CHAPTER 1

Motivation: representations of Lie groups

Sophus Lie was a Norwegian mathematician who lived from 1842 to 1899. Essentially single-handedly he discovered two fundamental classes of objects in modern mathematics, which now bear his name: Lie groups and Lie algebras. More importantly, he built a bridge between them; this is remarkable, because Lie groups seem to be part of differential geometry (in today's language) while Lie algebras seem to be purely algebraic. In this chapter we will discuss a small part of Lie's discovery.

1.1 Homomorphisms of general linear groups

Typically, Lie groups are infinite groups whose elements are invertible matrices with real or complex entries. So they are subgroups of the *general linear group*

$$GL_n = \{g \in \mathrm{Mat}_n \mid \det(g) \neq 0\},$$

where $\mathrm{Mat}_n = \mathrm{Mat}_n(\mathbb{C})$ denotes the set of $n \times n$ complex matrices for some positive integer n. Lie was interested in such groups because they give the symmetries of differential equations, but they have since found many other applications in areas such as differential geometry and harmonic analysis.

One of the most important algebraic problems concerning Lie groups is to classify a suitable class of *matrix representations* of a given Lie group G, i.e. group homomorphisms $G \to GL_m$ for various m. For the purposes of motivation, we concentrate on the case where G is the full general linear group GL_n; thus the problem can be stated (vaguely) as follows.

Problem 1.1.1. Describe all group homomorphisms $\Phi : GL_n \to GL_m$.

By definition, such a homomorphism is a map $\Phi : GL_n \to \mathrm{Mat}_m$ such that:

$$\Phi(1_n) = 1_m, \tag{1.1.1}$$

where 1_n denotes the $n \times n$ identity matrix, and

$$\Phi(gh) = \Phi(g)\Phi(h) \quad \text{for all } g, h \in GL_n. \tag{1.1.2}$$

(The case $h = g^{-1}$ of (1.1.2), combined with (1.1.1), forces $\Phi(g)$ to be invertible.) Such a map Φ is a collection of m^2 functions $\Phi_{ij} : GL_n \to \mathbb{C}$, where $\Phi_{ij}(g)$ is the (i, j) entry of the matrix $\Phi(g)$. Each function Φ_{ij} is in effect a function of n^2 variables, the entries of the input matrix g (the given domain consists of just the invertible matrices, so the function may or may not be defined for those choices of variables that give a zero determinant). So (1.1.1) and (1.1.2) amount to a complicated system of functional equations. To frame Problem 1.1.1 rigorously, we would have to specify what kinds of function are allowed as solutions – for example, continuous, differentiable, rational, or polynomial – but we will leave this undetermined for now and see what happens in some examples.

Example 1.1.2. The determinant $\det : GL_n \to \mathbb{C}^\times$ is one such homomorphism, if we make the obvious identification of $\mathbb{C}^\times$ with GL_1. The determinant of a matrix is clearly a polynomial function of the entries. ■

Example 1.1.3. The transpose map $GL_n \to GL_n : g \mapsto g^{\mathbf{t}}$ is not an example because it is an anti-automorphism rather than an automorphism: $(gh)^{\mathbf{t}}$ equals $h^{\mathbf{t}}g^{\mathbf{t}}$ and doesn't usually equal $g^{\mathbf{t}}h^{\mathbf{t}}$. But this means that the map $GL_n \to GL_n : g \mapsto (g^{\mathbf{t}})^{-1}$ is an example. The entries of $(g^{\mathbf{t}})^{-1}$ are rational functions of the entries of g: they are quotients of various $(n-1)\times(n-1)$ minors and the determinant. Therefore these functions are not defined on matrices with zero determinant. ■

Example 1.1.4. A map that is easily seen to satisfy (1.1.1) and (1.1.2) is the 'duplication' map

$$GL_2 \to GL_4 : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix}.$$

To produce something less trivial-looking, we could replace either copy of $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ with its conjugate $X\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)X^{-1}$, for some fixed $X \in GL_2$, or indeed we could conjugate the whole output matrix by some fixed $Y \in GL_4$. This is a superficial change that we could account for by introducing a suitable equivalence relation into the statement of Problem 1.1.1. Note that in this example the entries of the output matrix are linear functions of the entries of the input matrix. ■

