Math 128 Lecture 2.

Geometric proof of the Campbell-Baker-Hausdorff formula.

The formula.

The Campbell-Baker-Hausdorff formula says that

$$\log((\exp A)(\exp B)) = A + \int_0^1 \psi\left((\exp \operatorname{ad} A)(\exp t \operatorname{ad} B)\right) B\,dt. \quad (2)$$

where

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

or as formal power series.

$$\psi(1+u) = (1+u)\frac{\log(1+u)}{u} = (1+u)(1-\frac{u}{2}+\frac{u^2}{3}+\cdots) = 1+\frac{u}{2}-\frac{u^2}{6}+\cdots.$$

Strategy of the proof.

$$\log((\exp A)(\exp B)) = A + \int_0^1 \psi\left((\exp \mathrm{ad}\ A)(\exp t \mathrm{ad}\ B)\right) B dt. \quad (2)$$

Our strategy for the proof of (2) will be to prove a differential version of it:

$$\frac{d}{dt} \log\left((\exp A)(\exp tB)\right) = \psi\left((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)\right) B. \qquad (4)$$

Since $\log \exp A = A$ when $t = 0$, integrating (4) from 0 to 1 will prove (2). Let us define $\Gamma = \Gamma(t) = \Gamma(t, A, B)$ by

$$\Gamma = \log\left((\exp A)(\exp tB)\right). \qquad (5)$$

Then

$$\exp \Gamma = \exp A \exp tB$$

We want to prove

$$\frac{d}{dt}\log\left((\exp A)(\exp tB)\right) = \psi\left((\exp \operatorname{ad} A)(\exp t \operatorname{ad} B)\right)B. \tag{4}$$

Let

$$\Gamma = \log\left((\exp A)(\exp tB)\right).$$

so

$$\exp\Gamma = \exp A \exp tB$$

$$\begin{aligned}\frac{d}{dt}\exp\Gamma(t) &= \exp A\frac{d}{dt}\exp tB\\ &= \exp A(\exp tB)B\\ &= (\exp\Gamma(t))B \quad \text{so}\\ (\exp -\Gamma(t))\frac{d}{dt}\exp\Gamma(t) &= B.\end{aligned}$$

We will prove (4) by finding a general expression for

$$\exp(-C(t))\frac{d}{dt}\exp(C(t))$$

where $C = C(t)$ is a curve in the Lie algebra, $\mathbf{g}$, see (11) below.

Ad .

For any $g \in G$, define the linear transformation

$$\mathrm{Ad}\ g : \mathbf{g} \to \mathbf{g} : X \mapsto gXg^{-1}.$$

(In geometrical terms, this can be thought of as follows: (The differential of) Left multiplication by g carries $\mathbf{g} = T_I(G)$ into the tangent space, $T_g(G)$ to G at the point g. Right multiplication by g^{-1} carries this tangent space back to $\mathbf{g}$ and so the combined operation is a linear map of $\mathbf{g}$ into itself which we call Ad g. Notice that Ad is a representation in the sense that

$$\mathrm{Ad}\ (gh) = (\mathrm{Ad}\ g)(\mathrm{Ad}\ h) \quad \forall g, h \in G.$$

Ad and ad.

In particular, for any $A \in \mathbf{g}$, we have the one parameter family of linear transformations $\mathrm{Ad}(\exp tA)$ and

$$\begin{aligned}
\frac{d}{dt}\mathrm{Ad}\ (\exp tA)X &= (\exp tA)AX(\exp -tA) + (\exp tA)X(-A)(\exp -tA) \\
&= (\exp tA)[A,X](\exp -tA) \quad \text{so} \\
\frac{d}{dt}\mathrm{Ad}\ \exp tA &= \mathrm{Ad}(\exp tA) \circ \mathrm{ad}\ A.
\end{aligned}$$

