Y/m))$ in the notation as $\Delta(X/m, Y/m)$. This yields

$$\Delta(X,Y) = \exp\left(\frac{m-1}{m}X\right)\sum_{k=0}^{m-1}\exp\left(\frac{X}{m}\right)^{-k}\Delta\left(\frac{X}{m},\frac{Y}{m}\right)\exp\left(\frac{X}{m}\right)^{k}$$

Note each summand is in fact of the form of the Adjoint representation. Explicitly,

$$\begin{aligned}\exp\left(\frac{X}{m}\right)^{-k}\Delta\left(\frac{X}{m},\frac{Y}{m}\right)\exp\left(\frac{X}{m}\right)^{k} &= \mathrm{Ad}^k_{\exp(-X/m)}\left(\Delta\left(\frac{X}{m},\frac{Y}{m}\right)\right)\\ &= \exp\left(\mathrm{ad}^k_{-X/m}\left(\Delta\left(\frac{X}{m},\frac{Y}{m}\right)\right)\right)\end{aligned}$$

where the last equality follows from the fact that $\mathrm{Ad}_{\exp(X)} = \exp\circ\mathrm{ad}_X$ for all $X \in \mathfrak{g}$. As one last simplification we factor out $1/m$ from the Y/m in the argument of Δ and rewrite $\mathrm{ad}_{-X/m}$ as $-\mathrm{ad}_X/m$ to yield the equality

$$\Delta(X,Y) = \exp\left(\frac{m-1}{m}X\right)\frac{1}{m}\sum_{k=0}^{m-1}\exp\left(-\frac{\mathrm{ad}_X}{m}\right)^{k}\left(\Delta\left(\frac{X}{m},Y\right)\right)$$

Note the left hand side is independent of m, and thus we may take the limit as $m \to \infty$. There are two terms whose limits are relatively clear. Specifically,

$$\lim_{m\to\infty}\exp\left(\frac{m-1}{m}X\right) = \exp(X) \text{ and } \lim_{m\to\infty}\Delta\left(\frac{X}{m},Y\right) = \Delta(0,Y) = Y \tag{14}$$

where $\Delta(0,Y) = Y$ as $\Delta(0,Y)$ is the derivative of the exponential map at $X = 0$.

The remaining term is given by,

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\exp\left(-\frac{\mathrm{ad}_X}{m}\right)^{k} \tag{15}$$

Hall provides some pretty neat reasoning as to where this series should converge. Pretend momentarily that ad_X is simply some numeric quantity. If so, then the summation in Equation 15 is a geometric series. For sufficiently large m, the quantity $-\mathrm{ad}_X/m < 1$ so that we may say

$$\frac{1}{m}\sum_{k=0}^{m-1}\exp\left(-\frac{\mathrm{ad}_X}{m}\right)^{k} = \frac{1}{m}\frac{1-\exp(-\mathrm{ad}_X)}{1-\exp(-\mathrm{ad}_X/m)} \longrightarrow \frac{1-\exp(-\mathrm{ad}_X)}{\mathrm{ad}_X}$$

We reason that a similar statement should hold in our case where ad_X is a linear operator and not a number. That said, we begin by writing $\exp(-\mathrm{ad}_X/m)^k$ as $\exp(-k\mathrm{ad}_X/m)$ and expanding the exponential term within the sum in Equation 15.

$$\begin{aligned}\frac{1}{m}\sum_{k=0}^{m-1}\exp\left(-\frac{k}{m}\mathrm{ad}_X\right) &= \sum_{i=0}^{\infty}\frac{1}{m}\sum_{k=0}^{m-1}\frac{1}{i!}\left(-\frac{k}{m}\mathrm{ad}_X\right)^{i}\\ &= \sum_{i=0}^{\infty}\left(\frac{1}{m}\sum_{k=0}^{m-1}\left(\frac{k}{m}\right)^{i}\right)\frac{(-1)^i}{i!}\mathrm{ad}^i_X\end{aligned}$$

The term in the parentheses immediately to the right of the summation is the riemann sum approximation of $\int_0^1 x^i \, dx$ which in limit is $1/(i+1)$. Ideally one would like to bring the limit $m \to \infty$ inside the summation $\sum_{i=0}^{\infty}$, but one must always be wary of interchanging limits. That said, note we have the following bound

$$\left|\left| \frac{1}{m} \sum_{k=0}^{m-1} \left(\frac{k}{m} \right)^i \frac{(-1)^i}{i!} \mathrm{ad}_X^i \right|\right| \leq \frac{1}{i+1} \frac{1}{i!} ||\mathrm{ad}_X||^i \tag{16}$$