Example 1.1.5. More interesting is the map $\Psi : GL_2 \to GL_3$ defined by

$$\Psi\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a^2 & 2ab & b^2 \\ ac & ad+bc & bd \\ c^2 & 2cd & d^2 \end{pmatrix},$$

where the entries of the output are homogeneous polynomials of degree 2 in the entries of the input. It is clear that property (1.1.1) is satisfied. The proof of property (1.1.2) is as follows:

$$\Psi\begin{pmatrix} a & b \\ c & d \end{pmatrix}\Psi\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} a^2 & 2ab & b^2 \\ ac & ad+bc & bd \\ c^2 & 2cd & d^2 \end{pmatrix}\begin{pmatrix} e^2 & 2ef & f^2 \\ eg & eh+fg & fh \\ g^2 & 2gh & h^2 \end{pmatrix}$$

$$= \begin{pmatrix} (ae+bg)^2 & 2(ae+bg)(af+bh) & (af+bh)^2 \\ (ae+bg)(ce+dg) & \begin{matrix}(ae+bg)(cf+dh) \\ +(af+bh)(ce+dg)\end{matrix} & (af+bh)(cf+dh) \\ (ce+dg)^2 & 2(ce+dg)(cf+dh) & (cf+dh)^2 \end{pmatrix}$$

$$= \Psi\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix}\right).$$

At the moment this seems like an accident, and it is not clear how to find other such solutions of (1.1.1) and (1.1.2). ■

1.2 Multilinear algebra

The right context for explaining the above examples of homomorphisms, and for finding new examples, is the theory of multilinear algebra. If V is an n-dimensional vector space with chosen basis $v_1, \dots, v_n$ then the elements of GL_n correspond bijectively to invertible linear transformations of V: a matrix (a_{ij}) in GL_n corresponds to the unique linear map $\tau : V \to V$ such that

$$\tau(v_j) = \sum_{i=1}^{n} a_{ij} v_i \quad \text{for all } j. \tag{1.2.1}$$

If we have a way of constructing from V a new vector space W with basis $w_1, \dots, w_m$, and if this construction is sufficiently 'natural', then each linear transformation of V should induce a linear transformation of W and the resulting map $\Phi : GL_n \to GL_m$ of matrices should satisfy (1.1.1) and (1.1.2). This is one reason to be interested in Problem 1.1.1: the homomorphisms between general linear groups tell us something about natural constructions of vector spaces.

Example 1.2.1. A very important example of such a homomorphism occurs when W is the *dual space* V^*, consisting of all linear functions $f : V \to \mathbb{C}$. This is also n-dimensional: it has a basis $v_1^*, \dots, v_n^*$, where v_i^* is the unique linear function satisfying

$$v_i^*(v_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{1.2.2}$$

In other words, v_i^* is the function whose value on $a_1 v_1 + \cdots + a_n v_n \in V$ is the coefficient a_i. A general linear function $f : V \to \mathbb{C}$ can be written as $f(v_1)v_1^* + \cdots + f(v_n)v_n^*$. If τ is an invertible linear transformation of V then τ induces in a natural way an invertible linear transformation τ^* of V^*, defined by

$$\tau^*(f)(v) = f(\tau^{-1}(v)) \quad \text{for all } v \in V,\, f \in V^*. \tag{1.2.3}$$

(The transformation τ^{-1} on the right-hand side does indeed give the function that one would naturally expect, for the same reason that, in calculus, translating the graph of $y = f(x)$ one unit to the right gives the graph of $y = f(x-1)$.) To find the matrix of τ^* relative to the basis $v_1^*, \ldots, v_n^*$, observe that its (j,i) entry is the coefficient of v_j^* in $\tau^*(v_i^*)$; this is the same as $\tau^*(v_i^*)(v_j) = v_i^*(\tau^{-1}(v_j))$, the coefficient of v_i in $\tau^{-1}(v_j)$, i.e. the (i,j) entry of the matrix of τ^{-1} relative to $v_1, \ldots, v_n$. So, the map of matrices corresponding to $\tau \mapsto \tau^*$ is the inverse transpose map considered in Example 1.1.3. ■

Example 1.2.2. Take $W = V \oplus V = \{(v, v') \mid v, v' \in V\}$. Any linear transformation τ of V induces a linear transformation $\tau \oplus \tau$ of $V \oplus V$, defined by

$$(\tau \oplus \tau)(v, v') = (\tau(v), \tau(v')) \quad \text{for all } v, v' \in V. \tag{1.2.4}$$

