II. The exponential function

In this chapter we study one of the central tools in Lie theory: the matrix exponential function. This function has various applications in the structure theory of subgroups of matrix groups. First of all it is naturally linked to the one-parameter subgroups and it turns out that the local structure of a closed subgroup $G \subseteq \mathrm{GL}_n(\mathbb{K})$ is determined by its one-parameter subgroups. Moreover, it helps us to understand the global topology of various groups of matrices by refining the polar decomposition.

II.1. Smooth functions defined by power series

First we put the structure that we have on the space $M_n(\mathbb{K})$ of $(n \times n)$-matrices into a slightly more general context.

Definition II.1.1. (a) A vector space A together with a bilinear map $A \times A \to A, (x,y) \mapsto x \cdot y$ (called multiplication) is called an (*associative*) *algebra* if the multiplication is associative in the sense that

$$(x \cdot y) \cdot z = x \cdot (y \cdot z) \quad \text{for} \quad x, y, z \in A.$$

We write $xy := x \cdot y$ for the product of x and y in A.

The algebra A is called *unital* if it contains an element $\mathbf{1}$ satisfying $\mathbf{1}a = a\mathbf{1} = a$ for each $a \in A$.

(b) A norm $\|\cdot\|$ on an algebra A is called *submultiplicative* if

$$\|ab\| \leq \|a\| \cdot \|b\| \quad \text{for all} \quad a, b \in A.$$

Then the pair $(A, \|\cdot\|)$ is called a *normed algebra*. If, in addition, A is a complete normed space, then it is said to be a *Banach algebra*. ■

Remark II.1.2. Any finite-dimensional normed space is complete, so that each finite-dimensional normed algebra is a Banach algebra. ■

Example II.1.3. Endowing $M_n(\mathbb{K})$ with the operator norm with respect to the euclidean norm on $\mathbb{K}^n$ defines on $M_n(\mathbb{K})$ the structure of a unital Banach algebra. ■

Lemma II.1.4. *If A is a unital Banach algebra, then we endow the vector space $TA := A \oplus A$ with the norm*

$$\|(a,b)\| := \|a\| + \|b\|$$

and the multiplication

$$(a,b)(a',b') := (aa', ab' + a'b).$$

Then TA is a unital Banach algebra.

Writing $\varepsilon := (0,1)$, then each element of TA can be written in a unique fashion as $(a,b) = a + b\varepsilon$ and the multiplication satisfies

$$(a + b\varepsilon)(a' + b\varepsilon') = aa' + (ab' + a'b)\varepsilon.$$

In particular, $\varepsilon^2 = 0$.

Proof. That TA is a unital algebra is a trivial verification. That the norm is submultiplicative follows from

$$\begin{aligned} \|(a,b)(a',b')\| &= \|aa'\| + \|ab' + a'b\| \leq \|a\| \cdot \|a'\| + \|a\| \cdot \|b'\| + \|a'\| \cdot \|b\| \\ &\leq (\|a\| + \|b\|)(\|a'\| + \|b'\|) = \|(a,b)\| \cdot \|(a',b')\|. \end{aligned}$$

This proves that $(TA, \|\cdot\|)$ is a unital Banach algebra, the unit being $\mathbf{1} = (\mathbf{1}, 0)$. The completeness of TA follows easily from the completeness of A (Exercise).■

Lemma II.1.5. *Let $(c_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{K}$ and $r > 0$ with*

$$\sum_{n=0}^{\infty} |c_n| r^n \ < \ \infty.$$

Further let A be a finite-dimensional unital Banach algebra. Then

$$f\colon B_r(0) := \{x \in A\colon \|x\| < r\} \to A, \quad x \mapsto \sum_{n=0}^{\infty} c_n x^n$$

defines a smooth function. Its derivative is given by

$$\mathtt{d}f(x) = \sum_{n=0}^{\infty} c_n \mathtt{d}p_n(x),$$

where $p_n(x) = x^n$ is the nth power map whose derivative is given by

$$\mathtt{d}p_n(x)y = x^{n-1}y + x^{n-2}yx + \ldots + xyx^{n-2} + yx^{n-1}.$$

For $\|x\| < r$ and $y \in M_d(\mathbb{K})$ with $xy = yx$ we obtain in particular

$$\mathtt{d}p_n(x)y = nx^{n-1}y \quad \text{and} \quad \mathtt{d}f(x)y = \sum_{n=1}^{\infty} c_n n x^{n-1} y.$$

Proof. We observe that the series defining $f(x)$ converges for $\|x\| < r$ by the Comparison Test (for series in Banach spaces). We shall prove by induction over $k \in \mathbb{N}$ that all such functions f are C^k-functions.

Step 1: First we show that f is a C^1-function. We define $\alpha_n \colon A \to A$ by

$$\alpha_n(h) := x^{n-1}h + x^{n-2}hx + \ldots + xhx^{n-2} + hx^{n-1}.$$

Then α_n is a continuous linear map with $\|\alpha_n\| \leq n\|x\|^{n-1}$. Furthermore

$$p_n(x+h) = (x+h)^n = x^n + \alpha_n(h) + r_n(h),$$

where

$$\begin{aligned}\|r_n(h)\| &\leq \binom{n}{2}\|h\|^2\|x\|^{n-2} + \binom{n}{3}\|h\|^3\|x\|^{n-3} + \ldots + \|h\|^n \\ &= \sum_{k\geq 2}\binom{n}{k}\|h\|^k\|x\|^{n-k}.\end{aligned}$$

In particular $\lim_{h\to 0}\frac{\|r_n(h)\|}{\|h\|} = 0$, and therefore p_n is differentiable in x with $\mathtt{d}p_n(x) = \alpha_n$. The series

$$\beta(h) := \sum_{n=0}^{\infty} c_n\alpha_n(h)$$

converges absolutely in $\mathrm{End}(A)$ by the Ratio Test since $\|x\| < r$:

$$\sum_{n=0}^{\infty}|c_n|\|\alpha_n\| \leq \sum_{n=0}^{\infty}|c_n| \cdot n \cdot \|x\|^{n-1} < \infty.$$

We thus obtain a linear map $\beta(x) \in \mathrm{End}(A)$ for each x with $\|x\| < r$.

Now let h satisfy $\|x\| + \|h\| < r$, i.e., $\|h\| < r - \|x\|$. Then

$$f(x+h) = f(x) + \beta(x)(h) + r(h), \quad r(h) := \sum_{n=2}^{\infty} c_n r_n(h),$$

where

$$\begin{aligned}\|r(h)\| &\leq \sum_{n=2}^{\infty}|c_n|\|r_n(h)\| \leq \sum_{n=2}^{\infty}|c_n|\sum_{k=2}^{n}\binom{n}{k}\|h\|^k\|x\|^{n-k} \\ &\leq \sum_{k=2}^{\infty}\Big(\sum_{n=k}^{\infty}|c_n|\binom{n}{k}\|x\|^{n-k}\Big)\|h\|^k < \infty\end{aligned}$$

follows from $\|x\| + \|h\| < r$ because

$$\sum_k\sum_{n\geq k}|c_n|\binom{n}{k}\|x\|^{n-k}\|h\|^k = \sum_n|c_n|(\|x\| + \|h\|)^n \leq \sum_n|c_n|r^n < \infty.$$

Therefore the continuity of real-valued functions represented by a power series yields

$$\lim_{h\to 0}\frac{\|r(h)\|}{\|h\|} = \sum_{k=2}^{\infty}\Big(\sum_{n=k}^{\infty}|c_n|\binom{n}{k}\|x\|^{n-k}\Big)0^{k-1} = 0.$$

This proves that f is a C^1-function with the required derivative.

Step 2: To complete our proof by induction, we now show that if all functions f as above are C^k, then they are also C^{k+1}. In view of Step 1, this implies that they are smooth.

To set up the induction, we consider the Banach algebra TA from Lemma II.1.4 and apply Step 1 to this algebra to obtain a smooth function

$$F\colon \{x+\varepsilon h \in TA\colon \|x\|+\|h\| = \|x+\varepsilon h\| < r\} \to TA, \quad F(x+\varepsilon h) = \sum_{n=0}^{\infty} c_n\cdot(x+\varepsilon h)^n,$$

We further note that

$$(x+\varepsilon h)^n = x^n + \mathtt{d}p_n(x)h\cdot\varepsilon.$$

This implies the formula

$$F(x+\varepsilon h) = f(x) + \varepsilon\mathtt{d}f(x)h,$$

i.e., that the extension F of f to TA describes the first order Taylor expansion of f in each point $x\in A$. Our induction hypothesis implies that F is a C^k-function.

