LECTURE NOTES ON MATRIX GROUPS

C.C. REMSING

Contents

Date: February 2018.

Key words and phrases. Group of transformations, symmetric group, group action, permutation representation, Euclidean space, linear Lie group, matrix exponential, Lie algebra.

1. Groups of Transformations

Maps and groups • *Permutations of a finite set* • *Morphisms of groups* • *Cosets and quotient groups.*

1.1. **Maps and groups.** Let M be a non-empty set. A **transformation** of M is a function (map or mapping) from M to M. The *identity mapping* on M is denoted by id_{M}. Let M^{M} denote the set of all transformations of M. (An element $\alpha \in \mathsf{M}^{\mathsf{M}}$ is written symbolically as $\alpha : \mathsf{M} \to \mathsf{M}$ or $\mathsf{M} \overset{\alpha}{\to} \mathsf{M}$.) M^{M} is a *monoid* with identity element $\mathbf{1}_{\mathsf{M}} = \mathrm{id}_{\mathsf{M}}$.

Note: A **semigroup** $(\mathsf{M}, *)$ consists of a (non-empty) set M on which an *associative* binary operation $*$ is defined; that is, $(\alpha * \beta) * \gamma = \alpha * (\beta * \gamma)$ for all elements $\alpha, \beta, \gamma \in \mathsf{M}$. If there exists an element ϵ satisfying $\alpha * \epsilon = \alpha = \epsilon * \alpha$ for all $\alpha \in \mathsf{M}$, then the semigroup is called a **monoid** having identity element ϵ. This element can easily be seen to be unique, and is usually denoted by $\mathbf{1}_{\mathsf{M}}$. An element α in a monoid $(\mathsf{M}, *)$ is called *invertible* if there exists an element $\beta \in \mathsf{M}$ such that $\alpha * \beta = \mathbf{1}_{\mathsf{M}} = \beta * \alpha$. (Clearly, in that case β is also invertible.)

The natural operation on M^{M} is the usual composition of mappings; the composite mapping $\alpha \circ \beta$ is called the *product* of α and β (in this order) and is denoted simply by $\alpha\beta$. (In general, the product of mappings is *not* commutative.)

A map (transformation) $\alpha : \mathsf{M} \to \mathsf{M}$ is said to be **injective** (or one-to-one) if $x_1 \neq x_2$ implies $\alpha(x_1) \neq \alpha(x_2)$ $(x_1, x_2 \in \mathsf{M})$; it is said to be **surjective** (or onto) if for every $y \in \mathsf{M}$ there exists (at least one) $x \in \mathsf{M}$ such that $\alpha(x) = y$.

◇ **Exercise 1.** Let $\alpha \in \mathsf{M}^{\mathsf{M}}$. Show that the following statements are logically equivalent.

(a) α is injective.
(b) $\alpha(x_1) = \alpha(x_2)$ implies $x_1 = x_2$ $(x_1, x_2 \in \mathsf{M})$.
(c) For every $y \in \mathsf{M}$ there exists *at most* one $x \in \mathsf{M}$ such that $\alpha(x) = y$.

◇ **Exercise 2.** Let $\alpha, \beta \in \mathsf{M}^{\mathsf{M}}$. Show that

(a) if α and β are injective, then so is the product $\alpha\beta$;
(b) if α and β are surjective, then so is the product $\alpha\beta$.

◇ **Exercise 3.** Let $\alpha, \beta \in \mathsf{M}^{\mathsf{M}}$ such that $\beta\alpha = \mathbf{1}_{\mathsf{M}}$. Show that α is injective and β is surjective.

As in any monoid, an element $\alpha \in \mathsf{M}^{\mathsf{M}}$ is said to be **invertible** if there exists an element $\beta \in \mathsf{M}^{\mathsf{M}}$ such that $\alpha\beta = \mathbf{1}_{\mathsf{M}} = \beta\alpha$. If that is the case, β is called an *inverse* of α. If a mapping is invertible, then its inverse is unique (this is Problem 2), and is denoted by α^{-1}.

Proposition 1. *A transformation of* M *is invertible if and only if it is both injective and surjective (i.e., bijective).*

Proof. $(\Rightarrow)$ Suppose that $\alpha \in \mathsf{M}$ has an inverse $\beta = \alpha^{-1}$. Then

$$\beta\alpha = \mathbf{1}_{\mathsf{M}} \quad \text{and} \quad \alpha\beta = \mathbf{1}_{\mathsf{M}}.$$

These two conditions and **Exercise 2** give both injectivity and surjectivity of α.

$(\Leftarrow)$ Conversely, if we suppose that α is bijective, for any $y \in \mathsf{M}$ we can find a *unique* element $x \in \mathsf{M}$ such that $\alpha(x) = y$. Setting $\beta(y) := x$, we define a map $\beta : \mathsf{M} \to \mathsf{M}$ such that $\alpha\beta = \mathbf{1}_{\mathsf{M}} = \beta\alpha$. Thus $\alpha^{-1} = \beta$. □

Corollary 2. *If* $\alpha \in \mathsf{M}^{\mathsf{M}}$ *is invertible, then* α^{-1} *is also invertible and* $\left(\alpha^{-1}\right)^{-1} = \alpha$.

Proof. The equations

$$\alpha^{-1}\alpha = \mathbf{1}_{\mathsf{M}} \quad \text{and} \quad \alpha\alpha^{-1} = \mathbf{1}_{\mathsf{M}}$$

show that α is the inverse of α^{-1}. □

Corollary 3. *If* $\alpha, \beta \in \mathsf{M}^{\mathsf{M}}$ *are invertible, then the product (composition)* $\alpha\beta$ *is also invertible, and* $(\alpha\beta)^{-1} = \beta^{-1}\alpha^{-1}$.

Proof. The equations

$$(\alpha\beta)(\beta^{-1}\alpha^{-1}) = \alpha(\beta\beta^{-1})\alpha^{-1} = \alpha\alpha^{-1} = \mathbf{1}_{\mathsf{M}}$$
$$(\beta^{-1}\alpha^{-1})(\alpha\beta) = \beta^{-1}(\alpha^{-1}\alpha)\beta = \beta^{-1}\beta = \mathbf{1}_{\mathsf{M}}$$

imply that $\beta^{-1}\alpha^{-1}$ is the inverse of $\alpha\beta$. □

Let M be a non-empty set. An invertible transformation of M is called a **permutation** (or **symmetry**) of M. The collection of all permutations of M form a *group*, denoted by $\mathfrak{S}_{\mathsf{M}}$ and called the **symmetric group** on M. So

$$\mathfrak{S}_{\mathsf{M}} := \left\{\alpha \in \mathsf{M}^{\mathsf{M}} \,:\, \alpha \text{ invertible}\right\}. \tag{1}$$

Note: A monoid G all of whose elements are invertible is called a **group**. In other words, the following axioms must hold:

(G1) a binary operation $(g_1, g_2) \mapsto g_1g_2$ is defined on the set G;
(G2) this operation is associative: $(g_1g_2)g_3 = g_1(g_2g_3)$ for all $g_1, g_2, g_3 \in \mathsf{G}$;
(G3) G has a neutral (identity) element $\mathbf{1}_{\mathsf{G}} = \mathbf{1}$: $g\mathbf{1} = \mathbf{1}g = g$ for all $g \in \mathsf{G}$;
(G4) every element $g \in \mathsf{G}$ has an inverse g^{-1}: $gg^{-1} = g^{-1}g = \mathbf{1}$.

A group G is said to be **Abelian** if the group operation is *commutative* ($g_1 g_2 = g_2 g_1$ for all $g_1, g_2 \in \mathsf{G}$); the name "Abelian" is in honour of the Norwegian mathematician N.H. Abel (1802–1829). The term "group" itself was introduced by the French mathematician E. Galois (1811–1832) who considered finite groups of permutations.

A **subgroup** of a group G is a subset of G which itself forms a group under the group operation (multiplication). *A non-empty subset* H *of a group* G *is a subgroup if and only if* $g_1^{-1} g_2 \in \mathsf{H}$ *whenever* $g_1, g_2 \in \mathsf{H}$. If H is a subgroup of G we write $\mathsf{H} \leq \mathsf{G}$; H is said to be a *proper* subgroup if $\mathsf{H} \neq \mathsf{G}$. Any intersection of subgroups is a subgroup of G.

Here are some examples of groups.

