5. The Matrix Exponential

Definition and basic properties • Some useful formulas • The product and commutator formulas (optional) • The adjoint action.

5.1. **Definition and basic properties.** The exponential of a matrix plays a crucial role in the study of linear (Lie) groups. (It is the mechanism for passing information from the *Lie algebra* to the Lie group.) Let $A \in \Bbbk^{n\times n}$ and consider the *matrix series*

$$\sum_{k\geq 0} \frac{1}{k!}A^k = \mathbf{1} + A + \frac{1}{2!}A^2 + \frac{1}{3!}A^3 + \cdots$$

Note : This matrix series is a series in the complete normed vector space (in fact, algebra) $(\Bbbk^{n\times n}, \|\cdot\|)$, where $\|\cdot\|$ is the *operator norm* (induced by the Euclidean norm on $\Bbbk^n$). In a complete normed vector space, an absolutely convergent series $\sum_{k\geq 0} a_k$ (i.e., such that the series $\sum_{k\geq 0} \|a_k\|$ is convergent) is convergent, and

$$\left\| \sum_{k=0}^{\infty} a_k \right\| \leq \sum_{k=0}^{\infty} \|a_k\|.$$

(The converse is not true.) Also, every *rearrangement* of an absolutely convergent series is absolutely convergent, with same sum. Given two absolutely convergent series $\sum_{k\geq 0} a_k$ and $\sum_{k\geq 0} b_k$ (in a complete normed algebra), their *Cauchy product* $\sum_{k\geq 0} c_k$, where $c_k = \sum_{i+j=k} a_i b_j = a_0 b_k + a_1 b_{k-1} + \cdots + a_k b_0$ is also absolutely convergent, and

$$\sum_{k=0}^{\infty} c_k = \left(\sum_{k=0}^{\infty} a_k\right)\left(\sum_{k=0}^{\infty} b_k\right).$$

◇ **Exercise 71.** Show that the matrix series $\sum_{k\geq 0} \frac{1}{k!}A^k$ is *absolutely convergent.*

Let $\sum_{k=0}^{\infty} \frac{1}{k!}A^k$ denote the *sum* of the (absolutely) convergent matrix series $\sum_{k\geq 0} \frac{1}{k!}A^k$. We set

$$e^A = \exp(A) := \sum_{k=0}^{\infty} \frac{1}{k!}A^k.$$

This matrix is called the **matrix exponential** of A. Clearly, $\exp(0) = \mathbf{1}$. It follows that

$$\|\exp(A)\| \leq \|\mathbf{1}\| + \|A\| + \frac{1}{2!}\|A\|^2 + \cdots = e^{\|A\|}.$$

$\diamond$ **Exercise 72.** Given $A \in \Bbbk^{n\times n}$, show that

$$\|\exp(A) - \mathbf{1}\| \le e^{\|A\|} - 1.$$

$\diamond$ **Exercise 73.** Show that (for $\lambda, \mu \in \Bbbk$)

$$\exp((\lambda+\mu)A) = \exp(\lambda A)\exp(\mu A).$$

[Hint : These series are absolutely convergent. Think of the Cauchy product.]

It follows that

$$\mathbf{1} = \exp(0) = \exp((1+(-1))A) = \exp(A)\exp(-A)$$

and hence $\exp(A)$ is *invertible* with inverse $\exp(-A)$. So $\exp(A) \in \mathsf{GL}(n, \Bbbk)$.

Note : The "group property" $\exp((\lambda+\mu)A) = \exp(\lambda A)\exp(\mu A)$ may be rephrased by saying that, for fixed $A \in \Bbbk^{n\times n}$, the mapping $\lambda \mapsto \exp(\lambda A)$ is a (continuous) *homomorphism* from the additive group of scalars $\Bbbk$ into the general linear group $\mathsf{GL}(n, \Bbbk)$.

Definition 61. The mapping

$$\exp : \Bbbk^{n\times n} \to \mathsf{GL}(n, \Bbbk), \quad A \mapsto \exp(A)$$

is called the **matrix exponential map**.

Let $A \in \Bbbk^{n\times n}$ (with $\Bbbk$ either $\mathbb{R}$ or $\mathbb{C}$). Let $A^\dagger$ denote the transpose $A^\top$ when $\Bbbk = \mathbb{R}$, and the conjugate transpose A^* when $\Bbbk = \mathbb{C}$.

$\diamond$ **Exercise 74.** Show that

$$\exp(A)^\dagger = \exp(A^\dagger).$$

It is *not* true in general that $\exp(A+B) = \exp(A)\exp(B)$, although it is true if A and B commute. (This is a crucial point, with some significant consequences.)

Proposition 33. *If $A, B \in \Bbbk^{n\times n}$ commute, then*

$$\exp(A+B) = \exp(A)\exp(B).$$

Proof. We expand the series and perform a sequence of manipulations that are legitimate since these series are absolutely convergent :

$$\begin{aligned}
\exp(A)\exp(B) &= \left(\sum_{r=0}^{\infty}\frac{1}{r!}A^r\right)\left(\sum_{s=0}^{\infty}\frac{1}{s!}B^s\right)\\
&= \sum_{r,s=0}^{\infty}\frac{1}{r!s!}A^rB^s\\
&= \sum_{k=0}^{\infty}\left(\sum_{r=0}^{k}\frac{1}{r!(k-r)!}A^rB^{k-r}\right)\\
&= \sum_{k=0}^{\infty}\frac{1}{k!}\left(\sum_{r=0}^{k}\binom{k}{r}A^rB^{k-r}\right)\\
&= \sum_{k=0}^{\infty}\frac{1}{k!}(A+B)^k\\
&= \exp(A+B).
\end{aligned}$$

□

NOTE : We have made crucial use of the *commutativity* of A and B in the identity

$$\sum_{r=0}^{k}\binom{k}{r}A^rB^{k-r} = (A+B)^k.$$

In particular, for the (commuting) matrices λA and μA, we reobtain the property $\exp((\lambda+\mu)A) = \exp(\lambda A)\exp(\mu A)$. It is important to realize that, in fact, the following statements are equivalent (for $A, B \in \Bbbk^{n\times n}$):

(1) $AB = BA$.
(2) $\exp(\lambda A)\exp(\mu B) = \exp(\mu B)\exp(\lambda A)$ for all $\lambda, \mu \in \Bbbk$.
(3) $\exp(\lambda A + \mu B) = \exp(\lambda A)\exp(\mu B)$ for all $\lambda, \mu \in \Bbbk$.