Let $x_0\in A$ with $\|x_0\| < r$ and pick a basis $h_1,\ldots,h_d$ of A with $\|h_i\| < r-\|x_0\|$. Then all functions $x\mapsto \mathtt{d}f(x)h_i$ are defined and C^k on a neighborhood of x_0, and this implies that the function

$$B_r(0)\to \mathrm{Hom}(A,A), \quad x\mapsto \mathtt{d}f(x)$$

is C^k. This in turn implies that f is C^{k+1}. ■

The following proposition shows in particular that inserting elements of a Banach algebra in power series is compatible with composition.

Proposition II.1.6. (a) *On the set P_R of power series of the form*

$$f(z) := \sum_{n=0}^{\infty} a_n z^n, \quad a_n\in\mathbb{K}$$

and converging on the open disc $B_R(0) := \{z\in\mathbb{K}\colon |z| < R\}$, we define for $r < R$:

$$\|f\|_r := \sum_{n=0}^{\infty}|a_n|r^n.$$

Then $\|\cdot\|_r$ *is a norm with the following properties:*

(1) $\|\cdot\|_r$ *is submultiplicative:* $\|fg\|_r \leq \|f\|_r \|g\|_r$.

(2) *The polynomials* $f_N(z) := \sum_{n=0}^N a_n z^n$ *satisfy* $\|f - f_N\|_r \to 0$.

(3) *If* $A \in M_n(\mathbb{K})$ *satisfies* $\|A\| < R$, *then* $f(A) := \sum_{n=0}^\infty a_n A^n$ *converges. We further have*

$$\|f(A)\| \leq \|f\|_r \quad \text{for} \quad \|A\| \leq r < R$$

and for $f, g \in P_R$ *we have*

$$(f \cdot g)(A) = f(A)g(A).$$

(b) *If* $g \in P_S$ *with* $\|g\|_s < R$ *for all* $s < S$ *and* $f \in P_R$, *then* $f \circ g \in P_S$ *defines an analytic function on the open disc of radius* S, *and for* $A \in M_n(\mathbb{K})$ *with* $\|A\| < S$ *we have* $\|g(A)\| < R$ *and the Composition Formula*

$$f(g(A)) = (f \circ g)(A). \tag{1.1}$$

Proof. (1) First we note that P_R is the set of all power series $f(z) = \sum_{n=0}^\infty a_n z^n$ for which $\|f\|_r < \infty$ holds for all $r < R$. We leave the easy argument that $\|\cdot\|_r$ is a norm to the reader. If $\|f\|_r, \|g\|_r < \infty$ holds for $g(z) = \sum_{n=0}^\infty b_n z^n$, then the Cauchy Product Formula implies that

$$\|fg\|_r = \sum_{n=0}^\infty \Big| \sum_{k=0}^n a_k b_{n-k} \Big| r^n \leq \sum_{n=0}^\infty \sum_{k=0}^n |a_k|\, |b_{n-k}| r^k r^{n-k} = \|f\|_r \|g\|_r.$$

(2) follows immediately from $\|f - f_N\|_r = \sum_{n>N} |a_n| r^n \to 0$.

(3) The relation $\|f(A)\| \leq \|f\|_r$ follows from $\|a_n A^n\| \leq |a_n| r^n$ and the Comparison Test for absolutely convergent series in a Banach space. The relation $(f \cdot g)(A) = f(A)g(A)$ follows from the Cauchy Product Formula (Exercise II.1.3) because the series $f(A)$ and $g(A)$ converge absolutely.

(b) We may w.l.o.g. assume that $\mathbb{K} = \mathbb{C}$ because everything on the case $\mathbb{K} = \mathbb{R}$ can be obtained by restriction. Our assumption implies that $g(B_S(0)) \subseteq B_R(0)$, so that $f \circ g$ defines a holomorphic function on the open disc $B_S(0)$. For $s < S$ and $\|g\|_s < r < R$ we then derive

$$\|f \circ g\|_s \leq \sum_{n=0}^\infty \|a_n g^n\|_s \leq \sum_{n=0}^\infty |a_n| \|g\|_s^n \leq \|f\|_r.$$

For $s := \|A\|$ we obtain $\|g(A)\| \leq \|g\|_s < R$, so that $f(g(A))$ is defined. For $s < r < R$ we then have

$$\|f(g(A)) - f_N(g(A))\| \leq \|f - f_N\|_r \to 0.$$

Likewise

$$\|(f \circ g)(A) - (f_N \circ g)(A)\| \leq \|(f \circ g) - (f_N \circ g)\|_s \leq \|f - f_N\|_r \to 0,$$

and we get

$$(f \circ g)(A) = \lim_{N \to \infty} (f_N \circ g)(A) = \lim_{N \to \infty} f_N(g(A)) = f(g(A))$$

because the Composition Formula trivially holds if f is a polynomial. ■

Exercises for Section II.1

Exercise II.1.1. Let $X_1, \ldots, X_n$ be finite-dimensional normed spaces and $\beta: X_1 \times \ldots \times X_n \to Y$ an n-linear map.
(a) Show that β is continuous. Hint: Choose a basis in each space X_j and expand β accordingly.
(b) Show that there exists a constant $C \geq 0$ with

$$\|\beta(x_1, \ldots, x_n)\| \leq C\|x_1\| \cdots \|x_n\| \quad \text{for} \quad x_i \in X_i. \qquad \blacksquare$$

Exercise II.1.2. Let Y be a Banach space and $a_{n,m}$, $n, m \in \mathbb{N}$, elements in Y with

$$\sum_{n,m} \|a_{n,m}\| := \sup_{N \in \mathbb{N}} \sum_{n,m \leq N} \|a_{n,m}\| < \infty.$$

(a) Show that

$$A := \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{n,m} = \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{n,m}$$

and that both iterated sums exist.
(b) Show that for each sequence $(S_n)_{n \in \mathbb{N}}$ of finite subsets $S_n \subseteq \mathbb{N} \times \mathbb{N}$, $n \in \mathbb{N}$, with $S_n \subseteq S_{n+1}$ and $\bigcup_n S_n = \mathbb{N} \times \mathbb{N}$ we have

$$A = \lim_{n \in \mathbb{N}} \sum_{(j,k) \in S_n} a_{j,k}. \qquad \blacksquare$$

Exercise II.1.3. (Cauchy Product Formula) Let X, Y, Z be Banach space and $\beta: X \times Y \to Z$ a continuous bilinear map. Suppose that $x := \sum_{n=0}^{\infty} x_n$ is absolutely convergent in X and that $y := \sum_{n=0}^{\infty} y_n$ is absolutely convergent in Y. Then

$$\beta(x, y) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \beta(x_k, y_{n-k}).$$

Hint: Use Exercises II.1.1(b) and II.1.2(b). $\blacksquare$

II.2. Elementary properties of the exponential function

After the preparations of the preceding section, it is now easy to see that the matrix exponential function defines a smooth map on $M_n(\mathbb{K})$. In this section we describe some elementary properties of this function. As group theoretic consequences for $\mathrm{GL}_n(\mathbb{K})$, we show that it has no small subgroups and that all one-parameter groups are smooth and given by the exponential function.

For $x \in M_n(\mathbb{K})$ we define

$$e^x := \sum_{n=0}^{\infty} \frac{1}{n!} x^n. \tag{2.1}$$

The absolute convergence of the series on the right follows directly from the estimate

$$\sum_{n=0}^{\infty} \frac{1}{n!} \|x^n\| \leq \sum_{n=0}^{\infty} \frac{1}{n!} \|x\|^n = e^{\|x\|}$$

and the Comparison Test for absolute convergence of a series in a Banach space. We define the *exponential function of* $M_n(\mathbb{K})$ by

$$\exp\colon M_n(\mathbb{K}) \to M_n(\mathbb{K}), \quad \exp(x) := e^x.$$

Proposition II.2.1. *The exponential function* $\exp\colon M_d(\mathbb{K}) \to M_d(\mathbb{K})$ *is smooth. For* $xy = yx$ *we have*

$$\mathtt{d}\exp(x)y = \exp(x)y = y\exp(x) \tag{2.2}$$

and in particular

$$\mathtt{d}\exp(0) = \mathrm{id}_{M_n(\mathbb{K})}.$$

Proof. To verify the formula for the differential, we note that for $xy = yx$, Lemma II.1.5 implies that

$$\mathtt{d}\exp(x)y = \sum_{n=1}^{\infty} \frac{1}{n!} n x^{n-1} y = \sum_{n=0}^{\infty} \frac{1}{n!} x^n y = \exp(x)y.$$

For $x = 0$, the relation $\exp(0) = \mathbf{1}$ now implies in particular that $\mathtt{d}\exp(0)y = y$. ■

Lemma II.2.2. *Let* $x, y \in M_n(\mathbb{K})$.