The most obvious basis for $V \oplus V$ consists of

$$(v_1, 0), (v_2, 0), \ldots, (v_n, 0), (0, v_1), (0, v_2), \ldots, (0, v_n).$$

Relative to this basis, the matrix corresponding to $\tau \oplus \tau$ is exactly the block-diagonal duplication of the matrix of τ seen in Example 1.1.4; the conjugated versions mentioned there would arise if one used other bases of $V \oplus V$. ■

To explain Examples 1.1.2 and 1.1.5 similarly, we need the concept of the *tensor product*, which for finite-dimensional vector spaces can be explained fairly simply. Given two vector spaces V and W with respective bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$, the tensor product $V \otimes W$ is a vector space with basis $v_i \otimes w_j$ for all i, j with $1 \le i \le n$, $1 \le j \le m$. One can regard the elements $v_i \otimes w_j$ merely as symbols and $V \otimes W$ as the space of formal linear combinations of them. Note that the dimension of $V \otimes W$ is $(\dim V)(\dim W)$, in contrast with that of the direct sum $V \oplus W$, which is $\dim V + \dim W$. For arbitrary elements $v \in V$ and $w \in W$, we define the *pure tensor* $v \otimes w \in V \otimes W$ by the following rule:

$$\begin{gathered} \text{if } v = a_1 v_1 + \cdots + a_n v_n \quad \text{and} \quad w = b_1 w_1 + \cdots + b_m w_m \\ \text{then } v \otimes w = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j (v_i \otimes w_j). \end{gathered} \tag{1.2.5}$$

Note that if v happens to equal v_i and w happens to equal w_j then $v \otimes w$ does indeed equal the basis element $v_i \otimes w_j$, so our notation is consistent. Having made this definition, one can easily show that the tensor product does not depend on the chosen bases of V and W: for any other bases $v_1', \ldots, v_n'$ and $w_1', \ldots, w_m'$ the elements $v_i' \otimes w_j'$ form another, equally good, basis of $V \otimes W$. It is important to bear in mind that a general element of $V \otimes W$ is not a pure tensor: it is, of course, a linear combination of the basis elements $v_i \otimes w_j$ but the coefficients cannot usually be written in the form $a_i b_j$, as in (1.2.5).

So, we have another way to construct a new vector space from a vector space V: we can consider its tensor square $V^{\otimes 2} = V \otimes V$. Any linear transformation τ of V induces a linear transformation $\tau \otimes \tau$ of $V \otimes V$, defined on the basis elements by $(\tau \otimes \tau)(v_i \otimes v_j) = \tau(v_i) \otimes \tau(v_j)$. It is easy to see that in fact

$$(\tau \otimes \tau)(v \otimes v') = \tau(v) \otimes \tau(v') \quad \text{for any } v, v' \in V. \tag{1.2.6}$$

Example 1.2.3. Suppose that V is two-dimensional, with basis v_1, v_2. If the linear transformation $\tau : V \to V$ has matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ relative to this basis then, for instance,

$$\begin{aligned}
(\tau \otimes \tau)(v_1 \otimes v_1) &= \tau(v_1) \otimes \tau(v_1) \\
&= (av_1 + cv_2) \otimes (av_1 + cv_2) \\
&= a^2(v_1 \otimes v_1) + ac(v_1 \otimes v_2) + ac(v_2 \otimes v_1) + c^2(v_2 \otimes v_2).
\end{aligned}$$

This calculation gives the first column of the matrix of $\tau \otimes \tau$ relative to the basis $v_1 \otimes v_1, v_1 \otimes v_2, v_2 \otimes v_1, v_2 \otimes v_2$. The whole matrix is

$$\begin{pmatrix} a^2 & ab & ab & b^2 \\ ac & ad & bc & bd \\ ac & bc & ad & bd \\ c^2 & cd & cd & d^2 \end{pmatrix}.$$

The map sending $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to this matrix is a homomorphism from GL_2 to GL_4. To check (1.1.2) there is no need to make explicit matrix multiplications as in Example 1.1.5: the relation (1.1.2) follows from the fact that, for any linear transformations τ, τ' of V,

$$(\tau \circ \tau') \otimes (\tau \circ \tau') = (\tau \otimes \tau) \circ (\tau' \otimes \tau'), \tag{1.2.7}$$

which in turn follows because the two sides take the same values when evaluated on the basis elements. ■

