5. Matrix Lie Groups

(5.1) Definition. *The* **general linear group** *in dimension* n *is defined as the set of all invertible real* $(n \times n)$*-matrices, which can be written as*

$$\mathrm{GL}(n, \mathbb{R}) \ := \ \{g \in \mathbb{R}^{n \times n} \mid \det(g) \neq 0\}\,.$$

This is a group in the algebraic sense. (Namely, it is closed under the associative operation of matrix multiplication, contains the identity matrix $\mathbf{1}$ as a neutral element and contains with each element g also the inverse matrix g^{-1}.) Also, $\mathrm{GL}(n, \mathbb{R})$ is an open subset and hence an open submanifold of $\mathbb{R}^n$. Moreover, the group structure and the manifold structure are compatible in the sense that the algebraic operations (matrix multiplication and matrix inversion) are C^∞-mappings (even polynomial and rational functions) of the matrix coefficients. In mathematics, situations in which different structures (order structures, algebraic structures, topological structures, differentiable structures) occur and are compatible with each other give rise to particularly rich theories. This applies in particular to Lie groups.

(5.2) Definition. *A* **matrix Lie group** *is a closed subgroup of the general linear group* $\mathrm{GL}(n, \mathbb{R})$ *(where "closed" refers to the subspace topology which* $\mathrm{GL}(n, \mathbb{R})$ *inherits from* $\mathbb{R}^{n \times n}$*).*

(5.3) Examples. (a) The **special linear group** in dimension n is defined as

$$\mathrm{SL}(n, \mathbb{R}) \ := \ \{g \in \mathbb{R}^{n \times n} \mid \det(g) = 1\}.$$

(b) The **orthogonal group** in dimension n is defined as

$$\mathrm{O}(n, \mathbb{R}) \ := \ \{g \in \mathbb{R}^{n \times n} \mid g^T g = \mathbf{1}\}\,.$$

(c) The intersection of two matrix Lie groups is again a matrix Lie group. For example, the **special orthogonal group** is defined as

$$\mathrm{SO}(n) \ := \ \mathrm{O}(n) \cap \mathrm{SL}(n).$$

(d) The **affine group** on $\mathbb{R}^n$ is the group

$$\mathrm{Aff}(\mathbb{R}^n) \ := \ \{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{(n+1)\times(n+1)} \mid A \in \mathrm{GL}(n, \mathbb{R}),\, b \in \mathbb{R}^n\}\,.$$

It will turn out that each matrix Lie group is a submanifold of $\mathbb{R}^{n \times n}$. In the above examples this can be easily verified in each case, but we want to give a general proof. A key ingredient to this proof will be the generalization of the exponential function and the logarithm function to matrices.

(5.4) Definition. *Let $A \in \mathbb{C}^{n\times n}$. We define*

$$\exp(A) \;:=\; \sum_{n=0}^{\infty} \frac{1}{n!}A^n \;=\; \mathbf{1} + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots .$$

This is a valid definition because the series by which $\exp(A)$ is defined converges for all $A \in \mathbb{C}^{n\times n}$. This can be seen by choosing a norm $\|\cdot\|$ on $\mathbb{C}^{n\times n}$ satisfying $\|\mathbf{1}\| = 1$ and $\|AB\| \le \|A\|\,\|B\|$ for all $A, B \in \mathbb{C}^{n\times n}$; then

$$\sum_{n=0}^{\infty} \left\| \frac{1}{n!}A^n \right\| \;=\; \sum_{n=0}^{\infty} \frac{1}{n!}\|A^n\| \;\le\; \sum_{n=0}^{\infty} \frac{1}{n!}\|A\|^n \;=\; e^{\|A\|}.$$

The properties of the exponential function for matrices are listed in the following theorem.

(5.5) Theorem. *Let $X, Y, T \in \mathbb{C}^{n\times n}$.*

(a) *We have* $\exp(\mathbf{0}) = \mathbf{1}$.
(b) *If $XY = YX$ then* $\exp(X)\exp(Y) = \exp(X+Y)$.
(c) *The matrix* $\exp(X)$ *is invertible with* $\exp(X)^{-1} = \exp(-X)$.
(d) *If T is invertible then* $\exp(T^{-1}XT) = T^{-1}\exp(X)T$.
(e) *We have* $\det\bigl(\exp(X)\bigr) = e^{\mathrm{tr}(X)}$.
(f) *For fixed X, the mapping $t \mapsto \exp(tX)$ is differentiable with*

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp(tX) \;=\; X\exp(tX) \;=\; \exp(tX)X.$$

(g) *If (X_n) is any sequence of matrices with $X_n \to X$, then*

$$\exp(X) \;=\; \lim_{n\to\infty}\left(\mathbf{1} + \frac{1}{n}X_n\right)^n.$$

(h) *We have*

$$\exp(X+Y) \;=\; \lim_{n\to\infty}\left(\exp\left(\frac{X}{n}\right)\exp\left(\frac{Y}{n}\right)\right)^n.$$