(i) *If* $xy = yx$, *then* $\exp(x+y) = \exp x \exp y$.

(ii) $\exp(M_n(\mathbb{K})) \subseteq \mathrm{GL}_n(\mathbb{K})$, $\exp(0) = \mathbf{1}$, *and* $(\exp x)^{-1} = \exp(-x)$.

(iii) *For* $g \in \mathrm{GL}_n(\mathbb{K})$ *we have* $g e^x g^{-1} = e^{gxg^{-1}}$.

Proof. (i) Using the general form of the Cauchy–Product Formula (Exercise II.1.3), we obtain

$$\begin{aligned} \exp(x+y) &= \sum_{k=0}^{\infty} \frac{(x+y)^k}{k!} = \sum_{k=0}^{\infty} \frac{1}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} x^\ell y^{k-\ell} \\ &= \sum_{k=0}^{\infty} \sum_{\ell=0}^{k} \frac{x^\ell}{\ell!} \frac{y^{k-\ell}}{(k-\ell)!} = \Big(\sum_{p=0}^{\infty} \frac{x^p}{p!}\Big)\Big(\sum_{\ell=0}^{\infty} \frac{y^\ell}{\ell!}\Big). \end{aligned}$$

(ii) From (i) we derive in particular $\exp x \exp(-x) = \exp 0 = \mathbf{1}$, which implies (ii).

(iii) is a consequence of $gx^n g^{-1} = (gxg^{-1})^n$ and the continuity of the conjugation map $c_g(x) := gxg^{-1}$ on $M_n(\mathbb{K})$. ■

Remark II.2.3. (a) For $n = 1$, the exponential function

$$\exp\colon M_1(\mathbb{R}) \cong \mathbb{R} \to \mathbb{R}^\times \cong \mathrm{GL}_1(\mathbb{R}), \quad x \mapsto e^x$$

is injective, but this is not the case for $n > 1$. In fact,

$$\exp\begin{pmatrix} 0 & -2\pi \\ 2\pi & 0 \end{pmatrix} = \mathbf{1}$$

follows from

$$\exp\begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}, \quad t \in \mathbb{R}.$$

This example is nothing but the real picture of the relation $e^{2\pi i} = 1$. ■

Proposition II.2.4. *There exists an open neighborhood U of 0 in $M_n(\mathbb{K})$ such that the map*

$$\exp|_U\colon U \to \mathrm{GL}_n(\mathbb{K})$$

is a diffeomorphism onto an open neighborhood of $\mathbf{1}$ in $\mathrm{GL}_n(\mathbb{K})$.

Proof. We have already seen that exp is a smooth map, and that $\mathtt{d}\exp(\mathbf{0}) = \mathrm{id}_{M_n(\mathbb{K})}$. Therefore the assertion follows from the Inverse Function Theorem. ■

If U is as in Proposition II.2.4, we define

$$\log_V\colon = (\exp|_V)^{-1}\colon V \to U \subseteq M_n(\mathbb{K}).$$

We shall see below why this function deserves to be called a *logarithm function*.

The following corollary means that the group $\mathrm{GL}_n(\mathbb{K})$ contains no subgroups that are small in the sense that they lie arbitrarily close to the identity.

No Small Subgroup Theorem

Theorem II.2.5. *There exists an open neighborhood V of $\mathbf{1}$ in $\mathrm{GL}_n(\mathbb{K})$ such that $\{\mathbf{1}\}$ is the only subgroup of $\mathrm{GL}_n(\mathbb{K})$ contained in V.*

Proof. Let U be as in Proposition II.2.4 and assume furthermore that U is convex and bounded. We set $U_1 := \frac{1}{2}U$. Let $G \subseteq V := \exp U_1$ be a subgroup of $\mathrm{GL}_n(\mathbb{K})$ and $g \in G$. Then we write $g = \exp x$ with $x \in U_1$ and assume that $x \neq 0$. Let $k \in \mathbb{N}$ be maximal with $kx \in U_1$ (the existence of k follows from the boundedness of U). Then

$$g^{k+1} = \exp(k+1)x \in G \subseteq V$$

implies the existence of $y \in U_1$ with $\exp(k+1)x = \exp y$. Since $(k+1)x \in 2U_1 = U$ follows from $\frac{k+1}{2}x \in [0,k]x \subseteq U_1$, and $\exp|_U$ is injective, we obtain $(k+1)x = y \in U_1$, contradicting the maximality of k. Therefore $g = \mathbf{1}$. ■

A *one-parameter (sub)group* of a group G is a group homomorphism $\gamma\colon (\mathbb{R}, +) \to G$. The following result describes all differentiable one-parameter subgroups of $\mathrm{GL}_n(\mathbb{K})$.

One-parameter Group Theorem

Theorem II.2.6. *For each $x \in M_d(\mathbb{K})$ the map*

$$\gamma\colon (\mathbb{R},+) \to \mathrm{GL}_d(\mathbb{K}), \quad t \mapsto \exp(tx)$$

is a smooth group homomorphism solving the initial value problem

$$\gamma(0) = \mathbf{1} \quad \text{and} \quad \gamma'(t) = \gamma(t)x \quad \text{for } t \in \mathbb{R}.$$

Conversely, every continuous one-parameter group $\gamma\colon \mathbb{R} \to \mathrm{GL}_d(\mathbb{K})$ is of this form.

Proof. In view of Lemma II.2.2(i) and the differentiability of $\exp$, we have

$$\lim_{h\to 0} \frac{1}{h}\big(\gamma(t+h) - \gamma(t)\big) = \lim_{h\to 0} \frac{1}{h}\big(\gamma(t)\gamma(h) - \gamma(t)\big) = \gamma(t) \lim_{h\to 0} \frac{1}{h}\big(e^{hx} - \mathbf{1}\big) = \gamma(t)x.$$

Hence γ is differentiable with

$$\gamma'(t) = x\gamma(t) = \gamma(t)x.$$

From that it immediately follows that γ is smooth with $\gamma^{(n)}(t) = x^n\gamma(t)$ for each $n \in \mathbb{N}$.

Although we won't need it for the completeness of the proof, we first show that each one-parameter group $\gamma\colon \mathbb{R} \to \mathrm{GL}_d(\mathbb{K})$ which is differentiable in 0 has the required form. For $x := \gamma'(0)$, the calculation

$$\gamma'(t) = \lim_{s\to 0} \frac{\gamma(t+s) - \gamma(t)}{s} = \lim_{s\to 0} \gamma(t)\frac{\gamma(s) - \gamma(0)}{s} = \gamma(t)\gamma'(0) = \gamma(t)x$$

implies that γ is differentiable and solves the initial value problem

$$\gamma'(t) = \gamma(t)x, \qquad \gamma(0) = \mathbf{1}.$$

Therefore the Uniqueness Theorem for Linear Differential Equations implies that $\gamma(t) = \exp tx$ for all $t \in \mathbb{R}$.