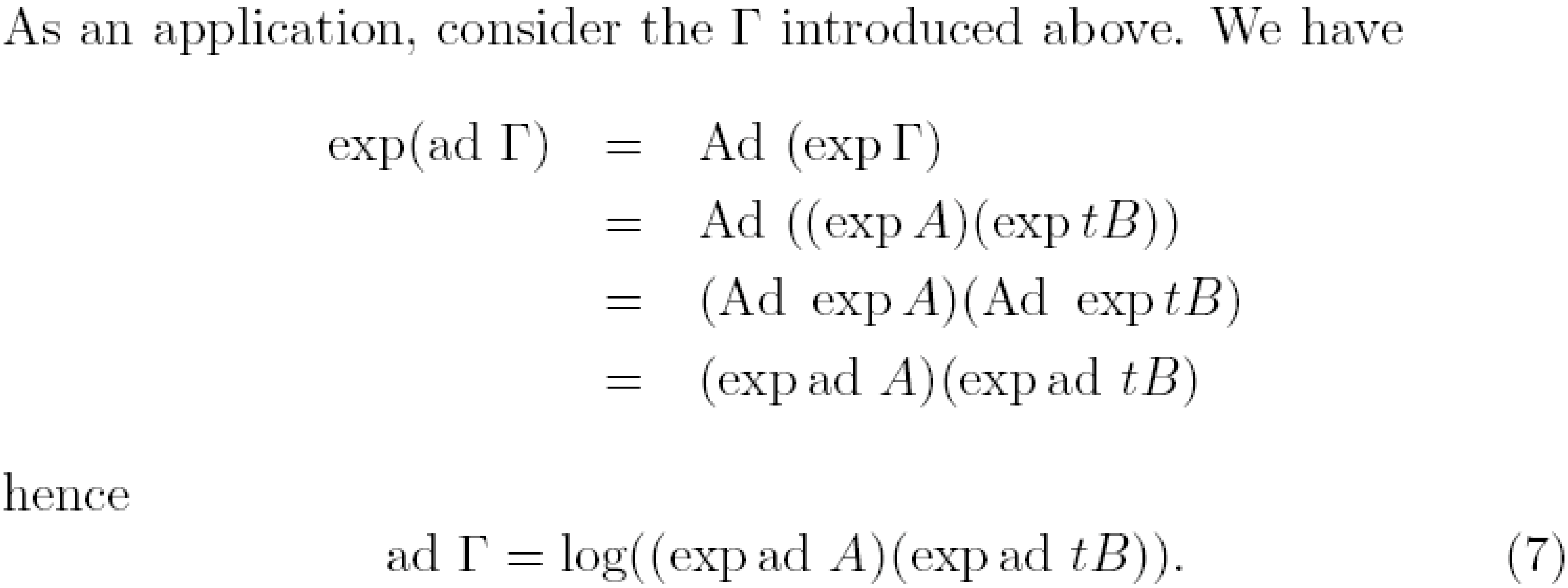
But $\mathrm{ad}\ A$ is a linear transformation acting on $\mathbf{g}$ and the solution to the differential equation

$$\frac{d}{dt}M(t) = M(t)\mathrm{ad}\ A, \quad M(0) = I$$

(in the space of linear transformations of $\mathbf{g}$) is $\exp t\,\mathrm{ad}\,A$. Thus $\mathrm{Ad}(\exp tA) = \exp(t\,\mathrm{ad}\,A)$. Setting $t = 1$ gives the important formula

$$\mathrm{Ad}\ (\exp A) = \exp(\mathrm{ad}\ A). \tag{6}$$

An application of exp(ad)=Ad(exp).



As an application, consider the Γ introduced above. We have

$$\begin{aligned}\exp(\mathrm{ad}\,\Gamma) &= \mathrm{Ad}\;(\exp\Gamma)\\ &= \mathrm{Ad}\;((\exp A)(\exp tB))\\ &= (\mathrm{Ad}\;\exp A)(\mathrm{Ad}\;\exp tB)\\ &= (\exp\mathrm{ad}\;A)(\exp\mathrm{ad}\;tB)\end{aligned}$$

hence

$$\mathrm{ad}\;\Gamma = \log((\exp\mathrm{ad}\;A)(\exp\mathrm{ad}\;tB)). \qquad (7)$$

The Maurer-Cartan form on a Lie group.

