LECTURE 8-9: THE BAKER-CAMPBELL-HAUSDORFF FORMULA

1. Taylor's expansion on Lie group

As we have seen,

$$[X, Y] = \mathrm{ad}X(Y).$$

So if G is an abelian group, then $c(g) : G \to G$ is the identity map for all $g \in G$. As a consequence, $\mathrm{ad}(X) \equiv 0$. It follows that the Lie algebra of an abelian Lie group is also abelian, i.e. $[X, Y] = 0$ for all $X, Y \in \mathfrak{g}$. Conversely, one can prove (using proposition 3.1 in lecture 6) that if G is connected and $\mathfrak{g}$ is abelian, then G is also abelian. In other words, the Lie bracket operation on $\mathfrak{g}$ measures the *non-commutativity* of the multiplication operation on G. In what follows we would like to characterize this quantitatively. In other words, we would like to find out the different between $\exp(X)\exp(Y)$ and $\exp(X + Y)$ for a general Lie group.

Let G be a Lie group and $X \in \mathfrak{g}$ a left invariant vector field on G. Then

$$(Xf)(a) = X_a f = dL_a X_e f = X_e(f \circ L_a) = \left.\frac{d}{dt}\right|_{t=0} f(a\exp(tX)).$$

for any $f \in C^\infty(G)$ and any $a \in G$. More generally, for any $t \in \mathbb{R}$,

$$(Xf)(a\exp(tX)) = \left.\frac{d}{ds}\right|_{s=0} f(a\exp(tX)\exp(sX)) = \left.\frac{d}{ds}\right|_{s=0} f(a\exp((t+s)X)) = \frac{d}{dt} f(a\exp(tX)).$$

Using this and induction, one can see that for any $k \geq 0$,

$$(X^k f)(a\exp(tX)) = \frac{d^k}{dt^k}\left(f(a\exp(tX))\right).$$

In particular,

$$(X^k f)(a) = \left.\frac{d^k}{dt^k}\right|_{t=0} f(a\exp(tX)).$$

The formulae above can be generalized to multi-variable case. In fact, if $X_1, \cdots, X_k \in \mathfrak{g}$, then

$$(X_1X_2f)(a) = \left.\frac{d}{dt_1}\right|_{t_1=0} (X_2f)(a\exp(t_1X_1)) = \left.\frac{d}{dt_1}\right|_{t_1=0} \left.\frac{d}{dt_2}\right|_{t_2=0} f(a\exp(t_1X_1)\exp(t_2X_2)),$$

and in general

$$(X_1 \cdots X_k f)(a) = \left.\frac{\partial^k}{\partial t_1 \cdots \partial t_k}\right|_{t_1=\cdots=t_k=0} f(a\exp(t_1X_1)\cdots\exp(t_kX_k)).$$

As a consequence, we get the following Taylor's expansion formula

Proposition 1.1. *If f is smooth on G, then for small $|t|$,*

$$f(\exp(tX_1)\cdots\exp(tX_n)) = f(e)+t\sum_i X_i f(e)+\frac{t^2}{2}\left\{\sum_i X_i^2 f(e)+2\sum_{i<j}X_iX_jf(e)\right\}+O(t^3).$$

We remark that the previous formulae hold for vector-valued functions as well. Our main result in this section is

Theorem 1.2. *Let $n \geq 1$ and $X_1, \cdots, X_n \in \mathfrak{g}$. Then for $|t|$ sufficiently small,*

$$\exp(tX_1)\cdots\exp(tX_n) = \exp(t\sum_{1\leq i\leq n} X_i + \frac{t^2}{2}\sum_{1\leq i<j\leq n}[X_i, X_j] + O(t^3)).$$

Proof. We apply proposition 1.1 to the inverse of the exponential map near e, i.e. the map f defined by

$$f(\exp(tX)) = tX$$

for t small enough. Then obviously, $f(e) = 0$. For any $X \in \mathfrak{g}$,

$$(Xf)(e) = \left.\frac{d}{dt}\right|_{t=0} f(\exp(tX)) = \left.\frac{d}{dt}\right|_{t=0}(tX) = X$$

and for any $n > 1$,

$$(X^n f)(e) = \left.\frac{d^n}{dt^n}\right|_{t=0} f(\exp(tX)) = \left.\frac{d^n}{dt^n}\right|_{t=0}(tX) = 0.$$