As can be seen in Example 1.2.3, there are two subspaces of $V \otimes V$ that are guaranteed to be preserved by all linear transformations of the form $\tau \otimes \tau$: these are the space of *symmetric tensors*, $\mathrm{Sym}^2(V)$, consisting of elements that are invariant under the interchange map $v_i \otimes v_j \mapsto v_j \otimes v_i$, and the space of *alternating tensors*, $\mathrm{Alt}^2(V)$, consisting of elements that change sign under this interchange map. By restricting $\tau \otimes \tau$ to these subspaces we obtain further homomorphisms of the type referred to in Problem 1.1.1.

Example 1.2.4. Continuing with V two-dimensional, as in Example 1.2.3, $\mathrm{Sym}^2(V)$ is three-dimensional with basis $v_1 \otimes v_1, v_1 \otimes v_2 + v_2 \otimes v_1, v_2 \otimes v_2$. The resulting homomorphism is exactly the map $\Psi : GL_2 \to GL_3$ of Example 1.1.5. By contrast, $\mathrm{Alt}^2(V)$ is one-dimensional, spanned by $v_1 \otimes v_2 - v_2 \otimes v_1$. If $\tau : V \to V$ has matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ then

$$\begin{aligned}(\tau \otimes \tau)(v_1 \otimes v_2 - v_2 \otimes v_1) &= (av_1 + cv_2) \otimes (bv_1 + dv_2) \\ &\quad - (bv_1 + dv_2) \otimes (av_1 + cv_2) \\ &= (ad - bc)(v_1 \otimes v_2 - v_2 \otimes v_1).\end{aligned}$$

So the resulting homomorphism is the determinant $\det : GL_2 \to GL_1$, as in Example 1.1.2. ■

In general, if V has a basis $v_1, \dots, v_n$ then an element of $V \otimes V$ lies in $\mathrm{Sym}^2(V)$ if and only if the coefficient of $v_i \otimes v_j$ equals the coefficient of $v_j \otimes v_i$ for all i, j. Hence $\mathrm{Sym}^2(V)$ has a basis consisting of the following elements:

$$v_i \otimes v_i \ \text{ for } 1 \leq i \leq n \quad \text{ and } \quad v_i \otimes v_j + v_j \otimes v_i \quad \text{ for } 1 \leq i < j \leq n.$$

An element of $V \otimes V$ lies in $\mathrm{Alt}^2(V)$ if and only if the coefficient of $v_i \otimes v_i$ is zero for every i and the coefficient of $v_i \otimes v_j$ is the negative of the coefficient of $v_j \otimes v_i$ for all $i \neq j$. Hence $\mathrm{Alt}^2(V)$ has a basis consisting of the elements

$$v_i \otimes v_j - v_j \otimes v_i \quad \text{for } 1 \leq i < j \leq n.$$

Clearly we have a direct sum decomposition,

$$V \otimes V = \mathrm{Sym}^2(V) \oplus \mathrm{Alt}^2(V), \tag{1.2.8}$$

and the dimensions of $\mathrm{Sym}^2(V)$ and $\mathrm{Alt}^2(V)$ are $\binom{n+1}{2}$ and $\binom{n}{2}$ respectively.

As well as tensor squares, one can define higher tensor powers in an entirely analogous way: $V^{\otimes 3} = V \otimes V \otimes V$, $V^{\otimes 4} = V \otimes V \otimes V \otimes V$, and so forth. If V has a basis $v_1, \dots, v_n$ then the k-fold tensor power $V^{\otimes k}$ has a basis consisting of the pure tensors:

$$v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k} \quad \text{for } 1 \leq i_1, \dots, i_k \leq n.$$