(i) *Writing $[X,Y] := XY - YX$, we have*

$$\exp([X,Y]) \;=\; \lim_{n\to\infty}\left(\exp\left(\frac{X}{n}\right)\exp\left(\frac{Y}{n}\right)\exp\left(-\frac{X}{n}\right)\exp\left(-\frac{Y}{n}\right)\right)^{n^2}.$$

Proof. Part (a) is trivial. For (b) we note that, since X and Y commute, the binomial formula

$$(X+Y)^n \;=\; \sum_{r+s=n} \frac{n!}{r!\,s!}X^rY^s$$

applies and yields

$$\begin{aligned}\exp(X+Y) &= \sum_{n=0}^{\infty}\frac{(X+Y)^n}{n!} = \sum_{n=0}^{\infty}\left(\sum_{r+s=n}\frac{1}{r!s!}X^rY^s\right)\\ &= \left(\sum_{r=0}^{\infty}\frac{X^r}{r!}\right)\left(\sum_{s=0}^{\infty}\frac{X^s}{s!}\right) = \exp(X)\exp(Y).\end{aligned}$$

Here the change in the order of summation is allowed due to the absolute convergence of the series. Using first (b) and then (a) we obtain $\exp(X)\exp(-X) = \exp\big(X+(-X)\big) = \exp(\mathbf{0}) = \mathbf{1}$, which yields (c). For (d) we note that

$$(T^{-1}XT)^n = T^{-1}XTT^{-1}XTT^{-1}XT\cdots T^{-1}XT = T^{-1}X^nT$$

for all $n \in \mathbb{N}_0$ and hence

$$\exp(T^{-1}XT) = \sum_{n=0}^{\infty}\frac{(T^{-1}XT)^n}{n!} = \sum_{n=0}^{\infty}\frac{T^{-1}X^nT}{n!} = T^{-1}\left(\sum_{n=0}^{\infty}\frac{X^n}{n!}\right)T = T^{-1}\exp(X)T.$$

To prove (e), we note that there is a transformation matrix T such that $T^{-1}XT$ is an upper triangular matrix whose diagonal elements are the eigenvalues $\lambda_1,\ldots,\lambda_n$ of X. Then

$$\begin{aligned}\det\big(\exp(X)\big) &= \det\big(T^{-1}\exp(X)T\big) = \det\big(\exp(T^{-1}XT)\big) = \det\left(\exp\begin{bmatrix}\lambda_1 & & \star\\ & \ddots & \\ 0 & & \lambda_n\end{bmatrix}\right)\\ &= \det\begin{bmatrix}e^{\lambda_1} & & \star\\ & \ddots & \\ 0 & & e^{\lambda_n}\end{bmatrix} = e^{\lambda_1}\cdots e^{\lambda_n} = e^{\lambda_1+\cdots+\lambda_n} = e^{\mathrm{tr}(X)}.\end{aligned}$$

Next, we note that taking derivatives termwise in the equation $\exp(tX) = \sum_{n=0}^{\infty}(t^n/n!)X^n$ results in

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp(tX) = \sum_{n=1}^{\infty}\frac{nt^{n-1}}{n!}X^n = X\sum_{n=1}^{\infty}\frac{t^{n-1}}{(n-1)!}X^{n-1} = X\exp(tX).$$

Since the resulting series is again uniformly convergent, taking the derivative termwise is justified, and (f) is established. To prove (g) we note that, due to the binomial formula, we can write

$$\left(\mathbf{1}+\frac{X_n}{n}\right)^n = \sum_{k=0}^{n}\binom{n}{k}\frac{X_n^k}{n^k} = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}\frac{X_n^k}{n^k} = \sum_{k=0}^{n}\frac{X_n^k}{k!}\frac{n-1}{n}\cdot\frac{n-2}{n}\cdots\frac{n-(k-1)}{n}$$

so that $(\mathbf{1}+X_n/n)^n = \sum_{k=0}^{\infty}F_k(n)$ where $F_k(n) := 0$ if $k > n$ and

$$F_k(n) := \frac{X_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right) \quad \text{if } k \le n.$$

The sequence (X_n) is bounded, say $\|X_n\| \le C$ for all n. Then $\|F_k(n)\| \le C^k/k!$ for all k uniformly in n, so that in the following sequence of equations the order of forming the limit and taking the series may be exchanged:

$$\lim_{n\to\infty}\left(\mathbf{1}+\frac{X_n}{n}\right)^n \;=\; \lim_{n\to\infty}\sum_{k=0}^{\infty}F_k(n) \;=\; \sum_{k=0}^{\infty}\lim_{n\to\infty}F_k(n) \;=\; \sum_{k=0}^{\infty}\frac{X^k}{k!} \;=\; \exp(X).$$

The remaining claims (h) and (i) follow easily from (g). To prove (h) we write $\exp(X/n) = \mathbf{1} + (X/n) + \mathrm{O}(1/n^2)$ (and similarly with Y instead of X) to find

$$\left(\exp\left(\frac{X}{n}\right)\exp\left(\frac{Y}{n}\right)\right)^n \;=\; \left(\mathbf{1}+\frac{X+Y}{n}+\mathrm{O}(1/n^2)\right)^n \;\to\; \exp(X+Y) \quad \text{as } n\to\infty.$$