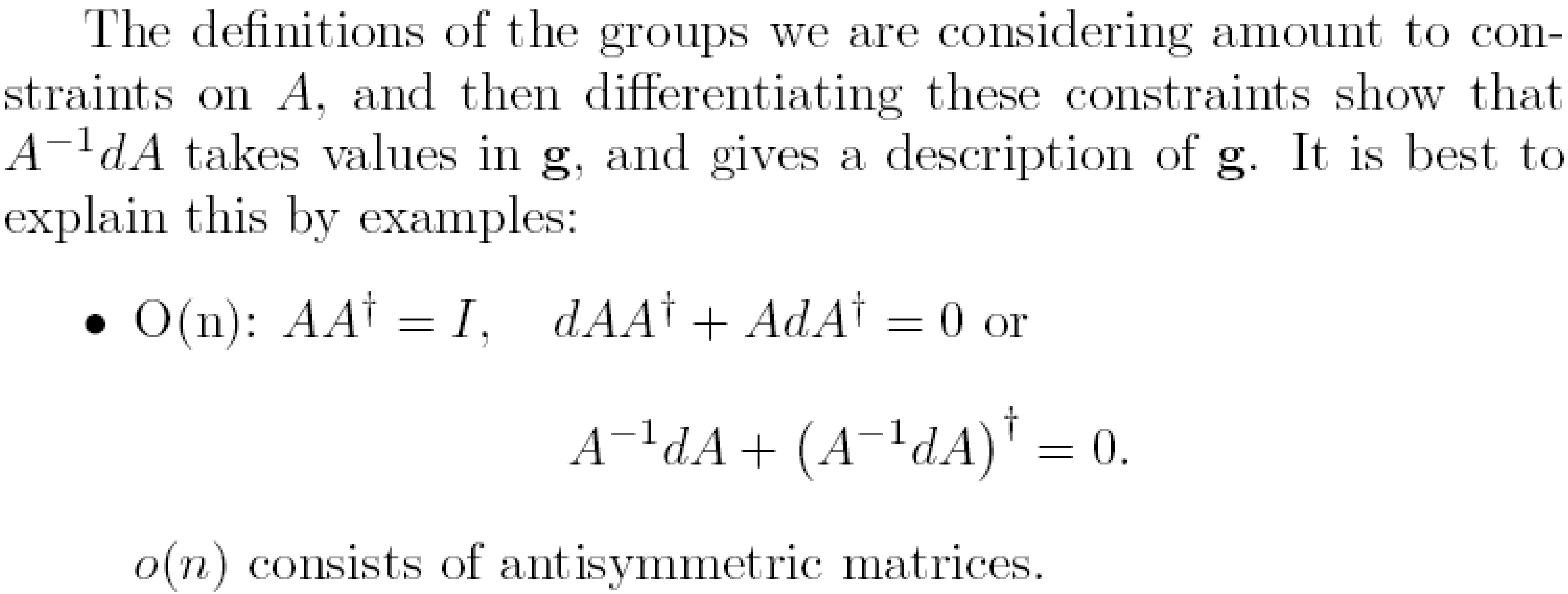
If G is a Lie group and $\gamma = \gamma(t)$ is a curve on G with $\gamma(0) = A \in G$, then $A^{-1}\gamma$ is a curve which passes through the identity at $t = 0$. Hence $A^{-1}\gamma'(0)$ is a tangent vector at the identity, i.e. an element of $\mathbf{g}$, the Lie algebra of G.

In this way, we have defined a linear differential form θ on G with values in $\mathbf{g}$. In case G is a subgroup of the group of all invertible $n \times n$ matrices (say over the real numbers), we can write this form as

$$\theta = A^{-1}dA.$$

We can then think of the A occurring above as a collection of n^2 real valued functions on G (the matrix entries considered as functions on the group) and dA as the matrix of differentials of these functions. The above equation giving θ is then just matrix multiplication. For simplicity, we will work in this case, although the main theorem, equation (8) below, works for any Lie group and is quite standard.

Examples of linear Lie algebras and their Maurer-Cartan forms.