It remains to show that each continuous one-parameter group γ of $\mathrm{GL}_d(\mathbb{K})$ is differentiable in 0. As in the proof of Theorem II.2.5, let U be a convex symmetric (i.e., $U = -U$) 0-neighborhood in $M_d(\mathbb{K})$ as in Proposition II.2.4 and $U_1 := \frac{1}{2}U$. Since γ is continuous in 0, there exists an $\varepsilon > 0$ such that $\gamma([-\varepsilon,\varepsilon]) \subseteq \exp(U_1)$. Then $\alpha(t) := (\exp|_U)^{-1}(\gamma(t))$ defines a continuous curve $\alpha\colon [-\varepsilon,\varepsilon] \to U_1$ with $\exp(\alpha(t)) = \gamma(t)$ for $|t| \leq \varepsilon$. For any such t we then have

$$\exp\big(2\alpha(\tfrac{t}{2})\big) = \exp(\alpha(\tfrac{t}{2}))^2 = \gamma(\tfrac{t}{2})^2 = \gamma(t) = \exp(\alpha(t)),$$

so that the injectitivy of $\exp$ on U yields

$$\alpha(\tfrac{t}{2}) = \frac{1}{2}\alpha(t) \quad \text{for} \quad |t| \leq \varepsilon.$$

Inductively, we thus obtain

$$\alpha(\frac{t}{2^k}) = \frac{1}{2^k}\alpha(t) \quad \text{for} \quad |t| \leq \varepsilon, k \in \mathbb{N}. \tag{2.3}$$

In particular, we obtain

$$\alpha(t) \in \frac{1}{2^k}U_1 \quad \text{for} \quad |t| \leq \frac{\varepsilon}{2^k}.$$

For $n \in \mathbb{Z}$ with $|n| \leq 2^k$ and $|t| \leq \frac{\varepsilon}{2^k}$ we now have $|nt| \leq \varepsilon$, $n\alpha(t) \in \frac{n}{2^k}U_1 \subseteq U_1$, and

$$\exp(n\alpha(t)) = \gamma(t)^n = \gamma(nt) = \exp(\alpha(nt)).$$

Therefore the injectivity of $\exp$ on U_1 yields

$$\alpha(nt) = n\alpha(t) \quad \text{for} \quad n \leq 2^k, |t| \leq \frac{\varepsilon}{2^k}. \tag{2.4}$$

Combining (2.3) and (2.4), leads to

$$\alpha(\frac{n}{2^k}t) = \frac{n}{2^k}\alpha(t) \quad \text{for} \quad |t| \leq \varepsilon, k \in \mathbb{N}, |n| \leq 2^k.$$

Since the set of all numbers $\frac{nt}{2^k}$, $n \in \mathbb{Z}$, $|n| \leq 2^k$, is dense in the interval $[-t, t]$, the continuity of α implies that

$$\alpha(t) = \frac{t}{\varepsilon}\alpha(\varepsilon) \quad \text{for} \quad |t| \leq \varepsilon.$$

In particular, α is smooth and of the form $\alpha(t) = tx$ for some $x \in M_d(\mathbb{K})$. Hence $\gamma(t) = \exp(tx)$ for $|t| \leq \varepsilon$, but then $\gamma(nt) = \exp(ntx)$ for $n \in \mathbb{N}$ leads to $\gamma(t) = \exp(tx)$ for each $t \in \mathbb{R}$. ∎

Exercises for Section II.2

Exercise II.2.1. Let $D \in M_n(\mathbb{K})$ be a diagonal matrix. Calculate its operator norm. ∎

Exercise II.2.2. If A is a Banach algebra with unit element $\mathbf{1}$ and $g \in A$ satisfies $\|g - \mathbf{1}\| < 1$, then g is invertible, i.e., there exists an element $h \in A$ with $hg = gh = \mathbf{1}$. Hint: For $x := \mathbf{1} - g$ the *Neumann series* $y := \sum_{n=0}^{\infty} x^n$ converges. Show that y is an inverse of g. ∎

Exercise II.2.3. (a) Calculate e^{tN} for $t \in \mathbb{K}$ and the matrix

$$N = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \cdot & 0 & 1 & 0 & \cdot \\ \cdot & & \cdot & \cdot & \cdot \\ \cdot & & & \cdot & 1 \\ 0 & & \dots & & 0 \end{pmatrix} \in M_n(\mathbb{K}).$$

(b) If A is a block diagonal matrix $\mathrm{diag}(A_1, \dots, A_k)$, then e^A is the block diagonal matrix $\mathrm{diag}(e^{A_1}, \dots, e^{A_k})$.
(c) Calculate e^{tA} for a matrix $A \in M_n(\mathbb{C})$ given in Jordan Normal Form. Hint: Use (a) and (b). ■

Exercise II.2.4. Recall that a matrix x is said to be *nilpotent* if x^d for some $d \in \mathbb{N}$ and y is called *unipotent* if $y - \mathbf{1}$ is nilpotent.

Let $a, b \in M_n(\mathbb{K})$ be commuting matrices.

(a) If a and b are nilpotent, then $a + b$ is nilpotent.
(b) If a and b are diagonalizable, then $a + b$ and ab are diagonalizable.
(c) If a and b are unipotent, then ab is unipotent. ■

Exercise II.2.5. (Jordan decomposition)
(a) (Additive Jordan decomposition) Show that each complex matrix $X \in M_n(\mathbb{C})$ can be written in a unique fashion as

$$X = X_s + X_n \quad \text{with} \quad [X_s, X_n] = 0,$$

where X_n is nilpotent and X_s diagonalizable. Hint: Existence (Jordan normal form), Uniqueness (what can you say about nilpotent diagonalizable matrices?).
(b) (Multiplicative Jordan decomposition) Show that each invertible complex matrix $g \in \mathrm{GL}_n(\mathbb{C})$ can be written in a unique fashion as

$$g = g_s g_u, \quad \text{with} \quad g_s g_u = g_u g_s,$$

where g_u is unipotent and g_s diagonalizable. Hint: Existence: Put $g_u := \mathbf{1} + g_s^{-1} g_n$.
(c) If $X = X_s + X_n$ is the additive Jordan decomposition, then $e^X = e^{X_s} e^{X_n}$ is the multiplicative Jordan decomposition of e^X.
(d) $A \in M_n(\mathbb{C})$ commutes with a diagonalizable matrix D if and only if A preserves all eigenspaces of D.
(e) $A \in M_n(\mathbb{C})$ commutes with X if and only if it commutes with X_s and X_n. Hint: If A commutes with X, it preserves the generalized eigenspaces of X (verify this!), and this implies that it commutes with X_s, which is diagonalizable and whose eigenspaces are the generalized eigenspaces of X. ■

Exercise II.2.6. Let $A \in M_n(\mathbb{C})$. Show that the set

$$e^{\mathbb{R}A} = \{e^{tA} : t \in \mathbb{R}\}$$

is bounded in $M_n(\mathbb{C})$ if and only if A is diagonalizable with purely imaginary eigenvalues. Hint: Choose a matrix $g \in \mathrm{GL}_n(\mathbb{C})$ for which $A' := gAg^{-1}$ is in Jordan normal form $A' = D + N$ (D diagonal and N strictly upper triangular). Then show that the boundedness of $e^{\mathbb{R}A}$ implies $N = 0$ and the boundedness of the subset $e^{\mathbb{R}D}$. ∎

Exercise II.2.7. Show that:
(a) $\exp(M_n(\mathbb{R}))$ is contained in the identity component $\mathrm{GL}_n(\mathbb{R})_+$ of $\mathrm{GL}_n(\mathbb{R})$. In particular the exponential function of $\mathrm{GL}_n(\mathbb{R})$ is not surjective because $\mathrm{GL}_n(\mathbb{R})$ is not connected.
(b) The exponential function

$$\exp : M_2(\mathbb{R}) \to \mathrm{GL}_2(\mathbb{R})_+$$

is not surjective. Hint: Use the Jordan normal form to derive some information on the eigenvalues of matrices of the form e^x which is not satisfied by all elements of $\mathrm{GL}_2(\mathbb{R})_+$. (Either the spectrum is contained in the positive axis or its consists of two mutually conjugate complex numbers). The matrix $g := \begin{pmatrix} -1 & 0 \\ 0 & -2 \end{pmatrix}$ is not contained in the image of $\exp$.
(c) Give also a direct argument why g is not of the form e^X. Hint: e^X commutes with X. ∎

Exercise II.2.8. Let $V \subseteq M_n(\mathbb{C})$ be a commutative subspace, i.e., an abelian Lie subalgebra. Then $A := e^V$ is an abelian subgroup of $\mathrm{GL}_n(\mathbb{C})$ and

$$\exp : (V, +) \to (A, \cdot)$$

is a group homomorphism whose kernel consists of diagonalizable elements whose eigenvalues are contained in $2\pi i\mathbb{Z}$. Hint: Lemma II.2.1, Exercise II.2.6. ∎

Exercise II.2.9. For $X, Y \in M_n(\mathbb{C})$ the following are equivalent:
(1) $e^X = e^Y$.
(2) $X_n = Y_n$ holds for the nilpotent Jordan components and $e^{X_s} = e^{Y_s}$. ∎

Exercise II.2.10. For $A \in M_n(\mathbb{C})$ we have $e^A = \mathbf{1}$ if and only if A is diagonalizable with all eigenvalues contained in $2\pi i\mathbb{Z}$. Hint: Exercise II.2.9. ∎

II.3. The logarithm function

In this section we apply the tools from Section II.1 to the logarithm series. Since this series has the radius of convergence 1, it defines a smooth function $B_1(\mathbf{1}) \to M_n(\mathbb{K})$, and we shall see that it provides a smooth inverse of the exponential function.