(1) *Groups of numbers.* Let $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ denote respectively the set of all integers, rational numbers, real numbers, and complex numbers. Each set becomes a group if we specify ordinary addition as the group operation. The sets $\mathbb{Q} \setminus \{0\}, \mathbb{R} \setminus \{0\}$, and $\mathbb{C} \setminus \{0\}$ are groups with respect to ordinary multiplication. All these groups are Abelian.

(2) *Groups of matrices.* Let $\Bbbk$ be a *field* (think of $\mathbb{R}$ or $\mathbb{C}$) and let $\mathsf{GL}\,(n, \Bbbk)$ denote the set of all *nonsingular* $n \times n$ matrices over $\Bbbk$. Taking matrix multiplication as the group operation we see that $\mathsf{GL}\,(n, \Bbbk)$ is a group. This group is called the **general linear group** (of degree n over $\Bbbk$).

(3) *Groups of linear transformations.* If V is an n-dimensional *vector space* over the field $\Bbbk$, let $\mathsf{GL}\,(\mathsf{V})$ denote the set of all bijective linear transformations of V. Then $\mathsf{GL}\,(\mathsf{V})$ is a group if the usual functional composition is specified as the group operation: $\alpha\beta(v) := \alpha(\beta(v))$ for $v \in \mathsf{V}$ and $\alpha, \beta \in \mathsf{GL}\,(\mathsf{V})$.

There is a close connection between the groups $\mathsf{GL}\,(\mathsf{V})$ and $\mathsf{GL}\,(n, \Bbbk)$. For, if a fixed ordered basis for V is chosen, each bijective linear transformation of V is associated with a nonsingular $n \times n$ matrix over $\Bbbk$. This correspondence is an isomorphism from $\mathsf{GL}\,(\mathsf{V})$ to $\mathsf{GL}\,(n, \Bbbk)$ (the reason being that when two linear transformations are composed, the product of the corresponding matrices represents the composite).

(4) *Groups of isometries.* Let M be a *metric space* (with a distance function $d : \mathsf{M} \times \mathsf{M} \to \mathbb{R}$). An **isometry** of M is a bijective mapping $\alpha : \mathsf{M} \to \mathsf{M}$ (i.e., a permutation of M) which preserves distances: $d\,(\alpha(x), \alpha(y)) = d(x, y)$ for all $x, y \in \mathsf{M}$. It is easy to verify that the set of all isometries of M is a group (with respect to the operation of functional composition). We shall write this group $\mathsf{Isom}\,(\mathsf{M})$.

Suppose that S is a non-empty subset of (the metric space) M. If α is an isometry, define $\alpha \cdot \mathsf{S}$ to be the set $\{\alpha(x) \,:\, x \in \mathsf{S}\}$. The **symmetry group** of S (with respect to M) is the set

$$\mathsf{Sym}\,(\mathsf{S}) := \{\alpha \in \mathsf{Isom}\,(\mathsf{M}) \,:\, \alpha \cdot \mathsf{S} = \mathsf{S}\}$$

of all isometries that leave S fixed as a set (together with functional composition). Again, it is clear that this is a group. The more "symmetrical" the set S is, the larger is its symmetry group. Thus we arrive at the fundamental idea of a group as a measure of the symmetry of a structure.

(5) *Group-valued functions.* Let M be a non-empty set and let G be a group. The set G^{M} of all functions $\alpha : \mathsf{M} \to \mathsf{G}$ is a group with the group operation defined pointwise:

$\alpha\beta(x) := \alpha(x)\beta(x)$ for $x \in \mathsf{M}$ and $\alpha, \beta \in \mathsf{G}^{\mathsf{M}}$. In particular, the set $\mathbb{R}^{\mathsf{M}}$ of all real-valued functions defined on M is a group; clearly, this group is Abelian.

Notice that the set $\Bbbk^{[n]\times[n]}$ of all $\Bbbk$-valued functions defined on the set $\mathsf{M} = [n] \times [n]$ (where $[n] := \{1, 2, \dots, n\}$) is precisely the set of $n \times n$ matrices over (the ring) $\Bbbk$. We shall write this (Abelian) group $\Bbbk^{n\times n}$.

Definition 1. Any subgroup of the symmetric group $\mathfrak{S}_{\mathsf{M}}$ is called a **transformation group** (or **permutation group**).

The *trivial group* $\{\mathbf{1}_{\mathsf{M}}\}$ and the symmetric group $\mathfrak{S}_{\mathsf{M}}$ itself are, of course, transformation groups. A collection $\mathsf{G} \subseteq \mathfrak{S}_{\mathsf{M}}$ of invertible transformations (permutations) of M is a transformation group if and only if $\alpha^{-1}\beta \in \mathsf{G}$ for all $\alpha, \beta \in \mathsf{G}$.

◇ **Exercise 4.**

(a) Let G be a group and $g \in \mathsf{G}$ is such that $\{g, g^2, g^3, \dots\}$ is *finite*. Show that there exists a positive integer k such that $g^k = \mathbf{1}$.

(b) Let G be a group and let S be a (non-empty) finite subset of G. Prove that S is a subgroup of G if and only if $g_1g_2 \in \mathsf{S}$ for all $g_1, g_2 \in \mathsf{S}$.

Example 2. Let $\Bbbk = \mathbb{R}$ or $\mathbb{C}$. In the general linear group $\mathsf{GL}\,(n, \Bbbk)$, consider the subset $\mathsf{SL}\,(n, \Bbbk)$ of matrices with determinant 1:

$$\mathsf{SL}\,(n, \Bbbk) := \{a \in \mathsf{GL}\,(n, \Bbbk) \,:\, \det a = 1\}\,.$$

Clearly, the **identity matrix** $\mathbf{1} = \mathrm{Id}_n \in \mathsf{SL}\,(n, \Bbbk)$. (The $n \times n$ matrix $\mathrm{Id}_n = [\delta_{ij}] \in \mathsf{SL}\,(n, \Bbbk)$ corresponds to the identity map $\mathrm{id}_{\Bbbk^n} : \Bbbk^n \to \Bbbk^n$.) The expression for the determinant of a product (i.e., $\det(ab) = \det a \cdot \det b$) implies that $\mathsf{SL}\,(n, \Bbbk)$ is a subgroup of $\mathsf{GL}\,(n, \Bbbk)$; it is called the **special linear group** (of degree n over $\Bbbk$).

The group $\mathsf{GL}\,(n, \Bbbk)$, which contains many other interesting groups (called **matrix groups**), has been for mathematicians of several generations a seemingly inexhaustible source of new ideas and unsolved problems.

Example 3. Transformations of the real line $\mathbb{R}$ of the form

$$\tau_{a,b} : x \mapsto ax + b \qquad (a, b \in \mathbb{R}, \quad a \neq 0)$$

are called **affine transformations**. Clearly, the inverse of any such transformation is another of the same form: $\tau_{a,b}^{-1} : x \mapsto \frac{1}{a}x - \frac{b}{a}$. The set $\mathsf{Aff}\,(1, \mathbb{R})$ of all these transformations is a group, called the **affine group** (of the real line).

◇ **Exercise 5.** Show that the product of two affine transformations $\tau_{a,b}$ and $\tau_{c,d}$ is also an affine transformation.

The group $\mathsf{Aff}\,(1,\mathbb{R})$ can be viewed as a group of 2×2 matrices over $\mathbb{R}$: the transformation $\tau_{a,b}$ corresponds to the matrix $\begin{bmatrix}1 & 0\\ b & a\end{bmatrix} \in \mathsf{GL}\,(2,\mathbb{R})$ because

$$\begin{bmatrix}1 & 0\\ b & a\end{bmatrix}\begin{bmatrix}1\\ x\end{bmatrix} = \begin{bmatrix}1\\ ax+b\end{bmatrix}.$$

◇ **Exercise 6.** Use the fact that

$$\begin{bmatrix}1 & 0\\ b & a\end{bmatrix}^{-1} = \begin{bmatrix}1 & 0\\ -\frac{b}{a} & \frac{1}{a}\end{bmatrix}$$

to work out the inverse of (the affine transformation) $\tau_{a,b}$.

The group $\mathsf{Aff}\,(1,\mathbb{R})$ contains the subgroup $\mathsf{GL}\,(1,\mathbb{R})$ of linear transformations (affine transformations which leave the point $x=0$ fixed), and the subgroup of "translations" $x\mapsto x+b$.