◇ **Exercise 75.** Compute (for $a, b \in \mathbb{R}$)

$$\exp\left(\begin{bmatrix} a & 0\\ 0 & a\end{bmatrix}\right), \quad \exp\left(\begin{bmatrix} a & -b\\ b & a\end{bmatrix}\right), \quad \exp\left(\begin{bmatrix} a & b\\ b & a\end{bmatrix}\right), \quad \exp\left(\begin{bmatrix} a & b\\ 0 & a\end{bmatrix}\right).$$

NOTE : Every real 2×2 matrix is *conjugate* to exactly one of the following types (with $a, b \in \mathbb{R},\ b \neq 0$) :

- $a\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix}$ (scalar).
- $a\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix} + b\begin{bmatrix} 0 & -1\\ 1 & 0\end{bmatrix}$ (elliptic).
- $a\begin{bmatrix} 1 & 0\\ 0 & 1\end{bmatrix} + b\begin{bmatrix} 0 & 1\\ 1 & 0\end{bmatrix}$ (hyperbolic).

- $a \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + b \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ (parabolic).

$\diamond$ **Exercise 76.**

(a) Show that if $A \in \mathbb{R}^{n \times n}$ is *skew-symmetric*, then $\exp(A)$ is orthogonal.
(b) Show that if $A \in \mathbb{C}^{n \times n}$ is *skew-Hermitian*, then $\exp(A)$ is unitary.

$\diamond$ **Exercise 77.** Let $A \in \Bbbk^{n \times n}$ and $B \in \mathsf{GL}(n, \Bbbk)$. Show that

$$\exp\left(BAB^{-1}\right) = B \exp(A) B^{-1}.$$

Deduce that if $B^{-1}AB = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, then

$$\exp(A) = B \operatorname{diag}\left(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}\right) B^{-1}.$$

$\diamond$ **Exercise 78.** Show (for $\lambda \in \mathbb{R}$)

$$\exp\left(\begin{bmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & \lambda \end{bmatrix}\right) = \begin{bmatrix} e^\lambda & e^\lambda & \frac{1}{2!}e^\lambda & \dots & \frac{1}{(n-1)!}e^\lambda \\ 0 & e^\lambda & e^\lambda & \dots & \frac{1}{(n-2)!}e^\lambda \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & e^\lambda \end{bmatrix}.$$

NOTE : When the matrix $A \in \Bbbk^{n \times n}$ is *diagonalizable* over $\mathbb{C}$ (i.e., $A = C \operatorname{diag}(\lambda_1, \dots, \lambda_n) C^{-1}$ for some $C \in \mathsf{GL}(n, \mathbb{C})$), we have

$$\exp(A) = C \operatorname{diag}\left(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}\right) C^{-1}.$$

This means that the problem of calculating the exponential of a diagonalizable matrix is solved once an explicit diagonalization is found. Many important types of matrices are indeed diagonalizable (over $\mathbb{C}$), including skew-symmetric, skew-Hermitian, orthogonal, and unitary matrices. However, there are also many non-diagonalizable matrices. If $A^k = 0$ for some positive integer k, then $A^\ell = 0$ for all $\ell \geq k$. In this case the matrix series which defines $\exp(A)$ terminates after the first k terms, and so can be computed explicitly. A general matrix A may be neither nilpotent nor diagonalizable. This situation is best discussed in terms of the *Jordan canonical form.*

For $\lambda \in \mathbb{C}$ and $r \geq 1$, we have the *Jordan block matrix*

$$J(\lambda, r) := \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{bmatrix} \in \mathbb{C}^{r \times r}.$$

The characteristic polynomial of $J(\lambda, r)$ is

$$\operatorname{char}_{J(\lambda,r)}(s) := \det\left(sI_r - J(\lambda, r)\right) = (s - \lambda)^r$$

and by the CAYLEY-HAMILTON THEOREM, $(J(\lambda, r) - \lambda I_r)^r = 0$, which implies that $(J(\lambda, r) - \lambda I_r)^{r-1} \neq O$ (and hence $\mathrm{char}_{J(\lambda,r)}(s) = \min_{J(\lambda,r)}(s) \in \mathbb{C}[s]$). The main result on Jordan form is the following : *Given* $A \in \mathbb{C}^{n\times n}$*, there exists a matrix* $P \in \mathsf{GL}(n, \mathbb{C})$ *such that*

$$P^{-1}AP = \begin{bmatrix} J(\lambda_1, r_1) & 0 & \dots & 0 \\ 0 & J(\lambda_2, r_2) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & J(\lambda_m, r_m) \end{bmatrix} \in \mathbb{C}^{n\times n}.$$

This form is unique except for the order in which the Jordan blocks $J(\lambda_i, r_i) \in \mathbb{C}^{r_i\times r_i}$ *occur.* (The elements $\lambda_1, \lambda_2, \dots, \lambda_m$ are the eigenvalues of A and in fact $\mathrm{char}_A(s) = (s - \lambda_1)^{r_1}(s - \lambda_2)^{r_2}\cdots(s - \lambda_m)^{r_m}$.)

Using the Jordan canonical form we can see that every matrix $A \in \mathbb{C}^{n\times n}$ can be written as $A = S + N$, where S is diagonalizable (over $\mathbb{C}$), N is nilpotent, and $SN = NS$.