To prove (i) we write $\exp(X/n) = \mathbf{1} + (X/n) + \big(X^2/(2n^2)\big) + \mathrm{O}(1/n^3)$ and then, after multiplying out,

$$\exp\left(\frac{X}{n}\right)\exp\left(\frac{Y}{n}\right)\exp\left(-\frac{X}{n}\right)\exp\left(-\frac{Y}{n}\right) \;=\; \mathbf{1}+\frac{XY-YX}{n^2}+\mathrm{O}(1/n^3),$$

from which (i) easily follows. ■

(5.6) Definition. *Let $A \in \mathbb{C}^{n\times n}$ such that $\|A-\mathbf{1}\| < 1$. Then we define*

$$\log(A) \;:=\; \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}(A-\mathbf{1})^n.$$

Somewhat more conveniently, we write $A = \mathbf{1} + X$ where $\|X\| < 1$ and then have

$$\log(\mathbf{1}+X) \;=\; \sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}X^n \;=\; X-\frac{1}{2}X^2+\frac{1}{3}X^3-\frac{1}{4}X^4\pm\cdots.$$

This definition is possible because the series converges absolutely. In fact,

$$\sum_{n=1}^{\infty}\left\|\frac{(-1)^{n+1}}{n}X^n\right\| \;=\; \sum_{n=1}^{\infty}\frac{1}{n}\|X^n\| \;\le\; \sum_{n=1}^{\infty}\frac{1}{n}\|X\|^n \;=\; -\ln(1-\|X\|)\,.$$

Note that whereas $\exp(X)$ is defined for all matrices X, the logarithm is only defined for matrices suffiently close to the identity matrix.

We now show that exp and log are locally inverses of each other.

(5.7) Theorem. (a) *If* $\|\exp(X)-\mathbf{1}\| < 1$ *then* $\log\big(\exp(X)\big) = X$.
(b) *If* $\|A-\mathbf{1}\| < 1$, $\|B-\mathbf{1}\| < 1$ *and* $\|AB-\mathbf{1}\| < 1$ *(so that* $\log(A)$, $\log(B)$ *and* $\log(AB)$ *are all defined) and if* $AB = BA$ *then* $\log(AB) = \log(A) + \log(B)$.

Proof. (a) Letting $\Phi(X) := \sum_{n=1}^{\infty} X^n/n!$, we have $\|\Phi(X)\| < 1$ and hence

$$\log\bigl(\exp(X)\bigr) \;=\; \log\bigl(\mathbf{1}+\Phi(X)\bigr) \;=\; \Phi(X) - \frac{1}{2}\Phi(X)^2 + \frac{1}{3}\Phi(X)^3 \mp \cdots .$$

The series on the right-hand side is absolutely convergent; hence we can rearrange terms (sorting by powers of X) and write this series as $\log\bigl(\exp(X)\bigr) = X + a_2X^2 + a_3X^3 + \cdots$ with certain coefficients a_k. But these coefficients must all be zero, as becomes clear by plugging in $X = x\mathbf{1}$ and comparing with the known result that $\log(\exp x) = x$ for real numbers x.

(b) Write $X := \log(A)$ and $Y := \log(B)$ so that $A = \exp(X)$ and $B = \exp(Y)$. Since the matrices A and B commute, so do the matrices $X = \log\bigl(\mathbf{1} + (A-\mathbf{1})\bigr) = \sum_{n=1}^{\infty}\bigl((-1)^{n+1}/n\bigr)\,(A-\mathbf{1})^n$ and $Y = \log\bigl(\mathbf{1}+(B-\mathbf{1})\bigr) = \sum_{n=1}^{\infty}\bigl((-1)^{n+1}/n\bigr)\,(B-\mathbf{1})^n$. Therefore, $\exp(X+Y) = \exp(X)\exp(Y) = AB$. Consequently, part (a) yields $\log(AB) = \log\bigl(\exp(X+Y)\bigr) = X+Y = \log(A)+\log(B)$, as claimed. ■

(5.8) Definition. *Let $G \subseteq \mathbb{R}^{n\times n}$ be a matrix Lie group. An* **infinitesimal group element** *for G is a matrix $X \in \mathbb{R}^{n\times n}$ such that there exist sequences $(g_1, g_2, g_3, \ldots)$ in G and $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots)$ in $\mathbb{R}^+$ such that $g_n \to \mathbf{1}$ and $(g_n - \mathbf{1})/\varepsilon_n \to X$ as $n \to \infty$. (Note that then necessarily $\varepsilon_n \to 0$ as $n \to \infty$ if $g_n \neq \mathbf{1}$ for all n.) The set of all infinitesimal group elements for G will be denoted by $L(G)$.*

Geometrically, an infinitesimal group element can be interpreted as a (scalar multiple of a) direction in which a sequence of group elements can approach the identity element **1**.