The definitions of the groups we are considering amount to constraints on A, and then differentiating these constraints show that $A^{-1}dA$ takes values in $\mathbf{g}$, and gives a description of $\mathbf{g}$. It is best to explain this by examples:

- O(n): $AA^\dagger = I, \quad dAA^\dagger + AdA^\dagger = 0$ or

$$A^{-1}dA + (A^{-1}dA)^\dagger = 0.$$

$o(n)$ consists of antisymmetric matrices.

Examples, continued.

- Sp(n): Let

$$J := \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}$$

and let Sp(n) consist of all matrices satisfying

$$AJA^{\dagger} = J.$$

Then

$$dAJa^{\dagger} + AJdA^{\dagger} = 0$$

or

$$(A^{-1}dA)J + J(A^{-1}dA)^{\dagger} = 0.$$

The equation $BJ + JB^{\dagger} = 0$ defines the Lie algebra sp(n).

Examples, continued.

- Let J be as above and define Gl(n,C) to consist of all invertible matrices satisfying

  $$AJ = JA.$$

  Then

  $$dAJ = JdA = 0.$$

  and so

  $$A^{-1}dAJ = A^{-1}JdA = JA^{-1}dA.$$

The Maurer Cartan equation.

We return to general considerations: Let us take the exterior derivative of the defining equation $\theta = A^{-1}dA$. For this we need to compute $d(A^{-1})$: Since

$$d(AA^{-1}) = 0$$

we have

$$dA \cdot A^{-1} + Ad(A^{-1}) = 0$$

or

$$d(A^{-1}) = -A^{-1}dA \cdot A^{-1}.$$

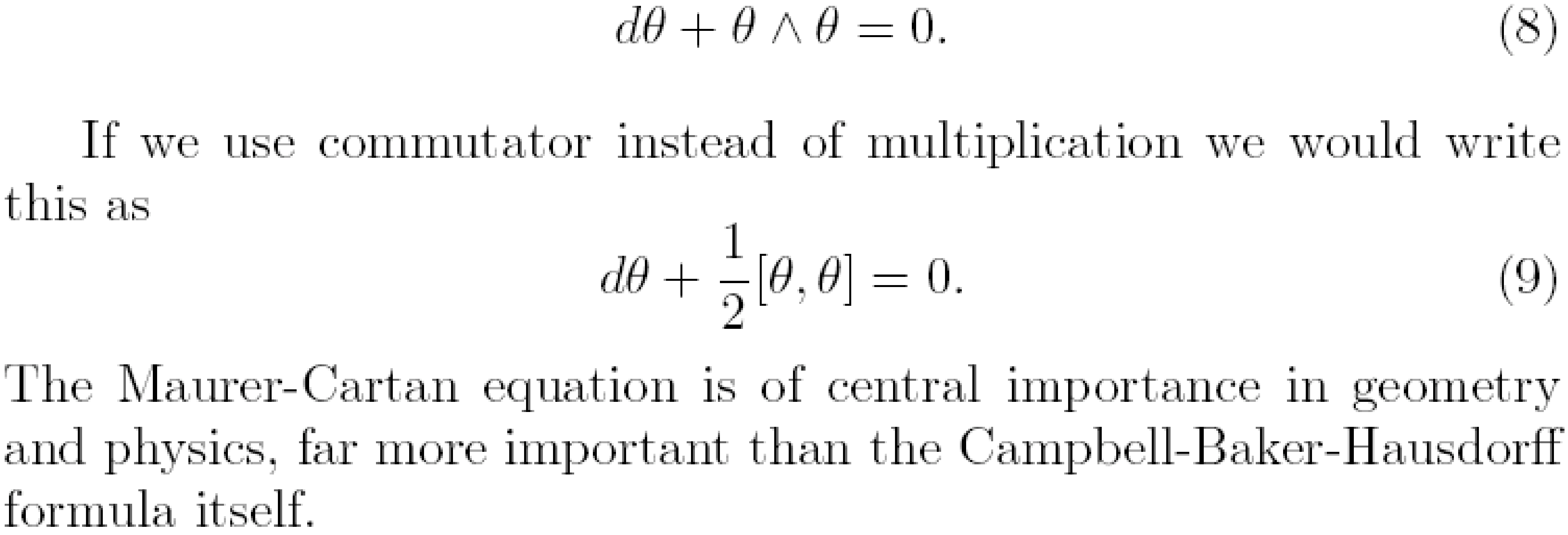
This is the generalization to matrices of the formula in elementary calculus for the derivative of $1/x$. Using this formula we get

$$d\theta = d(A^{-1}dA) = -(A^{-1}dA \cdot A^{-1}) \wedge dA = -A^{-1}dA \wedge A^{-1}dA$$

or the **Maurer-Cartan equation**

$$d\theta + \theta \wedge \theta = 0. \tag{8}$$

The Maurer Cartan equation, again.