Example 4. A **Möbius transformation** (or a *fractional linear transformation*) is a function μ of a complex variable z that can be written in the form

$$\mu(z) = \frac{az+b}{cz+d} \tag{2}$$

for some complex numbers a,b,c,d with $ad-bc\neq 0$.

The deceptively simple form of (2) conceals two problems. First, a Möbius transformation can be written in the form (2) in many ways (just as a rational number can be written as $\frac{r}{s}$ in many ways). The second problem stems from the fact that, for example, $z\mapsto\frac{1}{z-z_0}$ is not defined at the point z_0; this means that there is no subset of $\mathbb{C}$ on which all Möbius maps are defined. Informally, the first difficulty is resolved by saying that the 4-tuple (a,b,c,d) is determined within a (complex) scalar multiple. The second difficulty is resolved by joining an extra point (which is called the *point at infinity*) to $\mathbb{C}$; this new point is denoted by ∞.

We make the following (standard) conventions: if $c\neq 0$, we define $\mu(\infty) := \frac{a}{c}$ and $\mu(-\frac{d}{c}) := \infty$; if $c=0$, we define $\mu(\infty):=\infty$. The set $\mathbb{C}\cup\{\infty\}$ is called the **extended complex plane** (sometimes called the *complex projective line*) and is denoted by $\mathbb{C}_\infty$.

It turns out that *each Möbius transformation is a bijection of* $\mathbb{C}_\infty$ *onto itself* (i.e., a permutation of the set $\mathbb{C}_\infty$). The set $\mathsf{Möb}$ of all Möbius transformations (on $\mathbb{C}_\infty$) is a group, called the **Möbius group**.

◇ **Exercise 7.** Show that

(a) the product of two Möbius transformations is another Möbius transformation;
(b) each Möbius transformation has an inverse which is also a Möbius transformation.

1.2. **Permutations of a finite set.** Let M be a finite set with m elements. We may assume that $\mathsf{M} = \{1, 2, \dots, m\} =: [m]$. The group $\mathfrak{S}_{\mathsf{M}}$ is called the **symmetric group on** m **elements** and is denoted by $\mathfrak{S}_m$. The elements of $\mathfrak{S}_m$ are called permutations (of degree m).

◇ **Exercise 8.** Let M be a finite set with m elements. Show that

$$\left|\mathsf{M}^{\mathsf{M}}\right| = m^m \quad \text{and} \quad |\mathfrak{S}_m| = m!.$$

(The symbol $|\mathsf{S}|$ denotes the number of elements of the finite set S.)

It is customary, and convenient, to write a permutation $\pi \in \mathfrak{S}_m$ in the form

$$\pi = \begin{bmatrix} 1 & 2 & \dots & m \\ \pi(1) & \pi(2) & \dots & \pi(m) \end{bmatrix}$$

where the image $\pi(i)$ of i is placed in the second row underneath i in the first row; for example, the permutation (of degree four) such that $1 \mapsto 4, 2 \mapsto 2, 3 \mapsto 1$ and $4 \mapsto 3$ is denoted by

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 2 & 1 & 3 \end{bmatrix}.$$

Two permutations $\pi, \sigma \in \mathfrak{S}_m$ are multiplied by the usual rule for composing maps: $(\pi\sigma)(i) = \pi(\sigma(i))$.

Example 5. The elements of $\mathfrak{S}_3$ are

$$\mathbf{1} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix}.$$

Here are two representative computations:

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Notice that

$$\begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{bmatrix}.$$

Therefore the symmetric group $\mathfrak{S}_3$ is not Abelian. We can immediately say that $\mathfrak{S}_m$ *is not Abelian when* $m \geq 3$. Why ? (In 1770, J.L. Lagrange (1736–1813) studied the groups $\mathfrak{S}_2, \mathfrak{S}_3$ and $\mathfrak{S}_4$ in relation to the solutions of equations of degree $2, 3$ and 4.)

Permutations in $\mathfrak{S}_m$ can be decomposed into products of simpler permutations. Let $a_1, a_2, \ldots, a_k \in [m]$. A permutation $\pi \in \mathfrak{S}_m$ such that

$$a_1 \mapsto a_2 \mapsto a_3 \mapsto \cdots \mapsto a_k \mapsto a_1$$

and leaving all other integers in $[m]$ fixed, is called a **cyclic permutation** and is denoted by $(a_1\ a_2\ \ldots\ a_k)$. The number k is its *length* and a cyclic permutation of length k is called a k-**cycle**. If $a \in [m]$, then the 1-cycle (a) is the identity permutation $\mathbf{1} \in \mathfrak{S}_m$.

Example 6. The elements of $\mathfrak{S}_3$ are (in cycle notation)

$$\mathbf{1} = (1) = (2) = (3),\ (1\ 2),\ (1\ 3),\ (2\ 3),\ (1\ 2\ 3),\ (1\ 3\ 2).$$

Notice that the calculation in Example 5 becomes

$$(1\ 3\ 2)(1\ 3) = (1\ 2) \neq (2\ 3) = (1\ 3)(1\ 3\ 2).$$

Two cyclic permutations $\alpha = (a_1\ a_2\ \ldots\ a_r)$ and $\beta = (b_1\ b_2\ \ldots\ b_s)$ in $\mathfrak{S}_m$ are said to be **disjoint** if $a_i \neq b_j$ for all $i, j \in [m]$. For example, $(1\ 2\ 4)$ and $(3\ 5\ 6)$ are disjoint, but $(1\ 2\ 4)$ and $(3\ 4\ 6)$ are not. *Disjoint cycles commute*; that is, if α and β represent disjoint cycles, then $\alpha\beta = \beta\alpha$ (this is Problem 3).

Theorem 4. *Any permutation of a finite set is either a cycle or can be written as a product of pairwise disjoint cycles; and, except for the order in which the cycles are written, and the inclusion or omission of 1-cycles, this can be done in only one way.*

We shall omit the proof of this theorem, but it is illustrated in the following example.

Example 7. In each of the following equations the cycles on the right are pairwise disjoint.

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 1 & 5 & 2 \end{bmatrix} = (1\ 3)(2\ 4\ 5) = (2\ 4\ 5)(1\ 3)$$
$$(1\ 4\ 5)(2\ 3\ 5) = (1\ 4\ 5\ 2\ 3)$$
$$(1\ 6)(1\ 5)(1\ 4)(1\ 3)(1\ 2) = (1\ 2\ 3\ 4\ 5\ 6)$$
$$(1\ 2\ 3\ 4)^{-1} = (4\ 3\ 2\ 1) = (1\ 4\ 3\ 2)$$
$$(1\ 5\ 4\ 6\ 3)(4\ 3\ 6)(2\ 5) = (1\ 5)(2\ 4).$$

Note: A 2-cycle is called a **transposition**. For example, the transpositions in $\mathfrak{S}_3$ are $(1,2), (1,3)$ and $(2,3)$. It can be verified that *every permutation in $\mathfrak{S}_m$ is a transposition or a product of transpositions.* The decomposition of permutations into transpositions is not unique.

For example, we can write

$$\begin{aligned}(1\,2\,3\,4) &= (1\,4)(1\,3)(1\,2)\\ &= (1\,2)(2\,3)(3\,4)\\ &= (1\,2)(1\,4)(2\,3)(1\,4)(3\,4).\end{aligned}$$

In general, it can be proved that the number of transpositions needed is necessarily either even or odd, depending *only* on the given permutation. So, a permutation (in $\mathfrak{S}_m$) which can be expressed as the product of an even number of transpositions is called an **even permutation**; the others are **odd permutations**. Since

$$(a_1\ a_2\ \dots\ a_k) = (a_1\ a_k)(a_1\ a_{k-1})\cdots(a_1\ a_3)(a_1\ a_2)$$

a cyclic permutation is even precisely when its length is odd.

$\diamond$ **Exercise 9.** Show that the set of all even permutations in $\mathfrak{S}_m$ forms a subgroup of $\mathfrak{S}_m$ for each $m \geq 2$. (This subgroup is called the **alternating group** on m elements and is denoted by $\mathfrak{A}_m$.)

$\diamond$ **Exercise 10.** If α, β are elements of $\mathfrak{S}_m$, check that $\alpha\beta\alpha^{-1}\beta^{-1}$ always lies in $\mathfrak{A}_m$, and that $\alpha\beta\alpha^{-1}$ belongs to $\mathfrak{A}_m$ whenever β is an even permutation. Work out these elements when $m = 4, \alpha = (2\,1\,4\,3)$ and $\beta = (4\,2\,3)$.