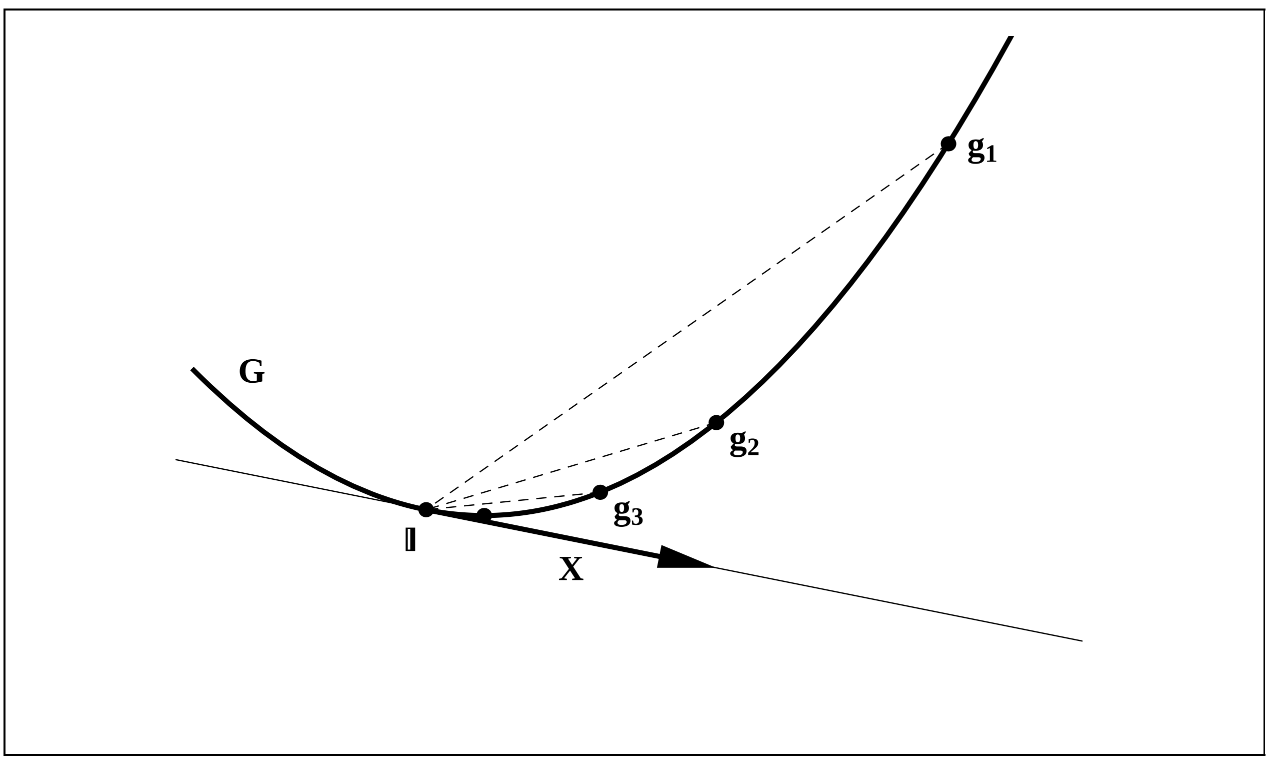

Fig. 5.1: Visualization of an infinitesimal element X for a group G.

Let us motivate this definition somewhat. Our goal will be to show that G is a submanifold of $\mathbb{R}^{n\times n}$ and that the tangent space $T_{\mathbf{1}}G$ of G at $\mathbf{1}$ consists exactly of the infinitesimal group elements of G. Assuming this for the moment, the connection is easily seen: If X is a tangent vector of G at $\mathbf{1}$, there is a smooth curve $t \mapsto \alpha(t) \in \mathbb{R}^{n\times n}$ such that $\alpha(t) \in G$ for all t and such that $\alpha(0) = \mathbf{1}$ and $\alpha'(0) = X$. In this case we can take $g_n := \alpha(1/n)$ and $\varepsilon_n := 1/n$ and then have

$$\lim_{n\to\infty} \frac{g_n - \mathbf{1}}{\varepsilon_n} \;=\; \lim_{n\to\infty} \frac{\alpha(1/n) - \alpha(0)}{1/n} \;=\; \alpha'(0).$$

What is more tricky is the other way round: Given an infinitesimal group element X, can we find a smooth function $t \mapsto \alpha(t)$ in G such that $\alpha(0) = \mathbf{1}$ and $\alpha'(0) = X$? The answer will be yes, and, in fact, the curve can be chosen to be $\alpha(t) := \exp(tX)$. This is the main result of the following theorem.

(5.9) Theorem. *Let G be a matrix Lie group and let X be an infinitesimal element for G. Then* $\exp(tX) \in G$ *for all* $t \in \mathbb{R}$.