So $\dim V^{\otimes k} = n^k$. (By convention, $V^{\otimes 1}$ is V itself.) The space of symmetric tensors $\mathrm{Sym}^k(V)$ consists of those elements of $V^{\otimes k}$ that are fixed under any permutation of the tensor factors. In other words, the coefficients of $v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k}$ and $v_{j_1} \otimes v_{j_2} \otimes \cdots \otimes v_{j_k}$ have to be the same whenever $j_1, \dots, j_k$ can be obtained by rearranging $i_1, \dots, i_k$. So $\mathrm{Sym}^k(V)$ has a basis consisting of all the elements

$$t_{(k_1,\dots,k_n)} := \sum_{\substack{1 \leq s_1, \dots, s_k \leq n, \\ k_i \text{ of the } s_j \\ \text{equal } i}} v_{s_1} \otimes \cdots \otimes v_{s_k},$$

where $(k_1, \dots, k_n)$ runs over all n-tuples of nonnegative integers such that $k_1 + \cdots + k_n = k$. (Note that $\mathrm{Sym}^1(V) = V$.) Hence

$$\dim \mathrm{Sym}^k(V) = \binom{n+k-1}{k}. \tag{1.2.9}$$

The space of alternating tensors $\mathrm{Alt}^k(V)$ consists of those elements of $V^{\otimes k}$ which change sign under any interchange of two tensor factors, so under a general permutation σ in the symmetric group S_k they are multiplied by the sign $\varepsilon(\sigma)$. This forces the coefficient of $v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k}$ to be zero whenever two of the i_j coincide, and if the i_j are all distinct then it forces the coefficient of $v_{i_{\sigma(1)}} \otimes v_{i_{\sigma(2)}} \otimes \cdots \otimes v_{i_{\sigma(k)}}$ to be $\varepsilon(\sigma)$ times the coefficient of $v_{i_1} \otimes v_{i_2} \otimes \cdots \otimes v_{i_k}$, for all permutations $\sigma \in S_k$. So $\mathrm{Alt}^k(V) = \{0\}$ for $k > n$ and, for $k \leq n$, $\mathrm{Alt}^k(V)$ has a basis consisting of the elements

$$u_{i_1,\dots,i_k} := \sum_{\sigma \in S_k} \varepsilon(\sigma)\, v_{i_{\sigma(1)}} \otimes v_{i_{\sigma(2)}} \otimes \cdots \otimes v_{i_{\sigma(k)}} \quad \text{for } 1 \leq i_1 < \cdots < i_k \leq n.$$

(Note that $\mathrm{Alt}^1(V) = V$.) Hence

$$\dim \mathrm{Alt}^k(V) = \binom{n}{k} \quad \text{(which is 0 if } k > n). \tag{1.2.10}$$

In particular, $\mathrm{Alt}^n(V)$ is one-dimensional, spanned by the alternating tensor

$$u_{1,\dots,n} = \sum_{\sigma \in S_n} \varepsilon(\sigma)\, v_{\sigma(1)} \otimes \cdots \otimes v_{\sigma(n)}.$$

The homomorphism $GL_n \to GL_1$ obtained by considering induced linear transformations of $\mathrm{Alt}^n(V)$ is always the determinant homomorphism, as in Example 1.2.4.

Many questions now arise. Purely from the dimensions we can see that, for $k \geq 3$, the sum of $\mathrm{Sym}^k(V)$ and $\mathrm{Alt}^k(V)$ is far from being the whole of $V^{\otimes k}$; so are there other subspaces of $V^{\otimes k}$ that are 'natural' in the sense of being preserved by any linear transformation of the form $\tau \otimes \cdots \otimes \tau$? When are the resulting homomorphisms $GL_n \to GL_m$ the same or equivalent (up to conjugation)? What do you get when you form tensor products like $\mathrm{Sym}^2(V) \otimes \mathrm{Alt}^3(V)$, or compose operations as in $\mathrm{Sym}^2(\mathrm{Alt}^3(V^*))$, or both? Such problems in multilinear algebra would seem easier to address if we had a satisfactory answer to Problem 1.1.1.

1.3 Linearization of the problem

One obstacle to solving Problem 1.1.1 directly is that the functions Φ_{ij} are in general *nonlinear*, so they cannot be specified simply by giving their values on basis elements. Lie's brilliant idea was to *linearize* Problem 1.1.1 by considering the derivative of $\Phi : GL_n \to GL_m$ at the identity matrix 1_n. By definition, this is the linear map

$$\begin{aligned} \varphi &: \mathrm{Mat}_n \to \mathrm{Mat}_m \\ &(a_{ij}) \mapsto \text{the matrix whose } (k,l) \text{ entry is } \sum_{i,j=1}^{n} a_{ij}(\partial^{ij}\Phi_{kl})(1_n), \end{aligned}$$

where ∂^{ij} denotes the partial derivative with respect to the (i,j)-entry. To make the idea work we obviously need these partial derivatives to exist, so this condition should be retrospectively added to Problem 1.1.1. Certainly the partial derivatives exist in all the examples we have seen so far, where the functions Φ_{kl} were polynomials or rational functions in the entries.