$\diamond$ **Exercise 11.** Show that the symmetry group of a rectangle which is not a square has four elements. By labeling the vertices $1, 2, 3, 4$, represent the symmetry group as a group of permutations on four elements. (This is the so-called **Klein four-group** V_4.)

1.3. **Morphisms of groups.** The question of deciding when we should regard two groups as being "the same" group is an important one. What we need is a formal way of identifying groups that have identical structures; the identification of two groups is given by a special mapping called an *isomorphism.*

Definition 8. Let $\mathsf{G}, \overline{\mathsf{G}}$ be groups. A mapping $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is an **isomorphism** if

(I1) Φ is a bijection;
(I2) $\Phi(gh) = \Phi(g)\Phi(h)$ for all $g, h \in \mathsf{G}$.

If such a Φ exists, we say that G and $\overline{\mathsf{G}}$ are *isomorphic* groups and we write $\mathsf{G} \cong \overline{\mathsf{G}}$.

Example 9. Consider the *exponential function* $\exp : \mathbb{R} \to \mathbb{R}^+, \quad x \mapsto e^x$ (here $\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$). It is known that this function is a bijection (from $\mathbb{R}$ onto $\mathbb{R}^+$). The crucial property $e^{x+y} = e^x e^y$ of $\exp$ is exactly the condition (I2) that is needed to show that $\exp$ is an isomorphism; thus the (additive) group $(\mathbb{R}, +)$ is isomorphic to the (multiplicative) group $(\mathbb{R}^+, \cdot)$.

◇ **Exercise 12.** Show that the (additive) group $\mathbb{R}$ is *not* isomorphic to the (multiplicative) group $\mathbb{R} \setminus \{0\}$.

◇ **Exercise 13.** Let $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ be an isomorphism. Show that

(a) if G is Abelian, then so is $\overline{\mathsf{G}}$;
(b) $\Phi(\mathbf{1}_{\mathsf{G}}) = \mathbf{1}_{\overline{\mathsf{G}}}$;
(c) $\Phi(g^{-1}) = \Phi(g)^{-1}$ for all $g \in \mathsf{G}$.

◇ **Exercise 14.** Show that if $\Phi_1 : \mathsf{G}_1 \to \mathsf{G}_2$ is an isomorphism, then so is $\Phi_1^{-1} : \mathsf{G}_2 \to \mathsf{G}_1$. If, in addition, $\Phi_2 : \mathsf{G}_2 \to \mathsf{G}_3$ is an isomorphism, then so is $\Phi_2 \circ \Phi_1 : \mathsf{G}_1 \to \mathsf{G}_3$.

The first *representation theorem* (for groups) was proved by A. Cayley (1821–1895) in 1878; it tells us that any group can be represented as (isomorphic to) something reasonably concrete: a group of permutations. In other words, the study of subgroups of symmetric groups is no less general than the study of all groups.

Theorem 5. (Cayley's Theorem) *Every group* G *is isomorphic to a permutation group on* G *(i.e., a subgroup of* $\mathfrak{S}_{\mathsf{G}}$*).*

Proof. Each element $a \in \mathsf{G}$ gives a permutation $L_a : \mathsf{G} \to \mathsf{G}$ defined by $L_a(g) := ag$. (L_a is *injective* because if $L_a(g_1) = L_a(g_2)$, then $ag_1 = ag_2$ giving $g_1 = a^{-1}ag_1 = a^{-1}ag_2 = g_2$. It is also *surjective* since if $h \in \mathsf{G}$, then $L_a(a^{-1}h) = aa^{-1}h = h$.) We call L_a the *left translation* by a. Let

$$\overline{\mathsf{G}} := \{L_a \,:\, a \in \mathsf{G}\} \subseteq \mathfrak{S}_{\mathsf{G}}.$$

We have

$$L_a(L_b(g)) = L_a(bg) = a(bg) = (ab)g = L_{ab}(g)$$

for all $g \in \mathsf{G}$. Therefore the product of two elements of $\overline{\mathsf{G}}$ lies in $\overline{\mathsf{G}}$. The identity element $\mathbf{1}_{\mathsf{G}}$ of $\mathfrak{S}_{\mathsf{G}}$ belongs to $\overline{\mathsf{G}}$, and the inverse of L_a in $\mathfrak{S}_{\mathsf{G}}$ is $L_{a^{-1}}$, which is also in $\overline{\mathsf{G}}$. This shows that $\overline{\mathsf{G}}$ is a *subgroup* of $\mathfrak{S}_{\mathsf{G}}$.

The correspondence

$$\Phi : \mathsf{G} \to \overline{\mathsf{G}}, \quad a \mapsto L_a$$

is certainly surjective, and it sends the multiplication of G to that of $\overline{\mathsf{G}}$ because $ab \mapsto L_{ab} = L_aL_b$. It is injective since if $L_a = L_b$, then $a = L_a(\mathbf{1}) = L_b(\mathbf{1}) = b$. Therefore, we have constructed an isomorphism between G and the subgroup $\overline{\mathsf{G}}$ of $\mathfrak{S}_{\mathsf{G}}$. □

Corollary 6. *Every finite group of order* m *is isomorphic to a subgroup of* $\mathfrak{S}_m$.

Proof. If the elements of G are labeled $1, 2, \ldots, m$ in some way, then each permutation of G induces a permutation of $[m] = \{1, 2, \ldots, m\}$. This gives an isomorphism from $\mathfrak{S}_{\mathsf{G}}$ to $\mathfrak{S}_m$, and the subgroup $\overline{\mathsf{G}} \leq \mathfrak{S}_{\mathsf{G}}$ is therefore isomorphic to a subgroup G' of $\mathfrak{S}_m$. As G is isomorphic to $\overline{\mathsf{G}}$, and the composition of two isomorphisms is an isomorphism (this is **Exercise 14**), G is isomorphic to G'. □

Note: Despite its simplicity, Cayley's Theorem has an important meaning for group theory. It shows the existence of a sort of "universal object" – the family $(\mathfrak{S}_n)_{n \in \mathbb{N}}$ of symmetric groups – in which all finite groups (considered up to isomorphism) live. The phrase "up to isomorphism" is typical, not only of group theory, but of all mathematics, which tends to consider at once all objects having common properties.

If $\mathsf{G} = \overline{\mathsf{G}}$ in the definition of an isomorphism, we have the concept of an isomorphism $\Phi : \mathsf{G} \to \mathsf{G}$ of a group G to itself. Such an isomorphism is called an **automorphism** of G. For example, the identity mapping $\mathrm{id}_{\mathsf{G}} = \mathbf{1}_{\mathsf{G}} : \mathsf{G} \to \mathsf{G}$ (*not* to be confused with the identity element $\mathbf{1}_{\mathsf{G}} = \mathbf{1}$ of G) is an automorphism. In general, a group G also has non-trivial automorphisms. It is easy to see that *the set* $\mathsf{Aut}\,(\mathsf{G})$ *of all automorphisms of a group* G *forms a group*, in fact, a subgroup of the symmetric group $\mathfrak{S}_{\mathsf{G}}$.

The group of automorphisms $\mathsf{Aut}\,(\mathsf{G})$ of a group G contains a very special subgroup, which is denoted by $\mathsf{Inn}\,(\mathsf{G})$ and is called the **group of inner automorphisms**. Its elements are the mappings

$$\mathcal{I}_a : \mathsf{G} \to \mathsf{G}, \quad g \mapsto aga^{-1}.$$

(The inner automorphism $\mathcal{I}_a$ is also referred to as the *conjugation map*; note that $\mathcal{I}_a = L_a \circ R_{a^{-1}}$, where $R_{a^{-1}} : g \mapsto ga^{-1}$ is the *right translation* by a^{-1}.)

◇ **Exercise 15.** Verify that the set $\mathsf{Inn}\,(\mathsf{G}) := \{\mathcal{I}_a \,:\, a \in \mathsf{G}\}$ is a subgroup of $\mathsf{Aut}\,(\mathsf{G})$.

The mapping

$$\mathcal{I} : \mathsf{G} \to \mathsf{Inn}\,(\mathsf{G}), \quad a \mapsto \mathcal{I}_a$$

satisfies property (I2) in the definition of an isomorphism: $\mathcal{I}(ab) = \mathcal{I}(a) \circ \mathcal{I}(b)$. However, property (I1) is not necessarily satisfied. For example, if G is an Abelian group, then $aga^{-1} = g$ for all $a, g \in \mathsf{G}$; that is, $\mathcal{I}_a = \mathbf{1}_{\mathsf{G}}$ for all $a \in \mathsf{G}$ and so $\mathsf{Inn}\,(\mathsf{G})$ only consists of the identity element $\mathbf{1}_{\mathsf{G}}$.