Proof. Assume first that $\|X\| < \ln(2)$ for some submultiplicative norm $\|\cdot\|$; then $\|\exp(X) - \mathbf{1}\| = \|\sum_{n=1}^{\infty} X^n/n!\| \le \sum_{n=1}^{\infty} \|X\|^n/n! = \exp(\|X\|) - 1 < \exp\big(\ln(2)\big) - 1 = 1$. Since X is an infinitesimal element for G, there are sequences (g_n) in G and (ε_n) in $\mathbb{R}^+$ such that $(g_n - \mathbf{1})/\varepsilon_n \to X$. Let k_n be the integer closest to $1/\varepsilon_n$; then $k_n(g_n - \mathbf{1}) \to X$ as $n \to \infty$. We claim that if n is sufficiently large then $\|g_n^{k_n} - \mathbf{1}\| < 1$, so that $\log(g_n^{k_n})$ exists. To prove this claim, write $R_n := k_n(g_n - \mathbf{1}) - X$ so that $R_n \to 0$ as $n \to \infty$. Then using (5.5)(g) we see that

$$g_n^{k_n} - \mathbf{1} \;=\; \left(\mathbf{1} + \frac{X + R_n}{k_n}\right)^{k_n} - \mathbf{1} \;\to\; \exp(X) - \mathbf{1},$$

which implies that $\|g_n^{k_n} - \mathbf{1}\| < 1$ if n is sufficiently large. Then $\log(g_n^{k_n})$ exists, and

$$\log(g_n^{k_n}) \;=\; k_n \log(g_n) = k_n(g_n - \mathbf{1}) - k_n(g_n - \mathbf{1}) \left[\frac{g_n - \mathbf{1}}{2} - \frac{(g_n - \mathbf{1})^2}{3} \pm \cdots\right]$$

where the term in parentheses is easily seen to tend to zero as $n \to \infty$. Consequently, $\log(g_n^{k_n}) \to X$ and hence, due to the continuity of the exponential function, $g_n^{k_n} = \exp\log(g_n^{k_n}) \to \exp(X)$. Since $g_n^{k_n} \in G$ for all n this implies $\exp(X) \in \overline{G} = G$. This shows that $\exp(X) \in G$ whenever $X \in L(G)$ such that $\|X\| < \ln(2)$. But this special case immediately implies the general case. To see this we note first that if X is an infinitesimal group element for G then so is tX for all $t \in \mathbb{R}$; this follows immediately from the definition and will be explicitly stated in the next theorem. Now let $X \in L(G)$ and $t \in \mathbb{R}$ be arbitrary. Choose $m \in \mathbb{N}$ so large that $\|tX/m\| < \ln(2)$; then $\exp(tX/m) \in G$ by what we have shown so far. But then also $\exp(tX) = \exp(tX/m)^m \in G$, since G is multiplicatively closed. ∎

Applying the remark preceding Theorem (5.9) with $\alpha(t) := \exp(tX)$, we see that the converse of Theorem (5.9) also holds. Hence a matrix $X \in \mathbb{R}^{n\times n}$ is an infinitesimal group element for G if and only if $\exp(tX) \in G$ for all $t \in \mathbb{R}$.

(5.10) Theorem. *Let G be a matrix Lie group, and let $L(G)$ be the set of its infinitesimal elements. Then $L(G)$ is a vector space. Moreover, if $X, Y \in L(G)$ then also $[X,Y] := XY - YX \in L(G)$. (We call $[X,Y]$ the* **Lie bracket** *of X and Y, and we call $L(G)$ the* **Lie algebra** *of the group G.)*

Proof. If we choose the constant sequences (g_n) and (ε_n) with $g_n := \mathbf{1}$ and $\varepsilon_n := 1$ or all n, we have $(g_n - \mathbf{1})/\varepsilon_n \to \mathbf{0}$; hence $\mathbf{0} \in L(G)$. If $X \in L(G)$ then also $-X \in L(G)$ because $(g_n - \mathbf{1})/\varepsilon_n \to X$ implies

$$\frac{g_n^{-1} - \mathbf{1}}{\varepsilon_n} \; = \; -g_n^{-1}\,\frac{g_n - \mathbf{1}}{\varepsilon_n} \; \to \; -\mathbf{1}^{-1}X \; = \; -X \quad \text{as } n \to \infty.$$

Moreover, if $X \in L(G)$ and $\lambda > 0$ then $\lambda X \in L(G)$, because $(g_n - \mathbf{1})/\varepsilon_n \to X$ implies

$$\frac{g_n - \mathbf{1}}{\varepsilon_n/\lambda} \; = \; \lambda\,\frac{g_n - \mathbf{1}}{\varepsilon_n} \; \to \; \lambda X \quad \text{as } n \to \infty.$$