◇ **Exercise 79.** Compute

$$\exp\left(\begin{bmatrix} \lambda & a & b \\ 0 & \lambda & c \\ 0 & 0 & \lambda \end{bmatrix}\right).$$

The exponential mapping $\exp : \mathbb{k}^{n\times n} \to \mathsf{GL}(n, \mathbb{k})$ is *continuous* (in fact, infinitely differentiable). Indeed, since any power A^k is a continuous mapping of A, the sequence of partial sums $\left(\sum_{k=0}^{r} \frac{1}{k!}A^k\right)_{r\geq 0}$ consists of continuous mappings. But the matrix series defining the exponential matrix converges *uniformly* on each set of the form $\{A : \|A\| \leq \rho\}$, and so the sum (i.e., the limit of its sequence of partial sums) is again continuous. By continuity (of the exponential mapping at the origin 0), there is a number $\delta > 0$ such that

$$\mathcal{B}_{\mathbb{k}^{n\times n}}(0, \delta) \subseteq \exp^{-1}\left(\mathcal{B}_{\mathsf{GL}(n,\mathbb{k})}(\mathbf{1}, 1)\right).$$

In fact we can actually take $\delta = \ln 2$ since

$$\exp\left(\mathcal{B}_{\mathbb{k}^{n\times n}}(0, \delta)\right) \subseteq \mathcal{B}_{\mathbb{k}^{n\times n}}\left(\mathbf{1}, e^{\delta} - 1\right).$$

Hence we have the following result

PROPOSITION 34. *The exponential mapping* $\exp : \mathbb{k}^{n\times n} \to \mathsf{GL}(n, \mathbb{k})$ *is injective when restricted to the open subset* $\mathcal{B}_{\mathbb{k}^{n\times n}}(0, \ln 2)$. *(Hence it is locally a diffeomorphism at the origin* 0.*)*

Let $A \in \mathbb{k}^{n\times n}$. For every $t \in \mathbb{R}$, the matrix series $\displaystyle\sum_{k\geq 0} \frac{t^k}{k!}A^k$ is (absolutely) *convergent* and we have

$$\sum_{k=0}^{\infty} \frac{t^k}{k!}A^k = \sum_{k=0}^{\infty} \frac{1}{k!}(tA)^k = \exp(tA).$$

So the mapping

$$\alpha : \mathbb{R} \to \Bbbk^{n\times n}, \quad t \mapsto \exp(tA)$$

is defined and *differentiable* with

$$\dot\alpha(t) = \sum_{k=1}^{\infty} \frac{t^{k-1}}{(k-1)!} A^k = \exp(tA)A = A\exp(tA).$$

NOTE : This mapping can be viewed as a *curve* in $\Bbbk^{n\times n}$. The curve is in fact *smooth* (i.e., infinitely differentiable) and satisfies the differential equation (in matrices) $\dot\alpha(t) = \alpha(t)A$ with initial condition $\alpha(0) = \mathbf{1}$. Also (for $t, s \in \mathbb{R}$),

$$\alpha(t+s) = \alpha(t)\alpha(s).$$

In particular, this shows that $\alpha(t)$ is always invertible with $\alpha(t)^{-1} = \alpha(-t)$.

$\diamond$ **Exercise 80.** Let $A, C \in \Bbbk^{n\times n}$. Show that the differential equation (in matrices) $\dot\alpha = \alpha\, A$ has a *unique* differentiable solution $\alpha : \mathbb{R} \to \Bbbk^{n\times n}$ for which $\alpha(0) = C$. (This solution is $\alpha(t) = C\exp(tA)$.) Furthermore, if C is invertible, then so is $\alpha(t)$ for $t \in \mathbb{R}$, hence α is in fact a curve in $\mathsf{GL}(n, \Bbbk)$.

5.2. **Some useful formulas.**

5.2.1. *First formula.* The following formula can be considered as another definition of the matrix exponential.

PROPOSITION 35. *Let $A \in \Bbbk^{n\times n}$. Then*

$$\exp(A) = \lim_{r\to\infty} \left(\mathbf{1} + \frac{1}{r}A\right)^r.$$

Proof. Consider the difference

$$\exp(A) - \left(\mathbf{1} + \frac{1}{r}A\right)^r = \sum_{k=0}^{\infty} \left(\frac{1}{k!} - \frac{1}{r^k}\binom{r}{k}\right) A^k.$$

This matrix series converges since the series for the matrix exponential $\exp(A)$ converges and $\left(\mathbf{1} + \frac{1}{r}A\right)^r$ is a polynomial. The coefficients in the rhs are nonnegative since

$$\frac{1}{k!} \geq \frac{r(r-1)\cdots(r-k+1)}{r\cdot r\cdots r}\,\frac{1}{k!}.$$

Therefore, setting $\|A\| = a$, we get

$$\left\|\exp(A) - \left(\mathbf{1} + \frac{1}{r}A^r\right)^r\right\| \leq \sum_{k=0}^{\infty} \left(\frac{1}{k!} - \frac{1}{r^k}\binom{r}{k}\right) a^k = e^a - \left(1 + \frac{a}{r}\right)^r$$

where the expression on the right approaches zero (as $r \to \infty$). The result now follows. $\square$

5.2.2. *Second formula.*

PROPOSITION 36. *Let* $A \in \Bbbk^{n\times n}$ *and* $\epsilon \in \mathbb{R}$. *Then*

$$\det(\mathbf{1} + \epsilon A) = 1 + \epsilon \operatorname{tr} A + O(\epsilon^2) \quad (\textit{as } \epsilon \to 0).$$

Proof. The determinant of $\mathbf{1} + \epsilon A$ equals the product of the eigenvalues of the matrix. But the eigenvalues of $\mathbf{1} + \epsilon A$ (with due regard for multiplicity) equal $1 + \epsilon \lambda_i$, where the λ_i are the eigenvalues of A. It follows that

$$\begin{aligned} \det(\mathbf{1} + \epsilon A) &= (1 + \epsilon\lambda_1)(1 + \epsilon\lambda_2)\cdots(1 + \epsilon\lambda_n) \\ &= 1 + \epsilon(\lambda_1 + \lambda_2 + \cdots + \lambda_n) + O(\epsilon^2) \\ &= 1 + \epsilon \operatorname{tr} A + O(\epsilon^2). \end{aligned}$$

□

NOTE : Whenever we have a mapping Z from some (open) interval (a, b), $a < 0 < b$ into a finite-dimensional normed vector space (e.g. $\Bbbk^{n\times n}$), then Z will often be denoted by $O(t^k)$ if $t \mapsto \frac{1}{t^k} Z(t)$ is bounded in an (open) neighborhood of the origin 0 (i.e. there are constants C_1 and C_2 such that

$$\|Z(t)\| \leq C_1 |t^k| \quad \text{for } |t| < C_2.)$$

Thus $O(t^k)$ may denote different mappings at different times. The big-O notation was first introduced in 1892 by PAUL G.H. BACHMANN (1837-1920) in a book on number theory, and is currently used in several areas of mathematics and computer science (including mathematical analysis and the theory of algorithms).