Note: One way to study a relatively large and complicated group is to study its smaller and less complicated subgroups. But it would also be useful to be able to study the group as a whole. Homomorphisms, which are more general than isomorphisms, can help to do just that. A homomorphism is a mapping from one group to another that preserves the group operation but is not necessarily one-to-one. Thus the image of a homomorphism can be smaller than the domain, but it will generally reflect some essential features of the domain. Even more importantly, subgroups and images of homomorphisms can be used together to show that most groups are built up from smaller component groups. The concept of "homomorphism" also extends to other algebraic structures (for instance, to rings, fields, modules and algebras); it is unquestionably one of the most important concepts in algebra.

Definition 10. A mapping $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ from the group G to the group $\overline{\mathsf{G}}$ is called a **homomorphism** if

$$\Phi(gh) = \Phi(g)\Phi(h) \qquad \text{for all } g, h \in \mathsf{G}.$$

Various properties of isomorphisms were checked (**Exercise 13**). Those arguments which do not use the fact that an isomorphism is a bijection are equally valid here. Therefore if $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is a homomorphism, then

(H1) $\Phi(\mathbf{1}_\mathsf{G}) = \mathbf{1}_{\overline{\mathsf{G}}}$.
(H2) $\Phi(g^{-1}) = \Phi(g)^{-1}$ for all $g \in \mathsf{G}$.
(H3) The image $\mathsf{Im}\,\Phi := \{\Phi(g) \,:\, g \in \mathsf{G}\}$ of G is a subgroup of $\overline{\mathsf{G}}$.

Definition 11. The **kernel** of the homomorphism $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is the set

$$\mathsf{Ker}\,\Phi := \{g \in \mathsf{G} \,:\, \Phi(g) = \mathbf{1}_{\overline{\mathsf{G}}}\}.$$

Theorem 7. *If $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is a homomorphism, then its kernel $\mathsf{Ker}\,\Phi$ is a subgroup of G. Moreover, Φ is one-to-one $\underline{\text{if and only if}}$ $\mathsf{Ker}\,\Phi = \{\mathbf{1}_\mathsf{G}\}$.*

Proof. Let $a, b \in \mathsf{Ker}\,\Phi$. Then we have

$$\Phi(a^{-1}b) = \Phi(a^{-1})\Phi(b) = \Phi(a)^{-1}\Phi(b) = \mathbf{1}_{\overline{\mathsf{G}}}\,\mathbf{1}_{\overline{\mathsf{G}}} = \mathbf{1}_{\overline{\mathsf{G}}}$$

hence $a^{-1}b \in \mathsf{Ker}\,\Phi$. This shows that $\mathsf{Ker}\,\Phi$ is a subgroup of G.

Since $\mathbf{1}_\mathsf{G} \in \mathsf{Ker}\,\Phi$, it is clear that if Φ is one-to-one, then $\mathsf{Ker}\,\Phi = \{\mathbf{1}_\mathsf{G}\}$. Why ?

Assume, on the other hand, that $\mathsf{Ker}\,\Phi = \{\mathbf{1}_\mathsf{G}\}$. If $a, b \in \mathsf{G}$ and $\Phi(a) = \Phi(b)$, then we have

$$\mathbf{1}_{\overline{\mathsf{G}}} = \Phi(b)^{-1}\Phi(b) = \Phi(a)^{-1}\Phi(b) = \Phi(a^{-1}b)$$

hence $a^{-1}b \in \mathsf{Ker}\,\Phi$ and so $a = b$. This proves that Φ is one-to-one. □

Let $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ be a homomorphism. In general, Φ is neither injective nor surjective. We can make Φ into a surjective mapping by replacing $\overline{\mathsf{G}}$ by $\mathsf{Im}\,\mathsf{G}$ (which is a subgroup of $\overline{\mathsf{G}}$). So the "main" difference between a homomorphism and a isomorphism is the presence of a non-trivial kernel $\mathsf{Ker}\,\Phi$ (which is, one might say, a measure of non-injectivity of Φ). If $\mathsf{Ker}\,\Phi = \{\mathbf{1}_\mathsf{G}\}$, then $\Phi : \mathsf{G} \to \mathsf{Im}\,\Phi$ is an isomorphism.

Let $\mathsf{H} = \mathsf{Ker}\,\Phi \leq \mathsf{G}$. We have (for $h \in \mathsf{H}$, $g \in \mathsf{G}$)

$$\Phi(ghg^{-1}) = \Phi(g)\,\Phi(h)\,\Phi(g)^{-1} = \Phi(g)\,\mathbf{1}_{\overline{\mathsf{G}}}\,\Phi(g)^{-1} = \mathbf{1}_{\overline{\mathsf{G}}}$$

i.e., $ghg^{-1} \in \mathsf{H}$; hence, $g\,\mathsf{H}\,g^{-1} \subseteq \mathsf{H}$. If we replace g by g^{-1} here, we obtain $g^{-1}\,\mathsf{H}\,g \subseteq \mathsf{H}$ so that $\mathsf{H} \subseteq g\,\mathsf{H}\,g^{-1}$. Thus

$$g\,\mathsf{H}\,g^{-1} = \mathsf{H} \quad \text{for all } g \in \mathsf{G}. \tag{3}$$

A subgroup which has this property is called a **normal subgroup** (or invariant subgroup); if H is a normal subgroup of G, we write $\mathsf{H} \trianglelefteq \mathsf{G}$. We have thereby proved

Theorem 8. *The kernel of a homomorphism is always a normal subgroup.*

It is a remarkable fact that the converse of this theorem holds; that is, not only is the kernel of a homomorphism a normal subgroup, but *every normal subgroup is the kernel of a homomorphism* (this is **Exercise 18**).

◇ **Exercise 16.** Let H be a subgroup of G. Show that the following statements are logically equivalent.

(a) $g\,\mathsf{H}\,g^{-1} = \mathsf{H}$ for all $g \in \mathsf{G}$.
(b) $g\,\mathsf{H} = \mathsf{H}\,g$ for all $g \in \mathsf{G}$.
(c) $g\,\mathsf{H}\,g^{-1} \subseteq \mathsf{H}$ for all $g \in \mathsf{G}$.

Note: The terms "surjective map" (map onto), "injective map" (one-to-one map or imbedding) and "bijective map" (one-to-one correspondence) which can be used for maps between any sets (with or without any structure) are often replaced by other terms when used for groups (the same happens for other mathematical structures). We use the terms *epimorphism* (homomorphism onto), *monomorphism* (homomorphism whose kernel is the identity element) and *isomorphism* (homomorphism which is both an epimorphism and a monomorphism). There is a tendency to replace the term *homomorphism* with the word *morphism*.

We now give some further examples of group homomorphisms.

Example 12. The function

$$\Phi : x \mapsto e^{2\pi i x}$$

is a homomorphism from the (additive) group $\mathbb{R}$ to the (additive) group $\mathbb{C}$. This homomorphism is neither injective nor surjective. It is very easy to see that the kernel is $\mathbb{Z}$ (the group of integers) and the image group is the *circle* $\mathbb{S}^1 := \{z \in \mathbb{C} \,:\, |z| = 1\} \leq \mathbb{C}$.

Example 13. The *determinant* function

$$\Phi : \mathsf{GL}\,(n, \Bbbk) \to \Bbbk \setminus \{0\}, \quad a \mapsto \Phi(a) := \det a$$

is a homomorphism from the general linear group $\mathsf{GL}\,(n, \Bbbk)$ to the (multiplicative) group $\mathbb{R} \setminus \{0\}$. By definition, we have that $\mathsf{SL}\,(n, \Bbbk) = \mathsf{Ker}\,\Phi$.

Example 14. Let π be a permutation in $\mathfrak{S}_n$, and let $\pi = \tau_1\tau_2\cdots\tau_k$ be any decomposition of π into a product of transpositions. Then the number $\epsilon(\pi) := (-1)^k$ is completely determined by π and does not depend on which decomposition is used. $\epsilon(\pi)$ is called the **signature** of π.