This shows that $L(G)$ is closed by multiplication with arbitrary scalars. To show that $L(G)$ is also closed under addition, we do not proceed directly from the definition of infinitesimal elements, but invoke Theorem (5.9). Given $X, Y \in L(G)$, we know from this theorem that $\exp(tX) \in G$ and $\exp(tY) \in G$ for all $t \in \mathbb{R}$. Hence $\alpha(t) := \exp(tX)\exp(tY)$ is a curve in G which obviously satisfies $\alpha(0) = \mathbf{1}$ and $\alpha'(0) = X + Y$ which, due to the remark preceding Theorem (5.9), shows that $X + Y \in L(G)$. Finally, if $X, Y \in L(G)$ then, for any fixed $s \in \mathbb{R}$, a curve in G is given by $\alpha_s(t) := \exp(sX)\exp(tY)\exp(-sX)$ which satisfies $\alpha_s(0) = \mathbf{1}$ and $\alpha_s'(0) = \exp(sX)Y\exp(-sX)$. Hence $\exp(sX)Y\exp(-sX)$ in $L(G)$ for all $s \in \mathbb{R}$, which implies that $\alpha(s) := \exp(sX)Y\exp(-sX)$ is a curve in $L(G)$. Since $L(G)$ is a vector space, the derivative $\alpha'(0) = XY - YX$ is then also an element of $L(G)$. ∎

(5.11) Theorem. *Let G be a matrix Lie group. Then there is a neighborhood of* $\mathbf{1}$ *in G which is mapped into $L(G)$ by the logarithm function.*

Proof. Assume that there is no such neighborhood. Then there is a sequence $(g_1, g_2, g_3, \ldots)$ in G with $g_n \to \mathbf{1}$ such that $\log(g_n) \notin L(G)$ for all $n \in \mathbb{N}$. Choose any inner product on the vector space $L(G)$ and consider the decomposition $\mathbb{R}^{n\times n} = L(G) \oplus L(G)^\perp$; then $\log(g_n) = X_n + Y_n$ with $X_n \in L(G)$ and $Y_n \in L(G)^\perp \setminus \{0\}$. Note that $g_n \to \mathbf{1}$ for $n \to \infty$ implies $X_n \to 0$ and $Y_n \to 0$. The sequence $Y_n/\|Y_n\|$ is bounded, hence possesses a convergent subsequence; without loss of generality, we may assume that $Y_n/\|Y_n\| \to Y$ for some Y. Then $Y \in L(G)^\perp$ and $\|Y\| = 1$, in particular $Y \neq 0$. For each n, the matrix $\gamma_n := \exp(-X_n)g_n = \exp(-X_n)\exp(X_n + Y_n)$ is an element of G. Now

$$(\star) \qquad \begin{aligned} \gamma_n \; &= \; \Big(\mathbf{1} - X_n + \frac{X_n^2}{2} \mp \cdots\Big)\Big(\mathbf{1} + X_n + Y_n + \frac{(X_n + Y_n)^2}{2} + \cdots\Big) \\ &= \; \mathbf{1} \, + \, Y_n \, + \, \text{higher-order terms.} \end{aligned}$$

It is important to notice that amongst the higher-order terms there are no pure powers of X_n, because those would have to show up in the expansion of $\exp(-X_n)\exp(X_n + Y_n)$

with $Y_n = 0$, which is simply $\exp(-X_n)\exp(X_n) = \mathbf{1}$. Using this fact, we conclude from $(\star)$ that

$$\lim_{n\to\infty} \frac{\gamma_n - \mathbf{1}}{\|Y_n\|} \;=\; \lim_{n\to\infty} \frac{Y_n}{\|Y_n\|} \;=\; Y,$$

which shows that Y is an infinitesimal element for G, contradicting the fact that $Y \in L(G)^\perp \setminus \{0\}$. Thus our assumption was false, and the theorem is established. ■

(5.12) Theorem (von Neumann 1929). *Let $G \subseteq \mathrm{GL}(n,\mathbb{R})$ be a matrix Lie group. Then G is a submanifold of $\mathbb{R}^{n\times n}$ with tangent space $T_{\mathbf{1}}G = L(G)$. Moreover, the tangent space T_gG at an arbitrary element $g \in G$ is given by $g \cdot T_{\mathbf{1}}G = \{gX \mid X \in T_{\mathbf{1}}G\}$.*

Proof. We showed that a neighborhood of $\mathbf{1}$ in G can be deformed into a neighborhood of $\mathbf{0}$ in the vector space $L(G)$ via the matrix logarithm. (Equivalently, the exponential function serves as a local parametrization around $\mathbf{1}$.) Since the left- and right-translations $X \mapsto gX$ and $X \mapsto Xg$ are analytic mappings, a local parametrization around an arbitrary element $g \in G$ is given by $X \mapsto g\exp(X)$ or $X \mapsto \exp(X)g$. This shows that G is a manifold. The rest has been already established. ■