$$d\theta + \theta \wedge \theta = 0. \tag{8}$$

If we use commutator instead of multiplication we would write this as

$$d\theta + \frac{1}{2}[\theta, \theta] = 0. \tag{9}$$

The Maurer-Cartan equation is of central importance in geometry and physics, far more important than the Campbell-Baker-Hausdorff formula itself.

Pulling back the M-C form to a plane.

Suppose we have a map $g : \mathbf{R}^2 \to G$, with s,t coordinates on the plane. Pull θ back to the plane, so

$$g^*\theta = g^{-1}\frac{\partial g}{\partial s}ds + g^{-1}\frac{\partial g}{\partial t}dt$$

Define

$$\alpha = \alpha(s,t) := g^{-1}\frac{\partial g}{\partial s}$$

and

$$\beta := \beta(s,t) = g^{-1}\frac{\partial g}{\partial t}$$

so that

$$g^*\theta = \alpha ds + \beta dt.$$

Pulling back the M-C equations.

$$g^*\theta = g^{-1}\frac{\partial g}{\partial s}ds + g^{-1}\frac{\partial g}{\partial t}dt$$

$$\alpha = \alpha(s,t) := g^{-1}\frac{\partial g}{\partial s}$$

$$\beta := \beta(s,t) = g^{-1}\frac{\partial g}{\partial t}$$

so that

$$g^*\theta = \alpha ds + \beta dt.$$

Then collecting the coefficient of $ds \wedge dt$ in the Maurer Cartan equation gives

$$\frac{\partial \beta}{\partial s} - \frac{\partial \alpha}{\partial t} + [\alpha, \beta] = 0. \tag{10}$$

This is the version of the Maurer Cartan equation we shall use in our proof of the Campbell Baker Hausdorff formula. Of course this version is completely equivalent to the general version, since a two form is determined by its restriction to all two dimensional surfaces.

Applying the M-C equations.

$$\frac{\partial \beta}{\partial s} - \frac{\partial \alpha}{\partial t} + [\alpha, \beta] = 0. \qquad (10)$$

Let $C(t)$ be a curve in the Lie algebra $\mathbf{g}$ and let us apply (10) to

$$g(s,t) := \exp[sC(t)]$$

so that

$$\begin{aligned} \alpha(s,t) &= g^{-1}\frac{\partial g}{\partial s} \\ &= \exp[-sC(t)]\exp[sC(t)]C(t) \\ &= C(t) \\ \beta(s,t) &= g^{-1}\frac{\partial g}{\partial t} \\ &= \exp[-sC(t)]\frac{\partial}{\partial t}\exp[sC(t)] \text{ so by (10)} \\ \frac{\partial \beta}{\partial s} - C'(t) + [C(t), \beta] &= 0. \end{aligned}$$

$$\frac{\partial \beta}{\partial s} - C'(t) + [C(t), \beta] \quad = \quad 0.$$

For fixed t consider the last equation as the differential equation (in s)

$$\frac{d\beta}{ds} = -(\text{ad } C)\beta + C', \quad \beta(0) = 0$$

where $C := C(t),\ C' := C'(t)$.