The function

$$\operatorname{sgn} : \mathfrak{S}_n \to \{-1, 1\}, \quad \pi \mapsto \epsilon(\pi)$$

is a (surjective) homomorphism from the symmetric group $\mathfrak{S}_n$ to the (multiplicative) group $\{-1, 1\} \leq \mathbb{R} \setminus \{0\}$. Clearly, the kernel of (the signature epimorphism) sgn is the alternating group $\mathfrak{A}_n$.

Example 15. Consider the function $\Phi : \mathfrak{S}_3 \to \mathsf{GL}\,(3, \mathbb{R})$ defined as follows:

$$\mathbf{1} \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1\,2) \mapsto \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1\,3) \mapsto \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$
$$(2\,3) \mapsto \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad (1\,2\,3) \mapsto \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad (1\,3\,2) \mapsto \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}.$$

It is easy to check that Φ is a monomorphism, and that for each $\pi \in \mathfrak{S}_3$ the determinant of $\Phi(\pi)$ is ± 1, depending on the *signature* of the permutation π.

In general, *there exists a monomorphism* $\Phi : \mathfrak{S}_n \to \mathsf{GL}\,(n, \mathbb{R})$ *such that the matrix* $\Phi(\pi)$, $\pi \in \mathfrak{S}_n$ *has determinant* $\epsilon(\pi)$. (Matrices of the form $\Phi(\pi)$, $\pi \in \mathfrak{S}_n$ are called **permutation matrices**.) The restriction of the monomorphism Φ to the (alternating) group $\mathfrak{A}_n$ is a monomorphism into $\mathsf{SL}\,(n, \mathbb{R})$. Given any finite group G, the composition $\Phi \circ L$ of the map $L : \mathsf{G} \to \mathfrak{S}_n$ (see Cayley's Theorem) and $\Phi : \mathfrak{S}_n \to \mathsf{GL}\,(n, \mathbb{R})$ gives a monomorphism $\mathsf{G} \to \mathsf{GL}\,(n, \mathbb{R})$.

Example 16. Let denote by μ_{abcd} the Möbius transformation $z \mapsto \frac{az+b}{cz+d}$. The function

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \mu_{abcd}$$

is a (surjective) homomorphism from the general linear group $\mathsf{GL}\,(2, \mathbb{C})$ to the Möbius group $\mathsf{M\ddot{o}b}$. It is not hard to compute the kernel, which turns out to be the set (subgroup) of *scalar matrices*: $\{\lambda \mathbf{1} \,:\, \lambda \in \mathbb{C} \setminus \{0\}\} \leq \mathsf{GL}\,(2, \mathbb{C})$ (this group is isomorphic to $\mathbb{C} \setminus \{0\}$).

1.4. **Cosets and quotient groups.** Let H be a subgroup of G.

Definition 17. A **left coset** of H in G is a set of the form $g\,\mathsf{H} := \{gh \,:\, h \in \mathsf{H}\}$. (The element g is called a **coset representative** for $g\,\mathsf{H}$.)

Similarly, we define a **right coset** $\mathsf{H}\,g$.

◇ **Exercise 17.** Let H be a subgroup of G and let $g \in \mathsf{G}$. Show that the following statements are logically equivalent.

(a) $g \in \mathsf{H}$.
(b) $g\,\mathsf{H} = \mathsf{H}$.
(c) $\mathsf{H}\,g = \mathsf{H}$.

If $\mathsf{H} = \mathsf{Ker}\,\Phi$ is the kernel of a homomorphism, then $g\,\mathsf{H} = \mathsf{H}\,g$ because H is normal in G (see Theorem 7 and **Exercise 16**). Note that *the subgroup* H *itself is a coset*: $\mathsf{H} = \mathbf{1}_{\mathsf{G}}\,\mathsf{H} = \mathsf{H}\,\mathbf{1}_{\mathsf{G}}$. However, none of the other cosets can be a (proper) subgroup because, if $g\,\mathsf{H}$ were a subgroup, then we would have $\mathbf{1}_{\mathsf{G}} \in g\,\mathsf{H}$, so that $\mathbf{1}_{\mathsf{G}} = gh$, $g = h^{-1}$ and hence $g\,\mathsf{H} = h^{-1}\,\mathsf{H} = \mathsf{H}$.

Theorem 9. *Let* H *be a subgroup of* G. *Then* G *is the union of its left cosets, and any two left cosets are either equal or disjoint. Further, the left cosets* $a\,\mathsf{H}$ *and* $b\,\mathsf{H}$ *are equal* *if and only if* $a^{-1}b \in \mathsf{H}$. *(Similar statements are true for right cosets.)*

Proof. Since each element $g \in \mathsf{G}$ is contained in the coset $g\,\mathsf{H}$, the set G is a union of left cosets of H: $\mathsf{G} = \bigcup g_i\,\mathsf{H}$. Suppose that two left cosets $a\,\mathsf{H}$ and $b\,\mathsf{H}$ have an element in common: $g = ah = bh'$. Then $b = ahh'^{-1}$, and any element bh'' of the coset $b\,\mathsf{H}$ has the form $a(hh'^{-1}h'')$. Thus $b\,\mathsf{H} \subseteq a\,\mathsf{H}$. We similarly prove that every element of $a\,\mathsf{H}$ is contained in $b\,\mathsf{H}$. Hence $a\,\mathsf{H} = b\,\mathsf{H}$.

Let $a, b \in \mathsf{G}$ such that $a^{-1}b \in \mathsf{H}$. Then $b \in a\,\mathsf{H}$ and thus $b\,\mathsf{H} \subseteq a\,\mathsf{H}$. We similarly show that $a\,\mathsf{H} \subseteq b\,\mathsf{H}$. Hence $a\,\mathsf{H} = b\,\mathsf{H}$. Conversely, assume that $a\,\mathsf{H} = b\,\mathsf{H}$. Then $ah = bh'$ implies $a^{-1}b = hh'^{-1} \in \mathsf{H}$. This concludes the argument. □

Corollary 10. *The partition of* G *into left cosets of* H *gives an equivalence relation on* G.

Note: Given a set M, a (binary) *relation* $\sim$ on M is called an **equivalence relation** if the following conditions hold for all $a, b, c \in \mathsf{M}$:

(ER1) $a \sim a$ (reflexivity);
(ER2) if $a \sim b$, then $b \sim a$ (symmetry);
(ER3) if $a \sim b$ and $b \sim c$, then $a \sim c$ (transitivity).

The subset $[a] := \{x \in \mathsf{M} \,:\, a \sim x\} \subseteq \mathsf{M}$ (of all elements equivalent to a given element a) is called the **equivalence class** containing a. Clearly, $a \in [a]$. The set of equivalence classes of $\sim$ form a partition of M. (A *partition* of a given set is a collection of non-empty, mutually disjoint subsets such that their union is the whole set.) Conversely, any partition $(C_i)_{i\in I}$ of M defines an equivalence relation $\sim$ on M by $a \sim b$ if and only if $a, b \in C_i$ (for some $i \in I$). If $a \in C_i$, then $[a] = C_i$ and so the sets C_i are precisely the equivalence classes for $\sim$. In particular, the partition of the group G into left cosets of its subgroup H induces an equivalence relation $\sim$ on G, defined by

$$a \sim b \quad \iff \quad a^{-1}b \in \mathsf{H}.$$

(Condition $a^{-1}b \in \mathsf{H}$ is logically equivalent to condition $a\,\mathsf{H} = b\,\mathsf{H}$.)

Normal subgroups are important because their (left) cosets form a group in a natural way.

Theorem 11. *If* H *is a normal subgroup of* G, *then the set of all left cosets of* H *in* G *forms a group.*

Proof. The product of two left cosets is again a left coset because

$$(a\,\mathsf{H})(b\,\mathsf{H}) = (ab)\,\mathsf{H} \tag{4}$$

for any two elements $a, b \in \mathsf{G}$. Accepting this for a moment, the coset $\mathbf{1}_{\mathsf{G}}\,\mathsf{H} = \mathsf{H}$ acts as an identity, and $(a^{-1})\,\mathsf{H}$ is the inverse of $a\,\mathsf{H}$ for each $a \in \mathsf{G}$. So we do indeed have a group.