Example 1.3.1. The derivative of the homomorphism $\Psi : GL_2 \to GL_3$ in Example 1.1.5 is the linear map $\psi : \mathfrak{gl}_2 \to \mathfrak{gl}_3$ given by

$$\psi \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2a & 2b & 0 \\ c & a+d & b \\ 0 & 2c & 2d \end{pmatrix}.$$

One might wonder whether ψ satisfies (1.1.2). To find out, we make the following calculation:

$$\begin{aligned} &\psi \begin{pmatrix} a & b \\ c & d \end{pmatrix} \psi \begin{pmatrix} e & f \\ g & h \end{pmatrix} - \psi \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} e & f \\ g & h \end{pmatrix} \right) \\ &= \begin{pmatrix} 2a & 2b & 0 \\ c & a+d & b \\ 0 & 2c & 2d \end{pmatrix} \begin{pmatrix} 2e & 2f & 0 \\ g & e+h & f \\ 0 & 2g & 2h \end{pmatrix} \\ &\quad - \begin{pmatrix} 2(ae+bg) & 2(af+bh) & 0 \\ ce+dg & (ae+bg)+(cf+dh) & af+bh \\ 0 & 2(ce+dg) & 2(cf+dh) \end{pmatrix} \\ &= \begin{pmatrix} 2ae & 2af+2be & 2bf \\ ce+ag & ah+bg+cf+de & df+bh \\ 2cg & 2ch+2dg & 2dh \end{pmatrix}. \end{aligned}$$

The result is not zero, but it is noticeable that it is symmetric in the two matrices used (it remains unchanged if we swap a with e, b with f, c with g, and d with h). In other words, ψ does satisfy the property

$$\psi(x)\psi(y) - \psi(xy) = \psi(y)\psi(x) - \psi(yx) \quad \text{for all } x, y \in \mathrm{Mat}_2.$$

■

Lie observed that this identity holds in general.

Proposition 1.3.2. *If $\Phi : GL_n \to GL_m$ is a group homomorphism such that all the partial derivatives $\partial^{ij}\Phi_{kl}$ exist at 1_n, and $\varphi : \mathrm{Mat}_n \to \mathrm{Mat}_m$ is the derivative of Φ at 1_n as above, then*

$$\varphi(xy - yx) = \varphi(x)\varphi(y) - \varphi(y)\varphi(x) \quad \textit{for all } x, y \in \mathrm{Mat}_n.$$

Proof. (The reader who is prepared to accept the result can safely skip this proof; it relies on moderately sophisticated calculus techniques which we will not need elsewhere.) Since Φ is a homomorphism, we have

$$\Phi(ghg^{-1}) = \Phi(g)\Phi(h)\Phi(g)^{-1} \quad \text{for all } g, h \in GL_n. \tag{1.3.1}$$

The first step is to differentiate both sides of this equation as functions of h, keeping g fixed, and evaluating the result at $h = 1_n$. For all $g \in GL_n$, the conjugation map $GL_n \to GL_n : h \mapsto ghg^{-1}$ is the restriction of the linear map $\mathrm{Mat}_n \to \mathrm{Mat}_n : y \mapsto gyg^{-1}$, so its derivative at 1_n is this linear map. By the chain rule the derivative of $GL_n \to GL_m : h \mapsto \Phi(ghg^{-1})$ at 1_n is $\mathrm{Mat}_n \to \mathrm{Mat}_m : y \mapsto \varphi(gyg^{-1})$. Reasoning similarly for the right-hand side of (1.3.1), we obtain the desired differentiated equation:

$$\varphi(gyg^{-1}) = \Phi(g)\varphi(y)\Phi(g)^{-1} \quad \text{for all } g \in GL_n, y \in \mathrm{Mat}_n. \tag{1.3.2}$$