Lemma II.3.1. *The series*

$$\log(\mathbf{1}+x) := \sum_{k=1}^{\infty}(-1)^{k+1}\frac{x^k}{k}$$

converges for $x \in M_d(\mathbb{K})$ *with* $\|x\| < 1$ *and defines a smooth function*

$$\log\colon B_1(\mathbf{1}) \to M_d(\mathbb{K}).$$

For $\|x\| < 1$ *and* $y \in M_d(\mathbb{K})$ *with* $xy = yx$ *we have*

$$(\mathtt{d}\log)(\mathbf{1}+x)y = (\mathbf{1}+x)^{-1}y.$$

Proof. The convergence follows from

$$\sum_{k=1}^{\infty}(-1)^{k+1}\frac{r^k}{k} = \log(1+r) < \infty$$

for $r < 1$, so that the smoothness follows from Lemma II.1.5.

If x and y commute, then the formula for the derivative in Lemma II.1.5 leads to

$$(\mathtt{d}\log)(\mathbf{1}+x).y = \sum_{k=1}^{\infty}(-1)^{k+1}x^{k-1}y = (\mathbf{1}+x)^{-1}y$$

(here we used the Neumann series; cf. Exercise II.2.2). ■

Proposition II.3.2. (a) *For* $x \in M_d(\mathbb{K})$ *with* $\|x\| < \log 2$ *we have*

$$\log(\exp x) = x.$$

(b) *For* $g \in \mathrm{GL}_d(\mathbb{K})$ *with* $\|g - \mathbf{1}\| < 1$ *we have* $\exp(\log g) = g$.

Proof. (a) We apply Proposition II.1.6 with $\exp \in P_S$, $S = \log 2$, $R = e^{\log 2} = 2$ and $\|\exp\|_s \leq e^s \leq e^S = 2$ for $s < S$. We thus obtain $\log(\exp x) = x$ for $\|x\| < \log 2$.

(b) Next we apply Proposition II.1.6 with $f = \exp$, $S = 1$ and $g(z) = \log(1+z)$ to obtain $\exp(\log g) = g$. ■

The exponential function on nilpotent matrices

Proposition II.3.3. *Let*

$$U := \{g \in \mathrm{GL}_d(\mathbb{K}) \colon (g - \mathbf{1})^d = 0\}$$

be the set of unipotent matrices and

$$N := \{x \in M_d(\mathbb{K}) \colon x^d = 0\}$$

the set of nilpotent matrices. Then $U = \mathbf{1} + N$ *and*

$$\exp_N := \exp|_N \colon N \to U$$

is a homeomorphism whose inverse is given by

$$\log_U \colon g \mapsto \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(g - \mathbf{1})^k}{k} = \sum_{k=1}^{d-1} (-1)^{k+1} \frac{(g - \mathbf{1})^k}{k}.$$

Proof. First we observe that for $x \in N$ we have

$$e^x - \mathbf{1} = xa \quad \text{with} \quad a := \sum_{n=1}^{d} \frac{1}{n!} x^{n-1}.$$

In view of $xa = ax$, this leads to $(e^x - \mathbf{1})^d = x^d a^d = 0$. Therefore $\exp_N(N) \subseteq U$. Similarly we obtain for $g \in U$ that

$$\log_U(g) = (g - \mathbf{1}) \sum_{k=1}^{d} (-1)^{k+1} \frac{(g - \mathbf{1})^{k-1}}{k} \in N.$$

For $x \in N$ the curve

$$F \colon \mathbb{R} \to M_d(\mathbb{K}), \quad t \mapsto \log_U \exp_N(tx)$$

is a polynomial function and Proposition II.3.2 implies that $F(t) = tx$ for $\|tx\| < \log 2$. This imples that $F(t) = tx$ for each $t \in \mathbb{R}$ and hence that $\log_U \exp_N(x) = F(1) = x$.

Likewise we see that for $g = \mathbf{1} + x \in U$ the curve

$$G \colon \mathbb{R} \to M_d(\mathbb{K}), \quad t \mapsto \exp_N \log_U(\mathbf{1} + tx)$$

is polynomial with $G(t) = \mathbf{1} + tx$ for $\|tx\| < 1$. Therefore $\exp_N \log_U(g) = F(1) = \mathbf{1} + x = g$. This proves that the functions $\exp_N$ and $\log_U$ are inverse to each other. ■

Corollary II.3.4. *Let* $X \in \mathrm{End}(V)$ *be a nilpotent endomorphism of the* $\mathbb{K}$*-vector space* V *and* $v \in V$*. Then the following are equivalent:*
(1) $X.v = 0$.
(2) $e^X.v = v$.

Proof. Clearly $X.v = 0$ implies $e^X.v = \sum_{n=0}^\infty \frac{1}{n!}X^n.v = v$.

If, conversely, $e^X.v = v$, then

$$X.v = \log(e^X).v = \sum_{k=1}^\infty (-1)^{k+1}\frac{(e^X - \mathbf{1})^k}{k}.v = 0. \qquad \blacksquare$$

The exponential function on hermitian matrices

For the following proof we recall that for a hermitian $d \times d$-matrix A we have

$$\|A\| = \max\{|\lambda|\colon \det(A - \lambda\mathbf{1}) = 0\}$$

(Exercise II.3.1).

Proposition II.3.5. *The restriction*

$$\exp_P := \exp|_{\mathrm{Herm}_d(\mathbb{K})}\colon \mathrm{Herm}_d(\mathbb{K}) \to \mathrm{Pd}_d(\mathbb{K})$$

is a diffeomorphism onto the open subset $\mathrm{Pd}_d(\mathbb{K})$ *of* $\mathrm{Herm}_d(\mathbb{K})$.

Proof. For $x^* = x$ we have $(e^x)^* = e^{x^*}$, which implies that $\exp x$ is hermitian if x is hermitian. Moreover, if $\lambda_1, \ldots, \lambda_n$ are the real eigenvalues of x, then $e^{\lambda_1}, \ldots, e^{\lambda_n}$ are the eigenvalues of e^x. Therefore e^x is positive definite for each hermitian matrix x.

If, conversely, $g \in \mathrm{Pd}_d(\mathbb{K})$, then let $v_1, \ldots, v_n$ be an orthonormal basis of eigenvectors for g with $g.v_j = \lambda_j v_j$. Then $\lambda_j > 0$ for each j, and we define $\log_H(g) \in \mathrm{Herm}_d(\mathbb{K})$ by $\log_H(g).v_j := (\log \lambda_j)v_j$, $j = 1, \ldots, n$. From this construction of the logarithm function it is clear that

$$\log_H \circ \exp_P = \mathrm{id}_{\mathrm{Herm}_d(\mathbb{K})} \quad \text{and} \quad \exp_P \circ \log_H = \mathrm{id}_{\mathrm{Pd}_d(\mathbb{K})}.$$

For two real numbers $x, y > 0$ we have

$$\log(xy) = \log x + \log y.$$

From this we obtain for $\lambda > 0$ the relation

$$\log_H(\lambda g) = (\log \lambda) \cdot \mathbf{1} + \log_H(g) \tag{3.1}$$

by following what happens on each eigenspace of g.