5.2.3. *Third formula.*

PROPOSITION 37. *Let* $\alpha : (a, b) \to \Bbbk^{n\times n}$ *be a curve. Then*

$$\left.\frac{d}{dt} \det \alpha(t)\right|_{t=0} = \operatorname{tr} \dot{\alpha}(0).$$

Proof. The operation $\partial := \left.\frac{d}{dt}\right|_{t=0}$ has the *derivation property*

$$\partial(\gamma_1\gamma_2) = (\partial\gamma_1)\gamma_2(0) + \gamma_1(0)\partial\gamma_2.$$

Put $\alpha(t) = [a_{ij}(t)]$ and notice that (when $t = 0$) $a_{ij} = \delta_{ij}$. Write C_{ij} for the *cofactor matrix* obtained from $\alpha(t)$ by deleting the i^{th} row and the j^{th} column. By expanding along the n^{th} row we get

$$\det \alpha(t) = \sum_{j=1}^{n} (-1)^{n+j} a_{nj} \det C_{nj}.$$

For $t = 0$ (since $\alpha(0) = \mathbf{1}$) we have

$$\det C_{nj} = \delta_{nj}.$$

Then

$$\begin{aligned}\partial \det \alpha(t) &= \sum_{j=1}^{n}(-1)^{n+j}((\partial a_{nj}) \det C_{nj} + a_{nj}(\partial \det C_{nj}))\\ &= \sum_{j=1}^{n}(-1)^{n+j}((\partial a_{nj}) \det C_{nj}) + (\partial \det C_{nn})\\ &= \partial a_{nn} + \partial \det C_{nn}.\end{aligned}$$

We can repeat this calculation with the $(n-1) \times (n-1)$ matrix C_{nn} and so on. This gives

$$\begin{aligned}\partial \det \alpha(t) &= \partial a_{nn} + \partial a_{n-1,n-1} + \partial \det C_{n-1,n-1}\\ &\vdots\\ &= \partial a_{nn} + \partial a_{n-1,n-1} + \cdots + \partial a_{11}\\ &= \operatorname{tr} \dot{\alpha}(0).\end{aligned}$$

□

5.2.4. *Liouville's formula.* We can now prove a remarkable (and very useful) result, known as Liouville's Formula. Three different proofs will be given.

Theorem 38 (Liouville's Formula). *For* $A \in \mathbb{k}^{n\times n}$ *we have*

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

First Proof (using the second definition of the exponential) : We have

$$\det \exp(A) = \det \lim_{r\to\infty}\left(\mathbf{1} + \frac{1}{r}A\right)^r = \lim_{r\to\infty} \det \left(\mathbf{1} + \frac{1}{r}A\right)^r$$

since the determinant function $\det : \mathbb{k}^{n\times n} \to \mathbb{k}$ is *continuous*. Moreover, by Proposition 36,

$$\det \left(\mathbf{1} + \frac{1}{r}A\right)^r = \left[\det \left(\mathbf{1} + \frac{1}{r}A\right)\right]^r = \left[1 + \frac{1}{r}\operatorname{tr} A + O\left(\frac{1}{r^2}\right)\right]^r \quad (\text{as } r \to \infty).$$

It only remains to note that (for any $a \in \mathbb{k}$)

$$\lim_{r\to\infty}\left[1 + \frac{a}{r} + O\left(\frac{1}{r^2}\right)\right]^r = e^a.$$

In particular, for $a = \operatorname{tr} A$, we get the desired result. □

Second Proof (using differential equations) : Consider the curve

$$\gamma : \mathbb{R} \to \mathsf{GL}(1, \mathbb{k}) = \mathbb{k}^\times, \quad t \mapsto \det \exp(tA).$$

Then (by PROPOSITION 37 applied to the curve γ)

$$\begin{aligned}\dot{\gamma}(t) &= \lim_{h\to 0}\frac{1}{h}\left[\det\,\exp\left((t+h)A\right) - \det\,\exp\left(tA\right)\right]\\ &= \det\,\exp\left(tA\right)\lim_{h\to 0}\frac{1}{h}\left[\det\,\exp\left(hA\right) - 1\right]\\ &= \det\,\exp\left(tA\right)\operatorname{tr} A\\ &= \gamma(t)\operatorname{tr} A.\end{aligned}$$

So γ satisfies the same differential equation and initial condition as the curve $t \mapsto e^{t\operatorname{tr} A}$. By the uniqueness of the solution (see **Exercise 80**), it follows that

$$\gamma(t) = \det\,\exp\left(tA\right) = e^{t\operatorname{tr} A}.$$

In particular, for $t = 1$, we get the desired result. □

THIRD PROOF (using Jordan canonical form) : If $B \in \mathsf{GL}\left(n, \Bbbk\right)$, then (see **Exercise 77**)

$$\begin{aligned}\det\,\exp\left(BAB^{-1}\right) &= \det\left(B\exp\left(A\right)B^{-1}\right)\\ &= \det\,B\cdot\det\,\exp(A)\cdot\det\,B^{-1}\\ &= \det\,\exp\left(A\right)\end{aligned}$$

and

$$e^{\operatorname{tr}\left(BAB^{-1}\right)} = e^{\operatorname{tr} A}.$$

So it suffices to prove the identity for BAB^{-1} for a *suitably chosen* invertible matrix B. Using for example the theory of *Jordan canonical forms*, there is a suitable choice of such a B for which

$$BAB^{-1} = D + N$$

with D diagonal and N *strictly upper triangular* (i.e., $N_{ij} = 0$ for $i \geq j$). Then N is *nilpotent* (i.e., $N^k = O$ for some $k \geq 1$). We have

$$\begin{aligned}\exp\left(BAB^{-1}\right) &= \sum_{k=0}^{\infty}\frac{1}{k!}(D+N)^k\\ &= \sum_{k=0}^{\infty}\frac{1}{k!}D^k + \sum_{k=0}^{\infty}\frac{1}{(k+1)!}\left((D+N)^{k+1} - D^{k+1}\right)\\ &= \exp\left(D\right) + \sum_{k=0}^{\infty}\frac{1}{(k+1)!}N(D^k + D^{k-1}N + \cdots + N^k).\end{aligned}$$