The remaining step is to differentiate both sides again but this time as functions of g, keeping y fixed and evaluating the result at $g = 1_n$. For any $y \in \mathrm{Mat}_n$, the derivative of $GL_n \to \mathrm{Mat}_n : g \mapsto gyg^{-1}$ at 1_n is $\mathrm{Mat}_n \to \mathrm{Mat}_n : x \mapsto xy - yx$, by virtue of the power series expansion

$$\begin{aligned}(1_n + tx)y(1_n + tx)^{-1} &= \sum_{i=0}^{\infty}(1_n + tx)y(-tx)^i \\ &= y + t(xy - yx) + t^2(yx^2 - xyx) + \cdots .\end{aligned} \tag{1.3.3}$$

So, again by the chain rule, the derivative of $GL_n \to \mathrm{Mat}_m : g \mapsto \varphi(gyg^{-1})$ at 1_n is $\mathrm{Mat}_n \to \mathrm{Mat}_m : x \mapsto \varphi(xy - yx)$. By similar reasoning, the derivative of $GL_n \to \mathrm{Mat}_m : g \mapsto \Phi(g)\varphi(y)\Phi(g)^{-1}$ at 1_n is $\mathrm{Mat}_n \to \mathrm{Mat}_m : x \mapsto \varphi(x)\varphi(y) - \varphi(y)\varphi(x)$. So the differentiation of (1.3.2) gives the result. □

Proposition 1.3.2 can be written more concisely using the bracket notation $[x, y] := xy - yx$. It says that

$$\varphi([x, y]) = [\varphi(x), \varphi(y)] \quad \text{for all } x, y \in \mathrm{Mat}_n. \tag{1.3.4}$$

The quantity $[x, y]$ is called the *commutator* of x and y, since it measures to what extent the commutation equation $xy = yx$ fails to hold.

We are now in the curious situation of being interested in linear maps $\varphi : \mathrm{Mat}_n \to \mathrm{Mat}_m$ that preserve the commutator, in the sense of (1.3.4), but do not necessarily preserve the actual multiplication. The structure on Mat_n given solely by the commutator is that of a *Lie algebra* and is our first example of the topic of this book. When Mat_n is viewed as a Lie algebra (i.e. when matrix multiplication is not under consideration, but commutators are) it is written as $\mathfrak{gl}_n$; the notation is meant to suggest that it is a linearized version of the group GL_n. Linear maps $\varphi : \mathfrak{gl}_n \to \mathfrak{gl}_m$ satisfying (1.3.4) are called *Lie algebra homomorphisms*. So the linearization of Problem 1.1.1 is the following.

Problem 1.3.3. Describe all Lie algebra homomorphisms $\varphi : \mathfrak{gl}_n \to \mathfrak{gl}_m$.

Mainly because it concerns linear maps, Problem 1.3.3 is easier to tackle than Problem 1.1.1. Nevertheless, it is not a trivial matter: much of the rest of this book will be devoted to refining the statement of Problem 1.3.3 and solving it.

1.4 Lie's theorem

It may now appear that we have carried out the customary mathematician's dodge of replacing the problem that we really intended to solve, Problem 1.1.1, with a problem that is more tractable, Problem 1.3.3. However, because of the bridge between Lie groups and Lie algebras the two problems are almost equivalent. In this section we will briefly explain why this is so.

The first key result is that the derivative φ determines the homomorphism Φ. This is surprising, because you would think there could be two homomorphisms $GL_n \to GL_m$ that had the same derivative at 1_n but different 'higher-order terms'. We need to use the fact that every element of GL_n can be written as $\exp(x)$ for some $x \in \mathrm{Mat}_n$, where $\exp$ is the matrix exponential defined by the (always convergent) series

$$\exp(x) = 1_n + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots .$$

Proposition 1.4.1. *Under the hypotheses of Proposition 1.3.2,*

$$\Phi(\exp(x)) = \exp(\varphi(x)) \quad \textit{for all } x \in \mathrm{Mat}_n.$$

So Φ is determined by φ.

Proof. Consider the Mat_m-valued functions of one variable

$$f(t) = \Phi(\exp(tx)) \quad \text{and} \quad g(t) = \exp(-\varphi(tx)).$$

We have $f(t+s) = f(t)f(s)$ and $g(s+t) = g(s)g(t)$. It follows that

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

and similarly $g'(t) = -\varphi(x)g(t)$. Hence

$$(f(t)g(t))' = f'(t)g(t) + f(t)g'(t) = f(t)\varphi(x)g(t) - f(t)\varphi(x)g(t) = 0,$$