which follows as the riemann sum is this particular case is strictly increasing towards its limit $1/(i+1)$. If one makes the identification of ad_X with an $n \times n$ matrix, then each term in the matrix associated to ad_X is bounded above the $||\mathrm{ad}_X||$. Since the limit as $i \to \infty$ of the right-hand side of Equation 16 converges, one may apply the dominated convergence theorem to interchange the limit $m \to \infty$ and $\sum_{i=0}^{\infty}$. This yields

$$\lim_{m \to \infty} \frac{1}{m} \sum_{k=0}^{m-1} \exp\left(-k \mathrm{ad}_X / m\right) = \sum_{i=0}^{\infty} \frac{(-1)^i}{(i+1)!} \mathrm{ad}_X^i = \frac{1 - e^{-\mathrm{ad}_X}}{\mathrm{ad}_X} \tag{17}$$

where the last equality follows by Equation 11. Combining Equation 14 and Equation 17 provide us

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

4 Proof of the Integral BCH Formula

In this section we provide one incarnation of the BCH Formula, namely an integral formula. Similar to the methods employed Equation 11, for a fixed linear operator $A : V \longrightarrow V$ on some vector space V, let

$$g(A) = \sum_{m=0}^{\infty} a_m (A-1)^m \text{ where } g(z) = \frac{\log(z)}{1 - \frac{1}{z}} = \sum_{m=0}^{\infty} a_m (z-1)^m$$

where we interpret the 1 in the equation for $g(A)$ as the identity map on V. The integral formulation of the BCH says for sufficiently small enough $X, Y \in \mathfrak{g}$,

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

To begin, let us define

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

If X and Y are sufficiently small, then $Z(t)$ may be defined for all $0 \leq t \leq 1$. We wish to compute $Z(1)$. By definition,

$$e^{Z(t)} = e^X e^{tY} \tag{18}$$

not to be confused with e^{X+tY} which was analyzed in Section 3. Taking the derivative of the above is a simple task relatively speaking and was addressed in Equation 9. In fact, we have the equality,

$$e^{-Z(t)}\frac{d}{dt}e^{Z(t)} = \left(e^X e^{tY}\right)^{-1} e^X e^{tY} Y = Y \tag{19}$$

On the other hand, Equation 13 provides us with

$$e^{-Z(t)}\frac{d}{dt}e^{Z(t)} = \left(\frac{1-e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}}\right)\left(\frac{dZ}{dt}\right) \tag{20}$$

Equating Equation 19 and Equation 20 yields

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

Provided one chooses X and Y sufficiently small, $Z(t)$ will also be small so that $(1 - e^{-\mathrm{ad}_{Z(t)}})/\mathrm{ad}_{Z(t)}$ will be close to the identity, and thus invertible. Consequently, we have

$$\frac{dZ}{dt} = \left(\frac{1-e^{-\mathrm{ad}_{Z(t)}}}{\mathrm{ad}_{Z(t)}}\right)^{-1}(Y) \tag{21}$$

Exponentiating $Z(t)$ as was done in Equation 18, and applying the Adjoint representation yields

$$\mathrm{Ad}_{e^{Z(t)}} = \mathrm{Ad}_{e^X} \circ \mathrm{Ad}_{e^{tY}} \implies e^{\mathrm{ad}_{Z(t)}} = e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} \tag{22}$$

which after taking logarithms yields

$$\mathrm{ad}_{Z(t)} = \log\left(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}\right) \tag{23}$$

Substituting both Equation 22 and Equation 23 into Equation 21 yields

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

We may rewrite $g(z)$ as in Equation 17 as

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

and as per consequence, considering the composition of $g(z)$ with the linear operator $e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}$ using the same arguments for convergence as in Equation 11.

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

Integrating the above from 0 to 1 and recalling that $Z(0) = X$ by definition yields the integral formulation of the BCH

$$Z(1) = \log\left(e^X e^Y\right) = X + \int_0^1 g\left(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}\right)(Y)\,dt \text{ where } g(z) = \frac{\log(z)}{1-\frac{1}{z}} \tag{24}$$

5 The Series BCH Formula

With the integral formulation of the BCH formula as in Equation 24, we may use taylor series approximations to obtain a series expansion of the BCH integral formula. That said, it worth noting at this point in its current incarnation, one should be able to see that Equation 24 provides us with ample evidence that the logarithm of the product may be written as a summation of terms involving X, Y, and their brackets, as this is true for both $\exp(\mathrm{ad}_X)$ and $\exp(t\mathrm{ad}_Y)$, so by linearity of the infinite summation of g, it will be true for the integral as well. That said it is nevertheless insightful and somewhat awe evoking to behold the series expansion.