Just why does (4) hold and what does it have to do with the hypothesis that H be a *normal* subgroup of G ? Each element of $(a\,\mathsf{H})(b\,\mathsf{H})$ has the form $ahbh'$ for some $h, h' \in \mathsf{H}$. Rewrite this as

$$ab\left(b^{-1}hb\right)h'$$

and notice that $b^{-1}hb \in \mathsf{H}$ precisely because H is a normal subgroup of G. Why ? Hence $b^{-1}hb = h''$ for some $h'' \in \mathsf{H}$, giving

$$ahbh' = ab\left(b^{-1}hb\right)h' = ab(h''h') \in (ab)\,\mathsf{H}.$$

Thus we have $(a\,\mathsf{H})(b\,\mathsf{H}) \subseteq (ab)\,\mathsf{H}$. The reverse inclusion is easier to check (and works for any subgroup H). Each element of $(ab)\,\mathsf{H}$ has the form abh for some $h \in \mathsf{H}$. Rewriting this as $(a\mathbf{1}_{\mathsf{G}})(bh)$ shows that it belongs to $(a\,\mathsf{H})(b\,\mathsf{H})$, and we deduce $(ab)\,\mathsf{H} \subseteq (a\,\mathsf{H})(b\,\mathsf{H})$. This completes the argument. □

The group of left cosets of H in G introduced above is called the **quotient group** (or factor group) of G by H and denoted by G/H. (Recall that the left cosets of H in G form a partition of G. Each of these cosets represents a *single element* in G/H and it is, in this sense, that we have "divided G by H".)

◇ **Exercise 18.** Show that if $\mathsf{H} \trianglelefteq \mathsf{G}$, then the mapping

$$\mathsf{G} \to \mathsf{G}/\mathsf{H}, \quad g \mapsto g\,\mathsf{H}$$

is a surjective homomorphism, and its kernel is H. (This homomorphism is called the **natural homomorphism** of G onto G/H.)

The natural homomorphism $\mathsf{G} \to \mathsf{G}/\mathsf{H}$ shows that each quotient group of a group G is a homomorphic image of G. The next theorem shows that the converse is also true: each homomorphic image of G is (isomorphic to) a quotient group of G.

Theorem 12. (First Isomorphism Theorem) *Let* $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ *be a homomorphism with* $\mathsf{Ker}\,\Phi = \mathsf{H}$. *Then the mapping*

$$\widehat{\Phi} : \mathsf{G}/\mathsf{H} \to \mathsf{Im}\,\Phi, \quad g\,\mathsf{H} \mapsto \Phi(g)$$

is an isomorphism. (Therefore $\mathsf{G}/\mathsf{H} \cong \mathsf{Im}\,\Phi$.*)*

Proof. If two cosets $a\,\mathsf{H}, b\,\mathsf{H} \in \mathsf{G}/\mathsf{H}$ are equal, then $a^{-1}b \in \mathsf{H}$. Applying Φ gives

$$\Phi(a^{-1}b) = \Phi(a)^{-1}\Phi(b) = \mathbf{1}_{\overline{\mathsf{G}}}$$

and therefore $\Phi(a) = \Phi(b)$. This means that the function $\widehat{\Phi} : g\,\mathsf{H} \mapsto \Phi(g)$ is *well defined.* Reversing the above computation shows that if $\Phi(a) = \Phi(b)$, then $a\,\mathsf{H} = b\,\mathsf{H}$. So $\widehat{\Phi}$ is injective. The function $\widehat{\Phi}$ is a homomorphism because

$$\widehat{\Phi}\,((a\,\mathsf{H})(b\,\mathsf{H})) = \widehat{\Phi}\,((ab)\,\mathsf{H}) = \Phi(ab) = \Phi(a)\Phi(b) = \widehat{\Phi}(a\,\mathsf{H})\widehat{\Phi}(b\,\mathsf{H})$$

for any cosets $a\,\mathsf{H}, b\,\mathsf{H} \in \mathsf{G}/\mathsf{H}$. Finally, the image of $\widehat{\Phi}$ is the same as the image of Φ. We have proved that $\widehat{\Phi}$ is an isomorphism from G/H to the image of Φ. □

NOTE: Let π denote the natural projection/homomorphism of G onto $\mathsf{G}/\mathsf{Ker}\,\Phi$, and ι denote the natural inclusion of $\mathsf{Im}\,\Phi$ into $\overline{\mathsf{G}}$. Schematically, the two ways (Φ and $\iota \circ \widehat{\Phi} \circ \pi$) of getting from G to $\overline{\mathsf{G}}$ give the same result for every element of G. This is described by saying that the following diagram commutes:

$$\begin{tikzcd}
\mathsf{G} \arrow[r, "\Phi"] \arrow[d, "\pi"'] & \overline{\mathsf{G}} \\
\mathsf{G}/\mathsf{Ker}\,\Phi \arrow[r, "\widehat{\Phi}"] & \mathsf{Im}\,\Phi \arrow[u, "\iota"']
\end{tikzcd}$$

Two special cases are particularly useful.

COROLLARY 13. *If $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is an epimorphism, then $\mathsf{G}/\mathsf{Ker}\,\Phi$ is isomorphic to $\overline{\mathsf{G}}$.*

COROLLARY 14. *Suppose $\Phi : \mathsf{G} \to \overline{\mathsf{G}}$ is an epimorphism. Then Φ is an isomorphism if and only if its kernel $\mathsf{Ker}\,\Phi$ consists just of the identity element of G.*

PROBLEMS (1–5)

(1) Let M be a *finite* set and let $\alpha : \mathsf{M} \to \mathsf{M}$. Prove that the following statements are logically equivalent.
  (a) α is injective.
  (b) α is surjective.
  (c) α is bijective (i.e., a permutation).

(2) Let $\alpha, \beta, \gamma \in \mathsf{M}^{\mathsf{M}}$ such that

$$\beta\alpha = \gamma\alpha = \mathbf{1}_{\mathsf{M}} \quad \text{and} \quad \alpha\beta = \alpha\gamma = \mathbf{1}_{\mathsf{M}}.$$

Deduce that $\beta = \gamma$. (This proves that each invertible mapping has a unique inverse.)

(3) Assume that α and β are disjoint cycles representing elements of $\mathfrak{S}_m$, say $\alpha = (a_1\ \ldots\ a_r)$ and $\beta = (b_1\ \ldots\ b_s)$ with $a_i \neq b_j$ for all $i, j \in [m]$.

(a) Compute $(\alpha\beta)(a_i)$ and $(\beta\alpha)(a_i)$ for $i \in [r]$.

(b) Compute $(\alpha\beta)(b_j)$ and $(\beta\alpha)(b_j)$ for $j \in [s]$.

(c) Compute $(\alpha\beta)(k)$ and $(\beta\alpha)(k)$ for $k \in [m]$ with $k \neq a_i$ and $k \neq b_j$ for all $i, j \in [m]$.

(d) What do parts (a),(b) and (c), taken together, prove about the relationship between $\alpha\beta$ and $\beta\alpha$?

(4) (Second Isomorphism Theorem) Let H, K be subgroups of G with K normal in G. Then HK is a subgroup of G, $\mathsf{H} \cap \mathsf{K}$ is a normal subgroup of H, and the quotient groups HK/K and $\mathsf{H}/\mathsf{H} \cap \mathsf{K}$ are isomorphic.

(5) Let G be a group.

(a) Given $x \in \mathsf{G}$, the element gxg^{-1} $(= \mathcal{I}_g(x))$ is known as the **conjugate** of x by g. The set of all conjugates of x, that is $C(x) := \{gxg^{-1} : g \in \mathsf{G}\}$ is known as the **conjugacy class** of x (in G). Show that the collection of all conjugacy classes in G constitutes a partition of G.

(b) Given $x \in \mathsf{G}$, define the **centralizer** of x (in G) by

$$\mathsf{Z}\,(x) := \left\{g \in \mathsf{G} : gxg^{-1} = x\right\}.$$

(Thus the centralizer of x consists of all elements that commute with x.)

(i) Show that, for each $x \in \mathsf{G}$, the centralizer $\mathsf{Z}\,(x)$ is a subgroup of G.

(ii) Show that conjugates gxg^{-1} and hxh^{-1} of x are equal if and only if $g\,\mathsf{Z}\,(x) = h\,\mathsf{Z}\,(x)$. (This means that there is a a one-to-one correspondence between the conjugacy classes of x in G and the set of left cosets of the centralizer of x.)

(c) The **center** of G consists of all those elements which commute with every element of G. It is usually denoted by $\mathsf{Z}\,(\mathsf{G})$ so that

$$\mathsf{Z}\,(\mathsf{G}) := \{g \in \mathsf{G} : gx = xg \text{ for all } x \in \mathsf{G}\}.$$

(i) Show that $\mathsf{Z}(\mathsf{G})$ is an Abelian subgroup of G, and is made up of the conjugacy classes which contain just one element.