The matrix

$$N(D^k + D^{k-1}N + \cdots + N^k)$$

is strictly upper triangular, and so

$$\exp\left(BAB^{-1}\right) = \exp\left(D\right) + N'$$

where N' is strictly upper triangular. Now, if $D = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$, we have

$$\begin{aligned}
\det \exp(A) &= \det \exp(BAB^{-1}) \\
&= \det \exp(D) \\
&= \det \operatorname{diag}(e^{\lambda_1}, e^{\lambda_2}, \dots, e^{\lambda_n}) \\
&= e^{\lambda_1} e^{\lambda_2} \cdots e^{\lambda_n} \\
&= e^{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \\
&= e^{\operatorname{tr} D} \\
&= e^{\operatorname{tr}(BAB^{-1})} \\
&= e^{\operatorname{tr} A}.
\end{aligned}$$

□

The exponential mapping

$$\exp : \Bbbk^{n\times n} \to \mathsf{GL}(n, \Bbbk)$$

is a basic link between the linear structure on $\Bbbk^{n\times n}$ and the multiplicative structure on $\mathsf{GL}(n, \Bbbk)$. Let G be a linear Lie subgroup of $\mathsf{GL}(n, \Bbbk)$. Applying Proposition 34, we may choose $\rho \in \mathbb{R}$ so that $0 < \rho \leq \frac{1}{2}$ and if $A, B \in \mathcal{B}_{\Bbbk^{n\times n}}(O, \rho)$, then $\exp(A)\exp(B) \in \exp\left(\mathcal{B}_{\Bbbk^{n\times n}}(O, \frac{1}{2})\right)$. Since $\exp$ is one-to-one on $\mathcal{B}_{\Bbbk^{n\times n}}(O, \rho)$, *there is a unique matrix* $C \in \Bbbk^{n\times n}$ *for which*

$$\exp(A)\exp(B) = \exp(C).$$

Note : There is a beautiful formula, the Baker-Campbell-Hausdorff formula which expresses C as a power series in A and B. To develop this completely would take too long. Specifically, (one form of) the B-C-H formula says that *if X and Y are sufficiently small, then*

$$\exp(X)\exp(Y) = \exp(Z) \quad \text{with}$$

$$Z = X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] - \frac{1}{12}[Y,[X,Y]] + \cdots$$

It is not supposed to be evident at the moment what "…" refers to. The only important point is that all the terms (in the expansion of Z) are given in terms of X and Y, Lie brackets of X and Y, Lie brackets of Lie brackets involving X and Y, etc. Then it follows that the mapping $\phi : \mathsf{G} \to \mathsf{GL}(n, \mathbb{R})$ "defined" by the relation

$$\phi(\exp(X)) = \exp(\phi(X))$$

is such that on elements of the form $\exp(X)$, with X sufficiently small, is a *group homomorphism.* Hence the B-C-H formula shows that *all the information about the group product, a least near the identity, is "encoded" in the Lie algebra.*

An interesting special case is the following : If $X, Y \in \mathbb{C}^{n\times n}$ and X, Y commute with their commutator (i.e., $[X,[X,Y]] = [Y,[X,Y]]$), then

$$\exp(X)\exp(Y) = \exp\left(X + Y + \frac{1}{2}[X,Y]\right).$$

◇ **Exercise 81.** Show by direct computation that for

$$X, Y \in \mathfrak{h}_3 = \left\{ \begin{bmatrix} 0 & a & b \\ 0 & 0 & c \\ 0 & 0 & 0 \end{bmatrix} : a, b, c \in \mathbb{R} \right\}$$

(the Lie algebra of the Heisenberg group H_3)

$$\exp(X)\exp(Y) = \exp\left(X + Y + \frac{1}{2}[X,Y]\right).$$

5.3. **The product and commutator formulas (optional).** We set

$$R = C - A - B \in \Bbbk^{n\times n}.$$

For $X \in \Bbbk^{n\times n}$, we have

$$\exp(X) = \mathbf{1} + X + R_1(X),$$

where the remainder term $R_1(X)$ is given by

$$R_1(X) = \sum_{k=2}^{\infty} \frac{1}{k!} X^k.$$

Hence

$$\|R_1(X)\| \leq \|X\|^2 \sum_{k=2}^{\infty} \frac{1}{k!} \|X\|^{k-2}$$

and therefore if $\|X\| < 1$, then

$$\|R_1(X)\| \leq \|X\|^2 \sum_{k=2}^{\infty} \frac{1}{k!} = \|X\|^2 (e-2) < \|X\|^2.$$

Now for $X = C \in \mathcal{B}_{\Bbbk^{n\times n}}(O, \frac{1}{2})$, we have

$$\exp(C) = \mathbf{1} + C + R_1(C)$$

with

$$\|R_1(C)\| < \|C\|^2.$$

Similar considerations lead to

$$\exp(C) = \exp(A)\exp(B) = \mathbf{1} + A + B + R_1(A,B),$$

where

$$R_1(A,B) = \sum_{k=2}^{\infty} \frac{1}{k!} \left(\sum_{r=0}^{k} \binom{k}{r} A^r B^{k-r} \right).$$

This gives

$$\begin{aligned}\|R_1(A,B)\| &\leq \sum_{k=2}^{\infty} \frac{1}{k!}\left(\sum_{r=0}^{k}\binom{k}{r}\|A\|^r\|B\|^{k-r}\right)\\ &= \sum_{k=0}^{\infty}\frac{1}{k!}\left(\|A\|+\|B\|\right)^k\\ &= \left(\|A\|+\|B\|\right)^2\sum_{k=2}^{\infty}\frac{1}{k!}\left(\|A\|+\|B\|\right)^{k-2}\\ &\leq \left(\|A\|+\|B\|\right)^2\end{aligned}$$

since $\|A\|+\|B\|<1$.