Recall that $g(z)$ as in Equation 24 is given by

$$\begin{aligned} g(z) = \frac{z\log(z)}{z-1} &= \frac{(1+(z-1))\left((z-1) - \frac{(z-1)^2}{2} + \frac{(z-1)^3}{3} - \dots\right)}{(z-1)} \\ &= (1+(z-1))\left(1 - \frac{(z-1)}{2} + \frac{(z-1)^2}{3} + \dots\right) \end{aligned}$$

Multiplying and combining like terms provides

$$g(z) = 1 + \frac{1}{2}(z-1) - \frac{1}{6}(z-1)^2 + \dots$$

For those interested in obtaining more terms in the series expansion, a closed form expression for $g(z)$ is given by

$$g(z) = 1 + \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m(m+1)}(z-1)^m$$

As we intend to compose $e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}$ in $g(z)$ expanded about $z=1$, we observe that

$$\begin{aligned} &e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y} - 1 \\ &= \left(1 + \mathrm{ad}_X + \frac{1}{2}\mathrm{ad}_X^2 + \dots\right)\left(1 + t\mathrm{ad}_Y + \frac{t^2}{2}\mathrm{ad}_Y^2 + \dots\right) - 1 \\ &= \mathrm{ad}_X + t\mathrm{ad}_Y + t\mathrm{ad}_X\mathrm{ad}_Y + \frac{1}{2}\mathrm{ad}_X^2 + \frac{t^2}{2}\mathrm{ad}_Y^2 + \dots \end{aligned} \tag{25}$$

Note here that the above expansion contains no zero-th order identity term, simply expressions involving adjoint operations or higher powers of them. As per consequence each power of Equation 25 will also enjoy the same property. Bearing this in mind we may evaluate

$$\begin{aligned} g\left(e^{\mathrm{ad}_X} e^{t\mathrm{ad}_Y}\right) &= 1 + \frac{1}{2}\left(\mathrm{ad}_X + t\mathrm{ad}_Y + t\mathrm{ad}_X\mathrm{ad}_Y + \frac{1}{2}\mathrm{ad}_X^2 + \frac{t^2}{2}\mathrm{ad}_Y^2 + \dots\right) \\ &\quad - \frac{1}{6}\left(\mathrm{ad}_X + t\mathrm{ad}_Y + \dots\right)^2 + \dots \\ &= 1 + \frac{1}{2}\mathrm{ad}_X + \frac{t}{2}\mathrm{ad}_Y + \frac{t}{2}\mathrm{ad}_X\mathrm{ad}_Y + \frac{1}{4}\mathrm{ad}_X^2 + \frac{t^2}{4}\mathrm{ad}_Y^2 \\ &\quad - \frac{1}{6}\left(\mathrm{ad}_X^2 + t^2\mathrm{ad}_Y^2 + t\mathrm{ad}_X\mathrm{ad}_Y + t\mathrm{ad}_Y\mathrm{ad}_X\right) + \text{ higher order terms} \end{aligned}$$

Next we apply the above linear transformation to the vector Y and integrate from $t=0$ to $t=1$, keeping in mind that terms ending in ad_Y vanish when evaluated at Y. Adding this result to X yields Equation 24 as

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

which is stated in Equation 1. This complete the proof of the BCH series formula.

6 Concluding Remarks

Some final thoughts regarding the BCH formula, in practice it's nearly impossible to calculate in full generality. The above sections are perhaps daunting and fraught with both long calculations and possibly typos, though I hope I've minimized the latter. That said, it is well advised to grapple with some of these in specific cases. Two non-trivial cases that are well documented are the heisenberg group and the orientation preserving affine group of the reals. As both are simply connected, Lie's Third Theorem says their lie group structure is just as good as me telling you their lie algebra. I recommend checking this out.

For the heisenberg group let $\mathfrak{g}$ be generated by X, Y and Z subject to the relations $[X,Y]=Z$ and all other generating brackets are trivial. On the other hand, the orientation preserving affine group over $\mathbb{R}$ has a lie algebra generated by X and Y subject to $[X,Y]=X$.

For those a bit more ambitious try and calculate the exponential of the lie group $E(1,1)$ whose lie algebra is generated by X, Y and Z subject to $[Z,X]=X$ and $[Z,Y]=-Y$ and $[X,Y]=0$. This group is realized as the identity component of orientation preserving isometries of minkowski plane, namely $\mathbb{R}^2$ equipped with the non-riemannian inner product $(x,y)=x^1y^1-x^2y^2$. Thankfully a lot of these three-dimensional lie groups provide excellent examples and are not too computationally intensive.

References

[1] Hans Duistermaat and Johan Kolk. Lie Groups. pages 333–336, 2000.

[2] William Goldman. Geometric Structures on Manifolds. pages 91–95, 2020.

[3] Brian Hall. Lie groups Lie Algebras and Representations an Elementary Introduction. pages 63–80, 2003.

[4] John Lee. Smooth Manifolds Second Edition. pages 522–523, 2000.

[5] G. Tuynman. The Derivation of the Exponential Map of Matrices. *Amer Math Monthly*, 102:818–819, 1995.