The relation

$$\log(x) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(x-\mathbf{1})^k}{k}$$

for $x \in \mathbb{R}$ with $|x-1| < 1$ implies that for $\|g - \mathbf{1}\| < 1$ we have

$$\log_H(g) = \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(g-\mathbf{1})^k}{k}.$$

This proves that $\log_H$ is smooth in $B_1(\mathbf{1}) \cap \mathrm{Herm}_d(\mathbb{K})$, hence in a neighborhood of g_0 if $\|g_0 - \mathbf{1}\| < 1$ (Lemma II.3.1), which means that for each eigenvalue μ of g_0 we have $|\mu - 1| < 1$ (Exercise II.2.1). If this condition is not satisfied, then we choose $\lambda > 0$ such that $\|\lambda g\| < 2$. Then $\|\lambda g - \mathbf{1}\| < 1$, and we obtain with (3.1) the formula

$$\log_H(g) = -(\log\lambda)\mathbf{1} + \log_H(\lambda g) = -(\log\lambda)\mathbf{1} + \sum_{k=1}^{\infty} (-1)^{k+1} \frac{(\lambda g - \mathbf{1})^k}{k}.$$

Therefore $\log_H$ is smooth on the whole open cone $\mathrm{Pd}_d(\mathbb{K})$, so that $\log_H = \exp_P^{-1}$ implies that $\exp_P$ is a diffeomorphism. ■

Corollary II.3.6. *The group* $\mathrm{GL}_d(\mathbb{K})$ *is homeomorphic to*

$$\mathrm{U}_d(\mathbb{K}) \times \mathbb{R}^{\dim_{\mathbb{R}}(\mathrm{Herm}_d(\mathbb{K}))} \quad \textit{with} \quad \dim_{\mathbb{R}}(\mathrm{Herm}_d(\mathbb{K})) = \begin{cases} \frac{d(d+1)}{2} & \textit{for } \mathbb{K} = \mathbb{R} \\ d^2 & \textit{for } \mathbb{K} = \mathbb{C} \end{cases}.$$

■

Exercises for Section II.3.

Exercise II.3.1. Show that for a hermitian matrix $A \in \mathrm{Herm}_n(\mathbb{K})$ and the euclidian norm $\|\cdot\|$ on $\mathbb{K}^n$ we have

$$\|A\| := \sup\{\|Ax\| \colon \|x\| \leq 1\} = \max\{|\lambda| \colon \mathrm{der}(A - \lambda\mathbf{1}) = 0\}.$$

Hint: Write $x \in \mathbb{K}^n$ as a sum $x = \sum_j x_j$, where $Ax_j = \lambda_j x_j$ and calculate $\|Ax\|^2$ in these term. ■

Exercise II.3.2. The exponential function

$$\exp \colon M_n(\mathbb{C}) \to \mathrm{GL}_n(\mathbb{C})$$

is surjective. Hint: Use the multiplicative Jordan decomposition: Each $g \in \mathrm{GL}_n(\mathbb{C})$ can be written in a unique way as $g = du$ with d diagonalizable and u unipotent with $du = ud$; see also Proposition II.3.3. ■

II.4. The Baker–Campbell–Hausdorff–Dynkin Formula

In this section we derive a formula which expresses the product $\exp x \exp y$ of two sufficiently small matrices x, y as the exponential image $\exp(x*y)$ of an element $x*y$ which can be described in terms of Lie brackets. The (local) multiplication $*$ is called the *Baker–Campbell–Hausdorff Multiplication* and the explicit series describing this product the *Dynkin series*.

The discussion of $x*y$ requires some preparation. We start with the *adjoint representation of* $\mathrm{GL}_n(\mathbb{K})$. This is the group homomorphism

$$\mathrm{Ad}\colon \mathrm{GL}_n(\mathbb{K}) \to \mathrm{Aut}(M_n(\mathbb{K})), \quad \mathrm{Ad}(g)x = gxg^{-1},$$

where $\mathrm{Aut}(M_n(\mathbb{K}))$ is the group of algebra automorphisms of $M_n(\mathbb{K})$. For $x \in M_n(\mathbb{K})$ we further define a linear map

$$\mathrm{ad}(x)\colon M_n(\mathbb{K}) \to M_n(\mathbb{K}), \quad \mathrm{ad}\, x(y) := [x, y].$$

Lemma II.4.1. *For each $x \in M_n(\mathbb{K})$ we have*

$$\mathrm{Ad}(\exp x) = \exp(\mathrm{ad}\, x).$$

Proof. We define the linear maps

$$L_x\colon M_n(\mathbb{K}) \to M_n(\mathbb{K}), \quad y \mapsto xy$$

and

$$R_x\colon M_n(\mathbb{K}) \to M_n(\mathbb{K}), \quad y \mapsto yx.$$

Then $L_x R_x = R_x L_x$ and $\mathrm{ad}\, x = L_x - R_x$. Therefore Lemma II.2.2(ii) leads to

$$\mathrm{Ad}(\exp x)y = e^x y e^{-x} = e^{L_x} e^{-R_x} y = e^{L_x - R_x} y = e^{\mathrm{ad}\, x} y. \qquad \blacksquare$$

THE DIFFERENTIAL OF THE EXPONENTIAL FUNCTION

Proposition II.4.2. *Let $x \in M_n(\mathbb{K})$ and $\lambda_{\exp x}(y) := (\exp x)y$ the left multiplication by $\exp x$. Then*

$$\mathtt{d}\exp(x) = \lambda_{\exp x} \circ \frac{\mathbf{1} - e^{-\mathrm{ad}\, x}}{\mathrm{ad}\, x}\colon M_n(\mathbb{K}) \to M_n(\mathbb{K}),$$

where the fraction on the right means $\Phi(\mathrm{ad}\, x)$ for the entire function

$$\Phi(z) := \frac{1 - e^{-z}}{z} = \sum_{k=1}^{\infty} \frac{(-z)^{k-1}}{k!}.$$

The series $\Phi(x)$ *converges for each* $x \in M_n(\mathbb{K})$*.*

Proof. First let $\alpha\colon [0,1] \to M_n(\mathbb{K})$ be a smooth curve. Then

$$\gamma(t,s) := \exp(-s\alpha(t))\frac{d}{dt}\exp(s\alpha(t))$$

defines a map $[0,1]^2 \to M_n(\mathbb{K})$ which is C^1 in each argument. We calculate

$$\begin{aligned}\frac{\partial\gamma}{\partial s}(t,s) &= \exp(-s\alpha(t))\cdot(-\alpha(t))\frac{d}{dt}\exp(s\alpha(t))\\ &\quad+\exp(-s\alpha(t))\cdot\frac{d}{dt}\Big(\alpha(t)\exp(s\alpha(t))\Big)\\ &= \exp(-s\alpha(t))\cdot(-\alpha(t))\frac{d}{dt}\exp(s\alpha(t))\\ &\quad+\exp(-s\alpha(t))\cdot\Big(\alpha'(t)\exp(s\alpha(t))+\alpha(t)\frac{d}{dt}\exp(s\alpha(t))\Big)\\ &= \mathrm{Ad}(\exp(-s\alpha(t)))\alpha'(t) = e^{-s\,\mathrm{ad}\,\alpha(t)}\alpha'(t).\end{aligned}$$

Integration over $[0,1]$ with respect to s now leads to

$$\gamma(t,1) = \gamma(t,0) + \int_0^1 e^{-s\,\mathrm{ad}\,\alpha(t)}\alpha'(t)\,ds = \int_0^1 e^{-s\,\mathrm{ad}\,\alpha(t)}\alpha'(t)\,ds.$$

Next we note that for $x \in M_n(\mathbb{K})$ we have

$$\begin{aligned}\int_0^1 e^{-s\,\mathrm{ad}\,x}\,ds &= \int_0^1\sum_{k=0}^\infty\frac{(-\,\mathrm{ad}\,x)^k}{k!}s^k\,ds = \sum_{k=0}^\infty(-\,\mathrm{ad}\,x)^k\int_0^1\frac{s^k}{k!}\,ds\\ &= \sum_{k=0}^\infty\frac{(-\,\mathrm{ad}\,x)^k}{(k+1)!} = \Phi(\mathrm{ad}\,x).\end{aligned}$$

We thus obtain for $\alpha(t) = x + ty$ the relation

$$\gamma(0,1) = \exp(-x)\mathtt{d}\exp(x)y = \int_0^1 e^{-s\,\mathrm{ad}\,x}y\,ds = \Phi(\mathrm{ad}\,x)y.$$ ∎