Combining the two ways of writing $\exp(C)$ from above, we have

$$C = A + B + R_1(C) - R_1(A,B)$$

and so

$$\begin{aligned}\|C\| &\leq \|A\|+\|B\|+\|R_1(A,B)\|+\|R_1(C)\|\\ &< \|A\|+\|B\|+\left(\|A\|+\|B\|\right)^2+\|C\|^2\\ &\leq 2\left(\|A\|+\|B\|\right)+\frac{1}{2}\|C\|\end{aligned}$$

since $\|A\|, \|B\|, \|C\| \leq \frac{1}{2}$. Finally this gives

$$\|C\| \leq 4\left(\|A\|+\|B\|\right).$$

We also have

$$\begin{aligned}\|R\| = \|C-A-B\| &\leq \|R_1(A,B)\|+\|R_1(C)\|\\ &\leq \left(\|A\|+\|B\|\right)^2+\left(4(\|A\|+\|B\|)\right)^2\\ &= 17\left(\|A\|+\|B\|\right)^2.\end{aligned}$$

We have proved the following result.

PROPOSITION 39. *Let* $A, B, C \in \mathcal{B}_{\Bbbk^{n\times n}}(O, \frac{1}{2})$ *such that* $\exp(A)\exp(B) = \exp(C)$. *Then* $C = A + B + R$, *where the remainder term* R *satisfies*

$$\|R\| \leq 17\left(\|A\|+\|B\|\right)^2.$$

We can refine this estimate (to second order). We only point out the essential steps (details will be omitted). Set

$$S = C - A - B - \frac{1}{2}[A,B] \in \Bbbk^{n\times n}$$

and write

$$\exp(C) = \mathbf{1} + C + \frac{1}{2}C^2 + R_2(C)$$

with

$$\|R_2(C)\| \leq \frac{1}{3}\|C\|^3.$$

Then

$$\begin{aligned}\exp(C) &= \mathbf{1} + A + B + \frac{1}{2}[A,B] + S + \frac{1}{2}C^2 + R_2(C)\\ &= \mathbf{1} + A + B + \frac{1}{2}(A^2 + 2AB + B^2) + T,\end{aligned}$$

where

$$T = S + \frac{1}{2}(C^2 - (A+B)^2) + R_2(C).$$

Also

$$\exp(A)\exp(B) = \mathbf{1} + A + B + \frac{1}{2}(A^2 + 2AB + B^2) + R_2(A,B)$$

with

$$\|R_2(A,B)\| \leq \frac{1}{3}\left(\|A\| + \|B\|\right)^3.$$

We see that

$$S = R_2(A,B) + \frac{1}{2}((A+B)^2 - C^2) - R_2(C)$$

and by taking norms we get

$$\begin{aligned}\|S\| &\leq \|R_2(A,B)\| + \frac{1}{2}\|(A+B)(A+B-C) + (A+B-C)C\| + \|R_2(C)\|\\ &\leq \frac{1}{3}\left(\|A\| + \|B\|\right)^3 + \frac{1}{2}(\|A\| + \|B\| + \|C\|)\|A+B-C\| + \frac{1}{3}\|C\|^3\\ &\leq 65\left(\|A\| + \|B\|\right)^3.\end{aligned}$$

The following estimation holds.

Proposition 40. *Let* $A, B, C \in \mathcal{B}_{\Bbbk^{n\times n}}(O, \frac{1}{2})$ *such that* $\exp(A)\exp(B) = \exp(C)$. *Then* $C = A + B + \frac{1}{2}[A,B] + S$, *where the remainder term* S *satisfies*

$$\|S\| \leq 65\left(\|A\| + \|B\|\right)^3.$$

We will derive two main consequences of Proposition 39 and Proposition 40. (These relate group operations in $\mathsf{GL}(n,\Bbbk)$ to the linear operations in $\Bbbk^{n\times n}$ and are crucial ingredients in the proof that *every linear Lie group is a Lie group.*)

Theorem 41 (Lie-Trotter Product Formula). *For* $U, V \in \Bbbk^{n\times n}$ *we have*

$$\exp(U+V) = \lim_{r\to\infty}\left(\exp\left(\frac{1}{r}U\right)\exp\left(\frac{1}{r}V\right)\right)^r.$$

(This formula relates addition in $\Bbbk^{n\times n}$ to multiplication in $\mathsf{GL}(n,\Bbbk)$.)

Proof. For *large* r we may take $A = \frac{1}{r}U$ and $B = \frac{1}{r}V$ and apply PROPOSITION 39 to give

$$\exp\left(\frac{1}{r}U\right)\exp\left(\frac{1}{r}V\right) = \exp\left(C_r\right)$$

with

$$\left\| C_r - \frac{1}{r}(U+V) \right\| \leq \frac{17\left(\|U\| + \|V\|\right)^2}{r^2}.$$

As $r \to \infty$,

$$\|rC_r - (U+V)\| \leq \frac{17\left(\|U\| + \|V\|\right)^2}{r} \to 0$$

and hence

$$rC_r \to U + V.$$

Since $\exp\left(rC_r\right) = \exp\left(C_r\right)^r$, the LIE-TROTTER PRODUCT FORMULA follows by continuity of the exponential mapping. □

THEOREM 42 (COMMUTATOR FORMULA). *For* $U, V \in \Bbbk^{n\times n}$ *we have*

$$\exp([U,V]) = \lim_{r\to\infty}\left(\exp\left(\frac{1}{r}U\right)\exp\left(\frac{1}{r}V\right)\exp\left(-\frac{1}{r}U\right)\exp\left(-\frac{1}{r}V\right)\right)^{r^2}.$$

(This formula relates the Lie bracket - or commutator - in $\Bbbk^{n\times n}$ to the *group commutator* in $\mathsf{GL}\,(n, \Bbbk)$.)