The fact that exp and log are mutually inverse diffeomorphisms between a neighborhood of $\mathbf{0}$ in $L(G)$ and a neighborhood of $\mathbf{1}$ in G implies that if $X, Y \in L(G)$ are sufficiently close to $\mathbf{0}$ there is an element $Z = F(X,Y)$ such that $e^X e^Y = e^Z$. The analytic function F thus defined is given by the equation

$$\begin{aligned} F(X,Y) &= \log(e^X e^Y) \;=\; \log\Big(\Big(\mathbf{1} + X + \frac{X^2}{2} + \cdots\Big)\Big(\mathbf{1} + Y + \frac{Y^2}{2} + \cdots\Big)\Big) \\ &= \log\big(\mathbf{1} + \Phi(X,Y)\big) \quad \text{where} \quad \Phi(X,Y) = X + Y + \frac{X^2}{2} + XY + \frac{Y^2}{2} + \cdots \\ &= X + Y + \frac{XY}{2} - \frac{YX}{2} + \text{higher-order terms.} \end{aligned}$$

Writing $F(X,Y) = F_1(X,Y) + F_2(X,Y) + F_3(X,Y) + \cdots$, where $F_k(X,Y)$ is the sum of all occurring monomials which are homogeneous of degree k, we see that

$$\begin{aligned} F_1(X,Y) &= X + Y, \\ F_2(X,Y) &= (1/2)\,(XY - YX) \;=\; (1/2)\,[X,Y], \\ F_3(X,Y) &= (1/12)\,(X^2Y + XY^2 + YX^2 + Y^2X - 2XYX - 2YXY) \\ &= (1/12)\,\big([X,[X,Y]] + [Y,[Y,X]]\big), \end{aligned}$$

and so on. We see that the first three terms in the series $F(X,Y) = \sum_{n=1}^\infty F_n(X,Y)$ are all *Lie expressions* in X and Y, i.e., linear combinations of X and Y and elements obtained by forming Lie brackets between previously constructed elements. In general, we say that a polynomial in n non-commuting variables $X_1, \ldots, X_n$ is a *Lie expression* in these variables if it is a linear combination of these variables and Lie brackets obtained from them (in a finite number of steps). We now prove that all polynomials $F_n(X,Y)$ are Lie expressions in X and Y. This result dates back to Campbell (1897), Baker (1905) and Hausdorff (1906), who gave the first complete proof. The proof given here is due to Eichler (1968); this proof avoids analytical tools and shows that the following theorem is of a purely algebraic character. (Our presentation follows John Stillwell, *Naive Lie Theory*, Springer 2008.)

(5.13) Campbell-Baker-Hausdorff Theorem. *Let $F(A,B)$ be the formal power series in two non-commuting variables defined by $e^A e^B = e^{F(A,B)}$, and let $F_k(A,B)$ be the homogeneous part of order k of $F(A,B)$, for any $k \geq 1$. Then $F_n(A,B)$ is a Lie expression in A and B, for each $n \geq 1$.*

Proof. We first collect some basic properties of F. First, since $e^{rA}e^{sA} = e^{(r+s)A}$, we have

$$(1) \qquad F(rA, sA) = (r+s)A$$

for all $r,s \in \mathbb{R}$. Second, we have $e^{F(-A,-B)} = e^{-A}e^{-B} = (e^B e^A)^{-1} = (e^{F(B,A)})^{-1} = e^{-F(B,A)}$ and therefore

$$(2) \qquad F(B,A) \; = \; -F(-A,-B).$$

Finally, we make use of the associative law, which gives $(e^A e^B)e^C = e^A(e^B e^C)$. Now the left-hand side of this equation is $e^{F(A,B)}e^C = e^{F\big(F(A,B),C\big)}$ whereas the right-hand side is $e^A e^{F(B,C)} = e^{F\big(A,F(B,C)\big)}$. Since both sides are equal, we conclude that

$$(3) \qquad F\big(F(A,B),C\big) \; = \; F\big(A,F(B,C)\big)$$

which means that

$$(4) \qquad \sum_{i=1}^{\infty} F_i \left(\sum_{j=1}^{\infty} F_j(A,B),\, C \right) \; = \; \sum_{i=1}^{\infty} F_i \left(A,\, \sum_{j=1}^{\infty} F_j(B,C) \right) .$$

We now prove the claim using induction on n. The case $n=1$ is trivial because $F_1(A,B) = A+B$. Let $n \geq 2$. By induction hypothesis we assume that $F_k(\cdot,\cdot)$ is a Lie expression in its arguments for all $k < n$. Considering how in (4) the terms $F_i\big(F_1(A,B)+F_2(A,B)+\cdots,\, C\big)$ and $F_i\big(A,\, F_1(B,C)+F_2(B,C)+\cdots\big)$ are formed, we find, using the induction hypothesis, that only the terms for $(i,j) = (1,n)$ and for $(i,j) = (n,1)$ yield terms which are not known to be Lie expressions. On the left-hand of (4) side these are the terms $F_n(A,B)+C$ and $F_n(A+B,C)$, whereas on the right-hand side these are the terms $A + F_n(B,C)$ and $F_n(A,B+C)$. Since A and C are themselves Lie expressions, this shows that