Notice

$$\sum_i X_i^2 + 2\sum_{i<j} X_iX_j = (X_1+\cdots+X_n)^2 + \sum_{i<j}[X_i, X_j],$$

it follows that

$$f(\exp(tX_1)\cdots\exp(tX_n)) = t\sum_i X_i + \frac{t^2}{2}\sum_{i<j}[X_i, X_j] + O(t^3).$$

On the other hand, by the definition of f,

$$\exp(tX_1)\cdots\exp(tX_n) = \exp(f(\exp(tX_1)\cdots\exp(tX_n))).$$

This completes the proof. □

In particular, we see

$$\exp(tX)\exp(tY) = \exp(tX + tY + \frac{t^2}{2}[X, Y] + O(|t|^3))$$

for $|t|$ small. So $[X, Y]$ dominates the difference between $\exp(X)\exp(Y)$ and $\exp(X + Y)$, and thus dominates the non-commutativity of the group multiplication.

2. The Baker-Campbell-Hausdorff Formula

Now the question is: What are the higher order terms in $O(t^3)$ above? For simplicity we will denote by log the inverse of exp near $0 \in \mathfrak{g}$. Let

$$\mu(X,Y) = \log(\exp(X)\exp(Y))$$

for X, Y close to $0 \in \mathfrak{g}$. We have seen above that

$$\mu(X,Y) = X + Y + \frac{1}{2}[X,Y] + O(|X|^3, |Y|^3)$$

for $|X|, |Y|$ small. A remarkable fact about the remainder terms is that they involves only Lie brackets! In other words, we have

Theorem 2.1 (The Baker-Campbell-Hausdorff formula (existence))**.** *For X and Y small, we have*

$$\mu(X,Y) = X + Y + \sum_{m \geq 2} P_m(X,Y),$$

where $P_m(A,B)$ is a Lie polynomial *of order m, i.e. $P_m(X,Y)$ is a combination of nested commutators in X, Y that involves $m-1$ Lie brackets.*

Although in application, the above existence result is sufficient, we will prove the following explicit formula:

Theorem 2.2 (Dynkin's formula)**.** *For X and Y small,*

$$\mu(X,Y) = X+Y+\sum_{k=1}^{\infty} \frac{(-1)^k}{k+1} \sum \frac{(-1)^{\sum_i (l_i+m_i)}}{l_1+\cdots+l_k+1} \frac{(\mathrm{ad}Y)^{l_1}}{l_1!} \circ \frac{(\mathrm{ad}X)^{m_1}}{m_1!} \circ \cdots \circ \frac{(\mathrm{ad}Y)^{l_k}}{l_k!} \circ \frac{(\mathrm{ad}X)^{m_k}}{m_k!}(Y),$$

where the second summation is over $l_1, \cdots, l_k, m_1, \cdots, m_k \geq 0, l_i + m_i > 0$.

As a consequence, we can write down $P_m(X,Y)$ for m small. Of course,

$$P_2(X,Y) = \frac{1}{2}[X,Y].$$

The next term $P_3(X,Y)$ comes from the following terms in Dynkins's formula:

(1) $k = 1, l_1 = 1, m_1 = 1 \Longrightarrow -\frac{1}{2}\frac{1}{2}[Y,[X,Y]]$;
(2) $k = 1, l_1 = 0, m_1 = 2 \Longrightarrow -\frac{1}{2}\frac{1}{1}\frac{1}{2}[X,[X,Y]]$;
(3) $k = 2, l_1 = 1, m_1 = 0, l_2 = 0, m_2 = 1 \Longrightarrow \frac{1}{3}\frac{1}{2}[Y,[X,Y]]$;
(4) $k = 2, l_1 = 0, m_1 = 1, l_2 = 0, m_2 = 1 \Longrightarrow \frac{1}{3}\frac{1}{1}[X,[X,Y]]$;

It follows

$$P_3(X,Y) = \frac{1}{12}([X,[X,Y]] - [Y,[X,Y]]).$$

Similarly one can calculate the next term and get

$$P_4(X,Y) = \frac{1}{24}[X,[Y,[Y,X]]].$$

To prove the Dynkin's formula, we will need the following formula that computes the differential of the exponential map at an arbitrary point.

Lemma 2.3. *For each $X \in \mathfrak{g}$,*

$$(d\exp)_X = (dL_{\exp X})_e \circ \phi(\mathrm{ad}X),$$

where ϕ is the function

$$\phi(z) = \frac{1-e^{-z}}{z} = \sum_{m=0}^{\infty} \frac{(-1)^m}{(m+1)!} z^m.$$

Proof of Dynkin's formula.