If we expand $\beta(s,t)$ as a formal power series in s (for fixed t):

$$\beta(s,t) = a_1 s + a_2 s^2 + a_3 s^3 + \cdots$$

and compare coefficients in the differential equation we obtain $a_1 = C'$, and

$$n a_n = -(\text{ad } C) a_{n-1}$$

or

$$\beta(s,t) = sC'(t) + \frac{1}{2} s(-\text{ad } C(t))C'(t) + \cdots + \frac{1}{n!} s^n (-\text{ad } C(t))^{n-1} C'(t) + \cdots$$

$$\beta(s,t) = sC'(t) + \frac{1}{2}s(-\mathrm{ad}\ C(t))C'(t) + \cdots + \frac{1}{n!}s^n(-\mathrm{ad}\ C(t))^{n-1}C'(t) + \cdots$$

If we define

$$\phi(z) := \frac{e^z - 1}{z} = 1 + \frac{1}{2!}z + \frac{1}{3!}z^2 + \cdots$$

and set $s = 1$ in the expression we derived above for $\beta(s,t)$ we get

$$\exp(-C(t))\frac{d}{dt}\exp(C(t)) = \phi(-\mathrm{ad}\ C(t))C'(t). \tag{11}$$

Now to the proof of the Campbell-Baker-Hausdorff formula. Suppose that A and B are chosen sufficiently near the origin so that

$$\Gamma = \Gamma(t) = \Gamma(t, A, B) := \log((\exp A)(\exp tB))$$

is defined for all $|t| \leq 1$. Then, as we remarked,

$$\exp \Gamma = \exp A \exp tB$$

so $\exp \mathrm{ad}\ \Gamma = (\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)$ and hence

$$\mathrm{ad}\ \Gamma = \log\left((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)\right).$$

$$\exp(-C(t))\frac{d}{dt}\exp(C(t)) = \phi(-\mathrm{ad}\ C(t))C'(t). \tag{11}$$

$$\Gamma = \Gamma(t) = \Gamma(t, A, B) := \log((\exp A)(\exp tB))$$

$\exp\Gamma = \exp A \exp tB$ **so** $\mathrm{ad}\ \Gamma = \log\left((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)\right).$

$$\begin{aligned}
\frac{d}{dt}\exp\Gamma(t) &= \exp A \frac{d}{dt}\exp tB \\
&= \exp A(\exp tB)B \\
&= (\exp\Gamma(t)B \quad \text{so} \\
(\exp -\Gamma(t))\frac{d}{dt}\exp\Gamma(t) &= B \text{ and therefore} \\
\phi(-\mathrm{ad}\ \Gamma(t))\Gamma'(t) &= B \quad \text{by (11) so} \\
\phi(-\log\left((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)\right))\Gamma'(t) &= B.
\end{aligned}$$

Conclusion of proof.

$$\phi(z) := \frac{e^z - 1}{z} \qquad \Gamma(t) = \Gamma(t, A, B) := \log((\exp A)(\exp tB))$$

$$\phi(-\log((\exp \mathrm{ad}\ A)(\exp t\ \mathrm{ad}\ B)))\Gamma'(t) \quad = \quad B.$$

Now for $|z - 1| < 1$

$$\begin{aligned} \phi(-\log z) &= \frac{e^{-\log z} - 1}{-\log z} \\ &= \frac{z^{-1} - 1}{-\log z} \\ &= \frac{z - 1}{z \log z} \quad \text{so} \\ \psi(z)\phi(-\log z) &\equiv 1 \quad \text{where } \psi(z) := \frac{z \log z}{z - 1} \text{ so} \\ \Gamma'(t) &= \psi((\exp \mathrm{ad}\ A)(\exp t \mathrm{ad}\ B))\, B. \qquad (4) \end{aligned}$$

This is what we had to prove. QED

The differential of the exponential and its inverse.

$$\exp(-C(t))\frac{d}{dt}\exp(C(t)) = \phi(-\mathrm{ad}\ C(t))C'(t). \qquad (11)$$

Once again, equation (11), which we derived from the Maurer-Cartan equation, is of significant importance in its own right, perhaps more than the use we made of it - to prove the Campbell-Baker-Hausdorff theorem. We will rewrite this equation in terms of more familiar geometric operations, but first some preliminaries:

The exponential map exp sends the Lie algebra $\mathbf{g}$ into the corresponding Lie group, and is a differentiable map. If $\xi \in \mathbf{g}$ we can consider the differential of exp at the point ξ:

$$d(\exp)_\xi : \mathbf{g} = T\mathbf{g}_\xi \to TG_{\exp\xi}$$

where we have identified $\mathbf{g}$ with its tangent space at ξ which is possible since $\mathbf{g}$ is a vector space. In other words, $d(\exp)_\xi$ maps the tangent space to $\mathbf{g}$ at the point ξ into the tangent space to G at the point $\exp(\xi)$.