$$(5) \qquad F_n(A,B) + F_n(A+B,C) \; \equiv \; F_n(B,C) + F_n(A,B+C)$$

where, in general, we write $p(A,B,C) \equiv q(A,B,C)$ if the polynomial expressions $p(A,B,C)$ and $q(A,B,C)$ differ only by a Lie expression. (Clearly, $\equiv$ is an equivalence relation on the set of all polynomials in three non-commuting variables.) Now we plug in $C := -B$ in (5) to obtain $F_n(A,B)+F_n(A+B,-B) \equiv F_n(A,0)+F_n(B,-B) = 0$ (where we used (1) in the last step) so that

$$F_n(A,B) \;\equiv\; -F_n(A+B,\,-B)\,. \tag{6}$$

Letting $A := -B$ in (5) and using (1) we find that $0 = F_n(-B,B)+F_n(0,C) \equiv F_n(-B,B+C)+F_n(B,C)$ and hence $F_n(B,C) \equiv -F_n(-B,B+C)$. Replacing B by A and then C by B in this formula yields

$$F_n(A,B) \;\equiv\; -F_n(-A,A+B). \tag{7}$$

Letting $C := -B/2$ in (5) we find that $F_n(A,B)+F_n(A+B,-B/2) \equiv F_n(A,B/2)+F_n(B,-B/2) = F_n(A,B/2)$, the last equality being a consequence of (1), and hence

$$F_n(A,B) \;\equiv\; F_n(A,B/2) - F_n(A+B,-B/2). \tag{8}$$

Letting $A := -B/2$ in (5), we find that $F_n(-B/2,B)+F_n(B/2,C) \equiv F_n(-B/2,B+C)+F_n(B,C)$; since $F_n(-B/2,B) = 0$ by (1), this can be rewritten in the form $F_n(B,C) \equiv F_n(B/2,C) - F_n(-B/2,B+C)$. Replacing B by A and then C by B in this formula, we find that

$$F_n(A,B) \;\equiv\; F_n(A/2,B) - F_n(-A/2,A+B). \tag{9}$$

Replacing A by $A/2$ in (8) yields

$$\begin{aligned} F_n(A/2,B) &\equiv F_n(A/2,B/2) - F_n\big((A/2)+B,-B/2\big) \\ &\overset{(6)}{\equiv} F_n(A/2,B/2) + F_n\big((A+B)/2,B/2\big) \\ &= \left(\frac{1}{2}\right)^n \big(F_n(A,B) + F_n(A+B,B)\big) \end{aligned} \tag{10}$$

where we used the fact that F_n is homogeneous of degree n in the last step. We now replace A by $-A/2$ and B by $A+B$ in (8) to obtain

$$\begin{aligned} F_n(-A/2,A+B) &\equiv F_n\big(-A/2,(A+B)/2\big) - F_n\big((A/2)+B,-(A+B)/2\big) \\ &\overset{(7)}{\equiv} -F_n(A/2,B/2) - F_n\big((A/2)+B,-(A+B)/2\big) \\ &\overset{(6)}{\equiv} -F_n(A/2,B/2) + F_n\big(B/2,(A+B)/2\big) \\ &= \left(\frac{1}{2}\right)^n \big(-F_n(A,B) + F_n(B,A+B)\big), \end{aligned} \tag{11}$$

where again we used that F_n is homogeneous of degree n in the last step. Plugging (10) and (11) into the right-hand side of (9) yields

$$F_n(A,B) \;=\; \left(\frac{1}{2}\right)^n \big(2F_n(A,B) + F_n(A+B,B) - F_n(B,A+B)\big). \tag{12}$$

Because of (2) we have $F_n(B, A+B) = -F_n\big(-(A+B), -B\big) = (-1)^{n+1} F_n(A+B, B)$; hence (12) becomes

$$\left(1 - \left(\frac{1}{2}\right)^{n-1}\right) F_n(A,B) \;\equiv\; \left(\frac{1}{2}\right)^{n} \left(1 + (-1)^n\right) F_n(A+B, B)\,. \tag{13}$$

If n is odd this implies $F_n(A,B) \equiv 0$, which means that $F_n(A,B)$ is a Lie expression; so in this case we are done. If n is even we replace A by $A - B$ in (13) to obtain

$$\left(1 - \left(\frac{1}{2}\right)^{n-1}\right) F_n(A-B, B) \;\equiv\; \left(\frac{1}{2}\right)^{n-1} F_n(A,B)\,. \tag{14}$$

Using (6), we have $F_n(A-B, B) \equiv -F_n(A, -B)$. Thus (14) becomes

$$-F_n(A, -B) \;\equiv\; \frac{(1/2)^{n-1}}{1 - (1/2)^{n-1}} F_n(A,B)\,. \tag{15}$$

Finally, replacing B by $-B$ in (15), we get

$$-F_n(A,B) \;\equiv\; \frac{(1/2)^{n-1}}{1-(1/2)^{n-1}} F_n(A,-B) \;\stackrel{(15)}{\equiv}\; -\left(\frac{(1/2)^{n-1}}{1-(1/2)^{n-1}}\right)^2 F_n(A,B) \tag{16}$$

and hence $F_n(A,B) \equiv 0$ unless $n = 2$, in which case the claim follows directly by observing that $F_2(A,B) = (1/2)\,[A,B]$. ∎