Write

$$Z(t) = \log(\exp(X)\exp(tY)).$$

Applying lemma 2.3, we get

$$\frac{d}{dt}(\exp Z(t)) = dL_{\exp X}\frac{d}{dt}(\exp tY) = dL_{\exp X} dL_{\exp tY}\phi(\mathrm{ad}(tY))(Y) = dL_{\exp Z(t)}(Y),$$

where we used the fact $\phi(\mathrm{ad}(tY))(Y) = Y$. On the other hand, by using lemma 2.3 directly,

$$\frac{d}{dt}(\exp Z(t)) = dL_{\exp Z(t)}\phi(\mathrm{ad}Z(t))\frac{dZ}{dt}.$$

It follows

$$\frac{dZ}{dt} = \frac{\mathrm{ad}Z(t)}{I - \exp(-\mathrm{ad}Z(t))}(Y) = \sum_{k\geq 0} \frac{1}{k+1}(I - \exp(-\mathrm{ad}Z(t)))^k(Y).$$

Notice that by the naturality of exp and by the definition of ad and Ad,

$$\begin{aligned} \exp(-\mathrm{ad}Z(t)) &= \mathrm{Ad}\exp(-Z(t)) = \mathrm{Ad}(\exp(-tY)\exp(-X)) \\ &= \mathrm{Ad}(\exp(-tY)) \circ \mathrm{Ad}(\exp(-X)) \\ &= \exp(-t\mathrm{ad}(Y)) \circ \exp(-\mathrm{ad}(X)). \end{aligned}$$

Thus

$$\begin{aligned} \frac{dZ}{dt} &= \sum_{k\geq 0} \frac{(I - \exp(-t\mathrm{ad}Y)\circ\exp(-\mathrm{ad}X))^k}{k+1}(Y) \\ &= \sum_{k\geq 0} \frac{(-1)^k}{k+1} \sum_{l_1,\cdots,l_k,m_1,\cdots,m_k\geq 0, l_i+m_i>0} t^{|l|}(-1)^{|l|+|m|}\frac{(\mathrm{ad}Y)^{l_1}}{l_1!}\frac{(\mathrm{ad}X)^{m_1}}{m_1!}\cdots\frac{(\mathrm{ad}Y)^{l_k}}{l_k!}\frac{(\mathrm{ad}X)^{m_k}}{m_k!}Y. \end{aligned}$$

where in the last step we used the fact that $\mathrm{ad}X \in \mathrm{End}(\mathfrak{g})$ is an element in a linear Lie group, and thus the exponential map is exactly the matrix exponential. Now the Dynkin's formula follows from termwise integration over t from 0 to 1. $\square$

3. The Derivative of the Exponential Map

Finally we prove lemma 2.3. We first show

Lemma 3.1. *Let $\gamma_1(t), \gamma_2(t)$ be smooth curves on G, and let $\gamma(t) = \gamma_1(t)\gamma_2(t)$, then*

$$\dot{\gamma}(t) = dL_{\gamma_1(t)}(\dot{\gamma}_2(t)) + dR_{\gamma_2(t)}(\dot{\gamma}_1(t)).$$

Proof. Notice the fact $\gamma(t) = \mu(\gamma_1(t), \gamma_2(t))$, where μ is the multiplication operation on G. So the formula above follows from the following formula we have proven,

$$d\mu_{a,b}(X_a, Y_b) = (dL_a)_b(Y_a) + (dR_b)_a(X_a).$$

□

More generally, by using induction one can easily see that if $\gamma_1(t), \cdots, \gamma_m(t)$ are smooth curves on G, and let $\gamma(t) = \gamma_1(t) \cdots \gamma_m(t)$, then

$$\dot{\gamma}(t) = \sum_{k=1}^{m} dL_{\gamma_1(t)} \cdots dL_{\gamma_{k-1}} dR_{\gamma_{k+1}(t)} \cdots dR_{\gamma_m(t)}(\dot{\gamma}_k(t)).$$

Now we are ready to prove lemma 2.3. For simplicity we will denote

$$\nu(X, Y) := \left.\frac{d}{dt}\right|_{t=0} \exp(X + tY) = (d\exp)_X(Y).$$

Obviously $\nu(X, Y)$ is linear in Y for each fixed X, and lemma 2.3 follows from

Lemma 3.2. *For any $X, Y \in \mathfrak{g}$,*

$$\left.\frac{d}{dt}\right|_{t=0} \exp(X + tY) = (dL_{\exp X})_e \circ \phi(\mathrm{ad}X)(Y).$$