Proof. For *large* r (as in the proof of THEOREM 41) we have

$$\exp\left(\frac{1}{r}U\right)\exp\left(\frac{1}{r}V\right) = \exp(C_r)$$

with (as $r \to \infty$)

$$rC_r \to U + V.$$

We also have

$$C_r = \frac{1}{r}(U+V) + \frac{1}{2r^2}[U,V] + S_r,$$

where

$$\|S_r\| \leq 65\frac{(\|U\| + \|V\|)^3}{r^3}.$$

Similarly (replacing U, V with $-U, -V$) we obtain :

$$\exp\left(-\frac{1}{r}U\right)\exp\left(-\frac{1}{r}V\right) = \exp(C_r'),$$

where

$$C_r' = -\frac{1}{r}(U+V) + \frac{1}{2r^2}[U,V] + S_r'$$

and

$$\|S_r'\| \leq 65\frac{(\|U\| + \|V\|)^3}{r^3}.$$

Combining these we get

$$\begin{aligned}\exp\left(\frac{1}{r}U\right)\exp\left(\frac{1}{r}V\right)\exp\left(-\frac{1}{r}U\right)\exp\left(-\frac{1}{r}V\right) &= \exp(C_r)\exp(C'_r)\\ &= \exp(E_r),\end{aligned}$$

where

$$\begin{aligned}E_r &= C_r + C'_r + \frac{1}{2}[C_r, C'_r] + T_r\\ &= \frac{1}{r^2}[U,V] + \frac{1}{2}[C_r, C'_r] + S_r + S'_r + T_r.\end{aligned}$$

One can verify that

$$\begin{aligned}[C_r, C'_r] &= \frac{1}{r^3}\,[U+V,[U,V]] + \frac{1}{r}\,[U+V, S_r + S'_r]\\ &\quad + \frac{1}{2r^2}\,[[U,V], S'_r - S_r] + [S_r, S'_r].\end{aligned}$$

All four of these terms have norm bounded by an expression of the form $\frac{\text{constant}}{r^3}$ so the same is true of $[C_r, C'_r]$. Also S_r, S'_r, T_r have similarly bounded norms. Setting

$$Q_r := r^2 E_r - [U,V]$$

we obtain (as $r \to \infty$)

$$\|Q_r\| = r^2\|E_r - \frac{1}{r^2}[U,V]\| \leq \frac{\text{constant}}{r} \to 0$$

and hence

$$\exp(E_r)^{r^2} = \exp\left([U,V] + Q_r\right) \to \exp([U,V]).$$

The Commutator Formula now follows using continuity of the exponential mapping. □

Note : If g, h are elements of a group, then the expression $ghg^{-1}h^{-1}$ is called the *group commutator* of g and h.

5.4. **The adjoint action.** There is one further concept involving the exponential mapping that is basic in Lie theory. It involves *conjugation*, which is generally referred to as the **adjoint action**. For $g \in \mathsf{GL}\,(n, \Bbbk)$ and $A \in \Bbbk^{n\times n}$, we can form the conjugate

$$\mathrm{Ad}_g(A) := g\,A\,g^{-1}.$$

◇ **Exercise 82.** Let $A, B \in \Bbbk^{n\times n}$ and $g, h \in \mathsf{GL}\,(n, \Bbbk)$. Show that (for $\lambda, \mu \in \Bbbk$)

(a) $\mathrm{Ad}_g(\lambda A + \mu B) = \lambda \mathrm{Ad}_g(A) + \mu \mathrm{Ad}_g(B)$.
(b) $\mathrm{Ad}_g([A,B]) = [\mathrm{Ad}_g(A), \mathrm{Ad}_g(B)]$.
(c) $\mathrm{Ad}_{gh}(A) = \mathrm{Ad}_g(\mathrm{Ad}_h(A))$.

In particular, $\mathrm{Ad}_g^{-1} = \mathrm{Ad}_{g^{-1}}$.

Formulas (a) an (b) say that Ad_g is an *automorphism* of the Lie algebra $\Bbbk^{n\times n}$, and formula (c) says the mapping

$$\mathrm{Ad} : \mathsf{GL}\,(n, \Bbbk) \to \mathsf{Aut}\,(\Bbbk^{n\times n}), \quad g \mapsto \mathrm{Ad}_g$$

is a *group homomorphism*. The mapping Ad is called the **adjoint representation** of $\mathsf{GL}\,(n, \Bbbk)$.

Formula (c) implies in particular that if $t \mapsto \exp{(tA)}$ is a one-parameter subgroup of $\mathsf{GL}\,(n, \Bbbk)$, then $\mathrm{Ad}_{\exp{(tA)}}$ is a one-parameter group (of linear transformations) in $\Bbbk^{n\times n}$. Observe that we can identify $\mathsf{Aut}\,(\Bbbk^{n\times n})$ with $\mathsf{GL}\,(n^2, \Bbbk)$ (and thus view $\mathsf{Aut}\,(\Bbbk^{n\times n})$ as a linear Lie group). Then (see Theorem 44)

$$\mathrm{Ad}_{\exp{(tA)}} = \exp{(t\mathcal{A})}$$

for some $\mathcal{A} \in \Bbbk^{n^2\times n^2} = \mathsf{End}\,(\Bbbk^{n\times n})$. Since

$$\begin{aligned} \mathcal{A}(B) &= \frac{d}{dt}\mathrm{Ad}_{\exp(tA)}(B)\Big|_{t=0} \\ &= \frac{d}{dt}\exp{(tA)}B\exp{(-tA)}\Big|_{t=0} \\ &= [A, B] \end{aligned}$$

by setting (for $A, B \in \Bbbk^{n\times n}$)

$$\mathrm{ad}\,A(B) := [A, B]$$

we have the following formula

$$\mathrm{Ad}_{\exp{(tA)}} = \exp{(t\,\mathrm{ad}\,A)}.$$

Explicitly, the formula says that

$$\exp{(tA)}B\exp{(-tA)} = \sum_{k=0}^{\infty}\frac{t^k}{k!}\,(\mathrm{ad}\,A)^k\,B.$$

(Here $(\mathrm{ad}\,A)^0 = A$ and $(\mathrm{ad}\,A)^k = \mathrm{ad}(\mathrm{ad}\,A)^{k-1}$ for $k \geq 1$.)