At $\xi = 0$ we have

$$d(\exp)_0 = \mathrm{id}$$

and hence, by the implicit function theorem, $d(\exp)_\xi$ is invertible for sufficiently small ξ. Now the Maurer-Cartan form, evaluated at the point $\exp\xi$ sends $TG_{\exp\xi}$ back to $\mathbf{g}$:

$$\theta_\xi : TG_{\exp\xi} \to \mathbf{g}.$$

Hence

$$\theta_\xi \circ d(\exp)_\xi : \mathbf{g} \to \mathbf{g}$$

and is invertible for sufficiently small ξ. We claim that

$$\tau(\mathrm{ad}\,\xi) \circ \left(\theta_{\exp\xi} \circ d(\exp_\xi)\right) = \mathrm{id} \tag{12}$$

where τ is as defined above in (3). Indeed, we claim that (12) is an immediate consequence of (11).

Recall the definition (3) of the function τ as $\tau(z) = 1/\phi(-z)$.

$$\tau(w) := \frac{w}{1 - e^{-w}}. \tag{3}$$

$\tau(w) := \frac{w}{1-e^{-w}}$ (3) so $\tau(z) = 1/\phi(-z)$ where $\phi(z) := \frac{e^z-1}{z}$

$$\exp(-C(t))\frac{d}{dt}\exp(C(t)) = \phi(-\mathrm{ad}\, C(t))C'(t). \qquad (11)$$

Multiply both sides of (11) by $\tau(\mathrm{ad}\, C(t))$ to obtain

$$\tau(\mathrm{ad}\, C(t))\exp(-C(t))\frac{d}{dt}\exp(C(t)) = C'(t). \qquad (13)$$

Choose the curve C so that $\xi = C(0)$ and $\eta = C'(0)$. Then the chain rule says that

$$\frac{d}{dt}\exp(C(t))_{|t=0} = d(\exp)_\xi(\eta).$$

Thus

$$\left(\exp(-C(t))\frac{d}{dt}\exp(C(t))\right)_{|t=0} = \theta_{\exp\xi}d(\exp)_\xi\eta,$$

the result of applying the Maurer-Cartan form θ (at the point $\exp(\xi)$) to the image of η under the differential of exponential map at $\xi \in \mathbf{g}$.

$$\tau(\operatorname{ad} C(t)) \exp(-C(t)) \frac{d}{dt} \exp(C(t)) = C'(t). \tag{13}$$

$$\left(\exp(-C(t)) \frac{d}{dt} \exp(C(t)) \right)_{|t=0} = \theta_{\exp \xi} d(\exp)_\xi \eta,$$

Then (13) at $t = 0$ translates into (12)

$$\tau(\operatorname{ad} \xi) \circ \left(\theta_{\exp \xi} \circ d(\exp_\xi)\right) = \mathrm{id} \tag{12}$$

A mystery.

$$\tau(\operatorname{ad}\xi) \circ (\theta_{\exp\xi} \circ d(\exp_\xi)) = \operatorname{id} \qquad (12)$$

In a sense, this formula is taking into account the non-commutativity of the group multiplication, so τ is helping to relate the non-commutative to the commutative.

But much earlier in mathematical history, τ was introduced to relate the discrete to the continuous: if we set

$$D_1 := \frac{d}{dh_1}, \quad D_2 := \frac{d}{dh_2},$$

then the Euler MacLaurin formula (see the notes) says that

$$\tau(D_1)\tau(D_2) \int_{a-h_1}^{b+h_2} f(x)dx \Big|_{h_1=h_2=0} = \sum_{k=a}^{b} f(k), \qquad (18)$$

for all functions of the form $x \mapsto p(x)e^{zx}$ where p is a polynomial and $|z| < 2\pi$.

What is the relation between these two roles ?