Proof. We notice that for any positive integer m,

$$\begin{aligned}
\nu(X, Y) &= \left.\frac{d}{dt}\right|_{t=0} \left[\exp(\frac{X}{m} + t\frac{Y}{m})\right]^m \\
&= \sum_{k=0}^{m-1} (dL_{\exp\frac{X}{m}})^{m-k-1}(dR_{\exp\frac{X}{m}})^k \nu(\frac{X}{m}, \frac{Y}{m}) \\
&= \frac{1}{m}(dL_{\exp\frac{X}{m}})^{m-1} \sum_{k=0}^{m-1} (dL_{\exp\frac{X}{m}})^{-k}(dR_{\exp\frac{X}{m}})^k \nu(\frac{X}{m}, Y).
\end{aligned}$$

Recall that the differential of the conjugation map $c(a) = L_a R_{a^{-1}}$ is Ad, so we get

$$(dL_{\exp\frac{X}{m}})^{-k}(dR_{\exp\frac{X}{m}})^k = \left[dc(\exp(\frac{-X}{m}))\right]^k = \left[\mathrm{Ad}(\exp\frac{-X}{m})\right]^k = \left[\exp(-\frac{\mathrm{ad}X}{m})\right]^k.$$

So we get, for every positive integer m,

$$\nu(X,Y) = (dL_{\exp\frac{X}{m}})^{m-1}\frac{1}{m}\sum_{k=0}^{m-1}\left[\exp(-\frac{\mathrm{ad}X}{m})\right]^k \nu(\frac{X}{m},Y).$$

Now the result follows since as $m\to\infty$,

$$(dL_{\exp\frac{X}{m}})^{m-1} = dL_{\exp\frac{(m-1)X}{m}} \to dL_{\exp X},$$
$$\nu(\frac{X}{m},Y)\to\nu(0,Y) = (d\exp)_0(Y) = Y,$$

and, since $\mathrm{ad}X\in\mathrm{End}(\mathfrak{g})$ is a matrix,

$$\begin{aligned}\frac{1}{m}\sum_{k=0}^{m-1}\left[\exp(-\frac{\mathrm{ad}X}{m})\right]^k &= \frac{1}{m}\sum_{k=0}^{m-1}\exp\left(-\frac{k}{m}\mathrm{ad}X\right)\\ &= \frac{1}{m}\sum_{k=0}^{m-1}\sum_{n=0}^{\infty}\frac{1}{n!}\left(-\frac{k}{m}\mathrm{ad}X\right)^n\\ &= \sum_{n=0}^{\infty}\left[\frac{1}{m}\sum_{k=0}^{m-1}\left(\frac{k}{m}\right)^n\right]\frac{(-1)^n}{n!}(\mathrm{ad}X)^n\\ &\to \sum_{n=0}^{\infty}\left[\int_0^1 x^n dx\right]\frac{(-1)^n}{n!}(\mathrm{ad}X)^n\\ &= \sum_{n=0}^{\infty}\frac{(-1)^n}{(n+1)!}(\mathrm{ad}X)^n.\end{aligned}$$

□

Finally we give several applications. Given the derivative of the exponential map exp at an arbitrary point, we are ready to answer the following question: at which points X the map exp is singular, i.e. $(d\exp)_X$ is not invertible? Since

$$(d\exp)_X = (dL_{\exp X})_e\circ\phi(\mathrm{ad}X)$$

and $dL_{\exp X})_e$ is always invertible, we see that $(d\exp)_X$ is not invertible if and only if the matrix $\phi(\mathrm{ad}X)\in\mathrm{End}(\mathfrak{g})$ is not invertible, i.e. 0 is not an eigenvalue of $\phi(\mathrm{ad}X)$. Since all eigenvalues of $\phi(\mathrm{ad}X)$ are of the form $\phi(\lambda) = \frac{1-e^{-\lambda}}{\lambda}$, where λ is an eigenvalue of $\mathrm{ad}X\in\mathrm{End}(\mathfrak{g})$, we conclude

Corollary 3.3. *The singular points of the exponential map* $\exp:\mathfrak{g}\to G$ *are precisely those* $X\in\mathfrak{g}$ *such that* $\mathrm{ad}X\in\mathrm{End}(\mathfrak{g})$ *has an eigenvalue of the form* $2\pi ik$*, with* $k\in\mathbb{Z}\setminus\{0\}$.

As an example, we see that if G is an abelian Lie group, then exp is non-singular everywhere. More generally, if $\mathfrak{g}$ is nilpotent, then $\exp:\mathfrak{g}\to G$ is non-singular everywhere.