Lemma II.4.3. *For*

$$\Phi(z) = \frac{1-e^{-z}}{z} := \sum_{k=1}^\infty(-1)^{k-1}\frac{z^{k-1}}{k!}, \qquad z \in \mathbb{C}$$

and

$$\Psi(z) = \frac{z\log z}{z-1} := z\sum_{k=1}^\infty\frac{(-1)^{k-1}}{k}(z-1)^{k-1} = z\sum_{k=0}^\infty\frac{(-1)^k}{k+1}(z-1)^k \quad \textit{for } |z-1|<1$$

we have

$$\Psi(e^z)\Phi(z) = 1 \quad \textit{for} \quad z \in \mathbb{C}, |z| < \log 2.$$

Proof. If $|z| < \log 2$, then $|e^z - 1| < 1$ and we obtain from $\log(e^z) = z$:

$$\Psi(e^z)\Phi(z) = \frac{e^z z}{e^z - 1} \frac{1 - e^{-z}}{z} = 1. \qquad \blacksquare$$

In view of the Composition Formula (1.1) (Proposition II.1.6), the same identity as in Lemma II.4.3 holds if we insert linear maps $L \in \mathrm{End}(M_n(\mathbb{K}))$ with $\|L\| < \log 2$ into the power series Φ and Ψ:

$$\Psi(\exp L)\Phi(L) = (\Psi \circ \exp)(L)\Phi(L) = ((\Psi \circ \exp) \cdot \Phi)(L) = \mathrm{id}_{M_n(\mathbb{K})}.$$

Here we use that $\|L\| < \log 2$ implies that all expressions are defined and in particular that $\|\exp L - \mathbf{1}\| < 1$, as a consequence of the estimate

$$\|\exp L - \mathbf{1}\| \leq e^{\|L\|} - 1. \tag{4.1}$$

The derivation of the BCH formula follows a similar scheme as the proof of Proposition II.4.2. Here we consider $x, y \in V_o := B(0, \log\sqrt{2})$. For $\|x\|, \|y\| < r$ the estimate (4.1) leads to

$$\begin{aligned}
&\|\exp x \exp y - \mathbf{1}\| = \|(\exp x - \mathbf{1})(\exp y - \mathbf{1}) + (\exp y - \mathbf{1}) + (\exp x - \mathbf{1})\| \\
&\quad \leq \|\exp x - \mathbf{1}\| \cdot \|\exp y - \mathbf{1}\| + \|\exp y - \mathbf{1}\| + \|\exp x - \mathbf{1}\| \\
&\quad < (e^r - 1)^2 + 2(e^r - 1) = e^{2r} - 1.
\end{aligned}$$

For $r < \log\sqrt{2} = \frac{1}{2}\log 2$ and $|t| \leq 1$ we obtain in particular

$$\|\exp x \exp ty - \mathbf{1}\| < e^{\log 2} - 1 = 1.$$

Therefore $\exp x \exp ty$ lies for $|t| \leq 1$ in the domain of the logarithm function (Lemma II.3.1). We therefore define for $t \in [-1, 1]$:

$$F(t) = \log(\exp x \exp ty).$$

To estimate the norm of $F(t)$, we note that for $g := \exp x \exp ty$, $|t| \leq 1$, and $\|x\|, \|y\| < r$ we have

$$\|\log g\| \leq \sum_{k=1}^{\infty} \frac{\|g - 1\|^k}{k} = -\log(1 - \|g - \mathbf{1}\|) < -\log(1 - (e^{2r} - 1)) = -\log(2 - e^{2r}).$$

For $r := \frac{1}{2}\log(2 - \frac{\sqrt{2}}{2}) < \frac{\log 2}{2} = \log\sqrt{2}$ this leads for $\|x\|, \|y\| < r$ to

$$\|F(t)\| < -\log(2 - e^{2r}) = \log(\tfrac{2}{\sqrt{2}}) = \log(\sqrt{2}). \tag{4.2}$$

Next we calculate $F'(t)$ with the goal to obtain the BCH formula as $F(1) = F(0) + \int_0^1 F'(t)\,dt$. For the derivative of the curve $t \mapsto \exp F(t)$ we get with Proposition II.4.2:

$$\begin{aligned}(\mathtt{d}\exp)\big(F(t)\big)F'(t) &= \frac{d}{dt}\exp(F(t)) = \frac{d}{dt}\exp x \exp ty \\ &= (\exp x \exp ty)y = (\exp F(t))y.\end{aligned}$$

Using Proposition II.4.2 again, we obtain

$$y = \big(\exp F(t)\big)^{-1}(\mathtt{d}\exp)\big(F(t)\big)F'(t) = \frac{\mathbf{1} - e^{-\operatorname{ad} F(t)}}{\operatorname{ad} F(t)}F'(t). \tag{4.3}$$

The next step is to rewrite (4.3) with the function Φ from Lemma II.4.3 as

$$\Phi\big(\operatorname{ad} F(t)\big)F'(t) = y.$$

We claim that $\|\operatorname{ad}(F(t))\| < \log 2$. From $\|ab - ba\| \leq 2\|a\|\,\|b\|$ we derive

$$\|\operatorname{ad} a\| \leq 2\|a\| \quad \text{for} \quad a \in M_n(\mathbb{K}).$$

Therefore

$$\|\operatorname{ad} F(t)\| \leq 2\|F(t)\| < 2\log(\sqrt{2}) = \log 2,$$

so that the discussion above and (4.3) lead to

$$F'(t) = \Psi\big(\exp(\operatorname{ad} F(t))\big)y. \tag{4.4}$$

Proposition II.4.4. *For $x, y \in M_n(\mathbb{K})$ with $\|x\|, \|y\| < \frac{1}{2}\log(2 - \frac{\sqrt{2}}{2})$ we have*

$$\log(\exp x \exp y) = x + \int_0^1 \Psi\big(\exp(\operatorname{ad} x)\exp(t \operatorname{ad} y)\big)y\,dt,$$

with Ψ as in Lemma II.4.3.

Proof. With (4.4) and the preceding remarks we get

$$\begin{aligned}F'(t) &= \Psi\big(\exp(\operatorname{ad} F(t))\big)y \\ &= \Psi\big(\operatorname{Ad}(\exp F(t))\big)y = \Psi\big(\operatorname{Ad}(\exp x \exp ty)\big)y \\ &= \Psi\big(\operatorname{Ad}(\exp x)\operatorname{Ad}(\exp ty)\big)y = \Psi\big(\exp(\operatorname{ad} x)\exp(\operatorname{ad} ty)\big)y.\end{aligned}$$

Moreover, we have

$$F(0) = \log(\exp x) = x.$$

By integration we therefore obtain the formula. ∎

Proposition II.4.5. *For $x, y \in M_n(\mathbb{K})$ and $\|x\|, \|y\| < \frac{1}{2}\log(2 - \frac{\sqrt{2}}{2})$ we have*

$$\begin{aligned} x * y &:= \log(\exp x \exp y) \\ &= x + \\ &\sum_{\substack{k,m\geq 0 \\ p_i+q_i>0}} \frac{(-1)^k}{(k+1)(q_1+\ldots+q_k+1)} \frac{(\operatorname{ad} x)^{p_1}(\operatorname{ad} y)^{q_1}\ldots(\operatorname{ad} x)^{p_k}(\operatorname{ad} y)^{q_k}(\operatorname{ad} x)^m}{p_1!q_1!\ldots p_k!q_k!m!} y. \end{aligned}$$

Proof. We only have to rewrite the expression in Proposition II.4.4:

$$\begin{aligned} &\int_0^1 \Psi\big(\exp(\operatorname{ad} x)\exp(\operatorname{ad} ty)\big) y\, dt \\ &= \int_0^1 \sum_{k=0}^{\infty} \frac{(-1)^k\big(\exp(\operatorname{ad} x)\exp(\operatorname{ad} ty) - \operatorname{id}\big)^k}{(k+1)} \big(\exp(\operatorname{ad} x)\exp(\operatorname{ad} ty)\big) y\, dt \\ &= \int_0^1 \sum_{\substack{k\geq 0 \\ p_i+q_i>0}} \frac{(-1)^k}{(k+1)} \frac{(\operatorname{ad} x)^{p_1}(\operatorname{ad} ty)^{q_1}\ldots(\operatorname{ad} x)^{p_k}(\operatorname{ad} ty)^{q_k}}{p_1!q_1!\ldots p_k!q_k!} \exp(\operatorname{ad} x) y\, dt \\ &= \sum_{\substack{k,m\geq 0 \\ p_i+q_i>0}} \frac{(-1)^k}{(k+1)} \frac{(\operatorname{ad} x)^{p_1}(\operatorname{ad} y)^{q_1}\ldots(\operatorname{ad} x)^{p_k}(\operatorname{ad} y)^{q_k}(\operatorname{ad} x)^m}{p_1!q_1!\ldots p_k!q_k!m!} y \int_0^1 t^{q_1+\ldots+q_k}\, dt \\ &= \sum_{\substack{k,m\geq 0 \\ p_i+q_i>0}} \frac{(-1)^k(\operatorname{ad} x)^{p_1}(\operatorname{ad} y)^{q_1}\ldots(\operatorname{ad} x)^{p_k}(\operatorname{ad} y)^{q_k}(\operatorname{ad} x)^m.y}{(k+1)(q_1+\ldots+q_k+1)p_1!q_1!\ldots p_k!q_k!m!}. \end{aligned}$$