Note : The mapping

$$\mathrm{ad} : \Bbbk^{n\times n} \to \mathsf{End}\,(\Bbbk^{n\times n}), \quad X \mapsto \mathrm{ad}\,X$$

is called the **adjoint representation** of (the Lie algebra) $\Bbbk^{n\times n}$. From the Jacobi identity for Lie algebras, we have

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

That is, $\mathrm{ad}\,X$ is a *derivation* of the Lie algebra $\Bbbk^{n\times n}$. The formula above gives the relation between the automorphism $\mathrm{Ad}_{\exp{(tX)}}$ of the Lie algebra $\Bbbk^{n\times n}$ and the derivation $\mathrm{ad}\,X$ of $\Bbbk^{n\times n}$. One also has

$$\exp{(t\mathrm{Ad}_g(X))} = g\exp{(tX)}g^{-1}.$$

Using this formula, we can see that $[X, Y] = 0$ <u>if and only if</u> $\exp{(tX)}$ and $\exp{(sY)}$ commute for arbitrary $s, t \in \mathbb{R}$.

Problems (26–32)

(26) A matrix $A \in \Bbbk^{n\times n}$ is **nilpotent** if $A^k = 0$ for some $k \geq 1$.
  (a) Prove that a nilpotent matrix is singular.
  (b) Prove that a *strictly upper triangular* matrix $A = [a_{ij}]$ (i.e. with $a_{ij} = 0$ whenever $i \geq j$) is nilpotent.
  (c) Find two nilpotent matrices whose product is *not* nilpotent.

(27) Suppose that $A \in \Bbbk^{n\times n}$ and $\|A\| < 1$.
  (a) Show that the matrix series
$$\sum_{k\geq 0} A^k = \mathbf{1} + A + A^2 + A^3 + \cdots$$
  converges (in $\Bbbk^{n\times n}$).
  (b) Show that the matrix $\mathbf{1} - A$ is *invertible* and find a formula for $(\mathbf{1} - A)^{-1}$.
  (c) If A is *nilpotent*, determine $(\mathbf{1} - A)^{-1}$ and $\exp(A)$.

(28) Let $A \in \Bbbk^{n\times n}$.
  (a) Prove that A is *nilpotent* if and only if all its eigenvalues are equal to zero.
  (b) The matrix A is called **unipotent** if $\mathbf{1} - A$ is nilpotent (i.e., $(\mathbf{1} - A)^k = 0$ for some $k \geq 1$). Prove that A is *unipotent* if and only if all its eigenvalues are equal to 1.
  (c) If A is a strictly upper triangular matrix, show that $\exp(A)$ is unipotent.

(29) Let $A \in \Bbbk^{n\times n}$. Show that the functional equation (in matrices) $\alpha(t+s) = \alpha(t)\alpha(s)$ has a *unique* differentiable solution $\alpha : \mathbb{R} \to \Bbbk^{n\times n}$ for which $\alpha(0) = \mathbf{1}$ and $\dot\alpha(0) = A$. (This solution is $\alpha(t) = \exp(tA)$.)

(30) If $A, B \in \Bbbk^{n\times n}$ commute, show that
$$\frac{d}{dt}\exp(A + tB)\bigg|_{t=0} = \exp(A)B = B\exp(A).$$
(This is a formula for the *derivative* of the exponential mapping exp at an arbitrary A, evaluated only at those B such that $AB = BA$. The general situation is more complicated.)

(31) Let $A, B \in \Bbbk^{n\times n}$.
  (a) Verify that
$$\operatorname{ad}[A, B] = \operatorname{ad} A \operatorname{ad} B - \operatorname{ad} B \operatorname{ad} A = [\operatorname{ad} A,\, \operatorname{ad} B]\,.$$
  (This means that $\operatorname{ad} : \Bbbk^{n\times n} \to \mathsf{End}(\Bbbk^{n\times n})$ is a *Lie algebra homomorphism.*)
  (b) Show by induction that
$$(\operatorname{ad} A)^n(B) = \sum_{k=0}^{n} \binom{n}{k} A^k B(-A)^{n-k}.$$

(c) Show by direct computation that

$$\exp(\operatorname{ad} A)(B) = \operatorname{Ad}_{\exp(A)}(B) = \exp(A)B\exp(-A).$$

(32) Let $\alpha : \mathbb{R} \to \Bbbk^{n\times n}$ be a differentiable curve in $\Bbbk^{n\times n}$. Prove the formula

$$\frac{d}{dt}\exp(\alpha(t)) = \exp(\alpha(t))\frac{1-\exp(-\operatorname{ad}\alpha(t))}{\operatorname{ad}\alpha(t)}\frac{d\alpha}{dt}.$$

(The fraction of linear transformations of $\Bbbk^{n\times n}$ is defined by its – everywhere convergent – power series

$$\frac{1-\exp(-\operatorname{ad} X)}{\operatorname{ad} X} := \sum_{k=0}^{\infty}\frac{(-1)^k}{(k+1)!}(\operatorname{ad} X)^k.)$$

This exercise (statement) may also be read as saying that *the differential of the matrix exponential map* $\exp : \Bbbk^{n\times n} \to \Bbbk^{n\times n}$ *at any* $X \in \Bbbk^{n\times n}$ *is the linear transformation* $d\exp_X = D\exp(X) : \Bbbk^{n\times n} \to \Bbbk^{n\times n}$ *given by*

$$d\exp_X Y = \exp(X)\frac{1-\exp(-\operatorname{ad} X)}{\operatorname{ad} X}Y.$$

(The statement, together with the Inverse Function Theorem, gives information on the local behaviour of the matrix exponential map: the Inverse Function Theorem says that $\exp$ has a local inverse around a point $X \in \Bbbk^{n\times n}$ at which its differential $d\exp_X$ is invertible, and the statement says that this is the case precisely when $(1-\exp(-\operatorname{ad} X))/\operatorname{ad} X$ is invertible, i.e., when zero is not an eigenvalue of this linear transformation of $\Bbbk^{n\times n}$.)