■

The power series in Proposition II.4.5 is called the *Dynkin Series*. We observe that it does not depend on the size n of the matrices we consider. For practical purposes it often suffices to know the first terms of the Dynkin series:

Corollary II.4.6. *Let $x, y \in M_n(\mathbb{K})$ and $\|x\|, \|y\| < \frac{1}{2}\log(2 - \frac{\sqrt{2}}{2})$. Then we have*

$$x * y = x + y + \frac{1}{2}[x, y] + \frac{1}{12}[x, [x, y]] + \frac{1}{12}[y, [y, x]] + \ldots$$

Proof. One has to collect the summands in Proposition II.4.5 corresponding to $p_1 + q_1 + \ldots + p_k + q_k + m \leq 2$. ■

Product and Commutator Formula

We have seen in Lemma II.1.1 that the exponential image of a sum $x + y$ can be computed easily if x and y commute. In this case we also have for the commutator $[x, y] := xy - yx = 0$ the formula $\exp[x, y] = \mathbf{1}$. The following proposition gives a formula for $\exp(x + y)$ and $\exp([x, y])$ in the general case.

If g, h are elements of a group G, then $(g,h) := ghg^{-1}h^{-1}$ is called their *commutator*. On the other hand we call for two matrices $A, B \in M_n(\mathbb{K})$ the expression

$$[A,B] := AB - BA$$

their *commutator bracket*.

Proposition II.4.7. *For $x, y \in M_d(\mathbb{K})$ the following assertions hold:*

(i) (Trotter Product Formula)

$$\lim_{k\to\infty} \left(e^{\frac{1}{k}x} e^{\frac{1}{k}y}\right)^k = e^{x+y}.$$

(ii) (Commutator Formula)

$$\lim_{k\to\infty} \left(e^{\frac{1}{k}x} e^{\frac{1}{k}y} e^{-\frac{1}{k}x} e^{-\frac{1}{k}y}\right)^{k^2} = e^{xy-yx}.$$

Proof. (i) From Corollary II.4.6 we derive

$$\lim_{k\to\infty} k \cdot \left(\frac{x}{k} * \frac{y}{k}\right) = x + y. \tag{4.5}$$

Applying the exponential function, we obtain (i).

(ii) We consider the function

$$\gamma(t) := tx * ty * (-tx) * (-ty),$$

which is defined for sufficiently small $t \in \mathbb{R}$ and smooth. In view of

$$\exp(x * y * (-x)) = \exp x \exp y \exp(-x) = \exp\big(\mathrm{Ad}(\exp x)y\big) = \exp(e^{\mathrm{ad}\, x} y)$$

(Lemma II.4.1), we have

$$x * y * (-x) = e^{\mathrm{ad}\, x} y, \tag{4.6}$$

and therefore the Chain Rule for Taylor Polynomials yields

$$\begin{aligned}\gamma(t) &= tx * ty * (-tx) * (-ty) = e^{t\,\mathrm{ad}\, x} ty * (-ty) \\ &= (ty + t^2[x,y] + \frac{t^3}{2}[x,[x,y]] + \ldots) * (-ty) \\ &= ty + t^2[x,y] - ty + [ty, -ty] + t^2 r(t) = t^2[x,y] + t^2 r(t),\end{aligned}$$

where $\lim_{t\to 0} r(t) = 0$. We therefore have

$$\gamma(0) = \gamma'(0) = 0 \quad \text{and} \quad \frac{\gamma''(0)}{2} = [x,y].$$

This leads to

$$\lim_{k\to\infty} k^2 \cdot \big(\frac{1}{k}x\big) * \big(\frac{1}{k}y\big) * \big(-\frac{1}{k}x\big) * \big(-\frac{1}{k}y\big) = \frac{\gamma''(0)}{2} = [x,y]. \tag{4.7}$$

Applying $\exp$ leads to the Commutator Formula. ∎

Exercises for Section II.4.

Exercise II.4.1. If $(V, \cdot)$ is an associative algebra, then we have $\mathrm{Aut}(V, \cdot) \subseteq \mathrm{Aut}(V, [\cdot, \cdot])$. ■

Exercise II.4.2. (a) $\mathrm{Ad} : \mathrm{GL}_n(\mathbb{K}) \to \mathrm{Aut}\big(M_n(\mathbb{K})\big)$ is a group homomorphism.

(b) For each Lie algebra $\mathfrak{g}$ the map $\mathrm{ad}\colon \mathfrak{g} \to \mathrm{der}(\mathfrak{g})$, $\mathrm{ad}\, x(y) := [x, y]$ is a homomorphism of Lie algebras. ■

Exercise II.4.3. Let V be a finite-dimensional vector space, $F \subseteq V$ a subspace and $\gamma : [0, T] \to V$ a continuous curve with $\gamma([0, T]) \subseteq F$. Then for all $t \in [0, T]$:

$$I_t := \int_0^t \gamma(\tau)\, d\tau \in F.$$

Hint: Use the linearity of the integral to see that every linear functional vanishing on F vanishes on I_t. Why does this imply the assertion? ■

Exercise II.4.4. On each finite-dimensional Lie algebra $\mathfrak{g}$ there exists a norm with

$$\|[x, y]\| \leq \|x\| \|y\| \qquad \forall x, y \in \mathfrak{g},$$

i.e., $\|\mathrm{ad}\, x\| \leq \|x\|$. Hint: If $\|\cdot\|_1$ is any norm on $\mathfrak{g}$, then the continuity of the bracket implies that $\|[x, y]\|_1 \leq C\|x\|_1\|y\|_1$. Modify $\|\cdot\|_1$ to obtain $\|\cdot\|$. ■

Exercise II.4.5. Let $\mathfrak{g}$ be a Lie algebra with a norm as in Exercise II.4.4. Then for $\|x\| + \|y\| < \ln 2$ the Dynkin series

$$x * y = x + \sum_{\substack{k,m \geq 0 \\ p_i + q_i > 0}} \frac{(-1)^k}{(k+1)(q_1 + \ldots + q_k + 1)} \frac{(\mathrm{ad}\, x)^{p_1} (\mathrm{ad}\, y)^{q_1} \ldots (\mathrm{ad}\, x)^{p_k} (\mathrm{ad}\, y)^{q_k} (\mathrm{ad}\, x)^m}{p_1! q_1! \ldots p_k! q_k! m!} y$$

converges absolutely. Hint: Show that

$$\|x * y\| \leq \|x\| + e^{\|x\|}\|y\| \sum_{k>0} \frac{1}{k+1} (e^{\|x\|+\|y\|} - 1)^k. \qquad ■$$

Exercise II.4.6. Prove Corollary II.4.6. ■

Exercise II.4.7. Let V and W be vector spaces and $q : V \times V \to W$ a skew-symmetric bilinear map. Then

$$[(v, w), (v', w')] := \big(0, q(v, v')\big)$$

is a Lie bracket on $\mathfrak{g} := V \times W$. For $x, y, z \in \mathfrak{g}$ we have $\big[x, [y, z]\big] = 0$. ■

Exercise II.4.8. Let $\mathfrak{g}$ be a Lie algebra with $\big[x,[y,z]\big] = 0$ for $x, y, z \in \mathfrak{g}$. Then

$$x * y := x + y + \frac{1}{2}[x,y]$$

defines a group structure on $\mathfrak{g}$. An example for such a Lie algebra is the three-dimensional *Heisenberg algebra*

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & x & y \\ 0 & 0 & z \\ 0 & 0 & 0 \end{pmatrix} : x, y, z \in \mathbb{K} \right\}. \qquad \blacksquare$$