(ii) Show that the group $\mathsf{Inn}\,(\mathsf{G})$ of inner automorphisms of G is isomorphic to the quotient group $\mathsf{G}/\mathsf{Z}(\mathsf{G})$.

2. Actions of Groups on Sets

Group actions • Orbits and stabilizers • Particular G*-sets • Examples of group actions.*

2.1. **Group actions.** A valuable technique in studying a group is to represent it in terms of something familiar and concrete: if the elements happen to be permutations or matrices, however, we may be able to obtain results by using this extra information.

Note: We began with (examples of) transformation groups, i.e., subgroups of the symmetric group $\mathfrak{S}_{\mathsf{M}}$ on a set M. This approach is consistent both with the historical path along which group theory developed and with the importance of transformation groups in other areas of mathematics. The so-called abstract theory of groups, which arose in a later era – the first half of the 20th century – has gone far beyond transformation groups, but many of the concepts of this theory bear the imprint of earlier times. In fact, the most common source of these concepts is the idea of a *realization* (or *representation*) of a given group G in $\mathfrak{S}_{\mathsf{M}}$, where M is some suitably chosen set.

Let G be a group and M a (non-empty) set.

Definition 18. A (left) **action** of G on M is a function $\theta : \mathsf{G} \times \mathsf{M} \to \mathsf{M}$ such that

(LA1)′ $\theta(g_2, \theta(g_1, x)) = \theta(g_2 g_1, x)$ for all $g_1, g_2 \in \mathsf{G}$ and $x \in \mathsf{M}$;
(LA2)′ $\theta(\mathbf{1}, x) = x$ for all $x \in \mathsf{M}$.

We write $g \cdot x$ in place of the pedantic notation $\theta(g, x)$. We can now write the above conditions (axioms) as follows:

(LA1) $g_2 \cdot (g_1 \cdot x) = (g_2 g_1) \cdot x$ for all $g_1, g_2 \in \mathsf{G}$ and $x \in \mathsf{M}$;
(LA2) $\mathbf{1} \cdot x = x$ for all $x \in \mathsf{M}$.

The set M is called a (left) **G-set**. One also says that the group G **acts** on the set M.

Note: In this definition, the elements of G act from the left. There is a "right" version of G-sets that is sometimes convenient. Define a **right action** of G on M to be a function $\tau : \mathsf{M} \times \mathsf{G} \to \mathsf{M}$, $(x, g) \mapsto x \cdot g$ such that

(RA1) $(x \cdot g_1) \cdot g_2 = x \cdot (g_1 g_2)$ for all $g_1, g_2 \in \mathsf{G}$ and $x \in \mathsf{M}$;
(RA2) $x \cdot \mathbf{1} = x$ for all $x \in \mathsf{M}$.

It is easy to see that *every right action* $\tau : \mathsf{M} \times \mathsf{G} \to \mathsf{M}$ *gives rise to a left action* $\theta : \mathsf{G} \times \mathsf{M} \to \mathsf{M}$ *if one defines* $\theta(g, x) := x \cdot g^{-1}$.

Any action of G on M induces an action of G on (the Cartesian product) $\mathsf{M}^k = \mathsf{M} \times \cdots \times \mathsf{M}$ (k factors) by the obvious rule:

$$g \cdot (x_1, \dots, x_k) := (g \cdot x_1, \dots, g \cdot x_k)\,.$$

◇ **Exercise 19.** Let G be a group and let $\theta_1 : \mathsf{G} \times \mathsf{M}_1 \to \mathsf{M}_1$ and $\theta_2 : \mathsf{G} \times \mathsf{M}_2 \to \mathsf{M}_2$ be actions of G on the sets M_1 and M_2, respectively. Define

$$\mathsf{M}_1 + \mathsf{M}_2 := (\mathsf{M}_1 \times \{1\}) \cup (\mathsf{M}_2 \times \{2\})$$

and $\theta^{\vee} : \mathsf{G} \times (\mathsf{M}_1 + \mathsf{M}_2) \to \mathsf{M}_1 + \mathsf{M}_2$ by

$$(g, (x, i)) \mapsto (g \cdot x, i)$$

for $i = 1, 2,\ x \in \mathsf{M}_i$ and $g \in \mathsf{G}$. Show that $\theta^{\vee}$ is an action (of G on the *sum* $\mathsf{M}_1 + \mathsf{M}_2$).

◇ **Exercise 20.** Let G be a group and let $\theta_1 : \mathsf{G} \times \mathsf{M}_1 \to \mathsf{M}_1$ and $\theta_2 : \mathsf{G} \times \mathsf{M}_2 \to \mathsf{M}_2$ be actions of G on the sets M_1 and M_2, respectively. Define $\theta^{\wedge} : \mathsf{G} \times (\mathsf{M}_1 \times \mathsf{M}_2) \to \mathsf{M}_1 \times \mathsf{M}_2$ by

$$(g, (x_1, x_2)) \mapsto (g \cdot x_1, g \cdot x_2)$$

for $x_1 \in \mathsf{M}_1, x_2 \in \mathsf{M}_2$ and $g \in \mathsf{G}$. Show that $\theta^{\wedge}$ is an action (of G on the *product* $\mathsf{M}_1 \times \mathsf{M}_2$).

There is also an induced action of G on the set of all subsets $\mathcal{P}(\mathsf{M})$ of M. We set $g \cdot \emptyset := \emptyset$ and, if S is a non-empty subset of M, then we set $g \cdot \mathsf{S} := \{g \cdot x \,:\, x \in \mathsf{S}\}$.

◇ **Exercise 21.** Let M and N be sets, and let θ be an action of the group G on the set M. Consider the set N^{M} of all N-valued functions defined on M. Show that the correspondence

$$(g, F) \mapsto g \cdot F := F \circ \theta_{g^{-1}}$$

for $g \in \mathsf{G}$ and $F : \mathsf{M} \to \mathsf{N}$, defines an action (of G on the function set N^{M}). (For $\mathsf{M} = \mathbb{R}$ this gives the induced action on the function set $\mathfrak{F}_{\mathsf{M}} = \mathbb{R}^{\mathsf{M}}$; for $\mathsf{N} = \{0, 1\}$ this gives the induced action on the the power set $\mathcal{P}(\mathsf{M}) = \{0, 1\}^{\mathsf{M}}$.)

◇ **Exercise 22.** Let L and M be sets, and let τ be an action of the group G on the set M. Consider the set M^{L} of all M-valued functions defined on L. Show that the correspondence

$$(g, C) \mapsto g \cdot C := \tau_g \circ C$$

for $g \in \mathsf{G}$ and $C : \mathsf{L} \to \mathsf{M}$, defines an action (of G on the function set M^{L}). (For $\mathsf{L} = \mathbb{R}$ this gives the induced action on the set $\mathfrak{C}_{\mathsf{M}} = \mathbb{M}^{\mathbb{R}}$ of M-valued parametrised curves.)

Example 19. Let G be a subgroup of the symmetric group $\mathfrak{S}_{\mathsf{M}}$ on M: $\mathsf{G} \leq \mathfrak{S}_{\mathsf{M}}$ (G is a transformation group). Then the function

$$\mathsf{G} \times \mathsf{M} \ni (\alpha, x) \mapsto \alpha(x) \in \mathsf{M}$$

is an action of G on M; this is the most frequent case. For example, G can be defined as a subgroup of $\mathfrak{S}_{\mathsf{M}}$ satisfying certain conditions.

EXAMPLE 20. (The **regular representation**) Given a group G, we can make G into a G-set (i.e., take M to be G) by defining $g \cdot x$ to be the *group product*: the function

$$\mathsf{G} \times \mathsf{G} \ni (g, x) \mapsto gx \in \mathsf{G}$$

is an action of G on itself. The map $L_a : \mathsf{G} \to \mathsf{G}, \quad g \mapsto ag$ is the left translation by a. Our action (by left translations) induces an action of G on the set of subsets of G. In particular, let $\mathsf{H} \leq \mathsf{G}$. It is clear that the function (denoted λ^{H})

$$\mathsf{G} \times \mathsf{G}/\mathsf{H} \ni (g, a\,\mathsf{H}) \mapsto g(a\,\mathsf{H}) := (ga)\,\mathsf{H}$$

is an action of G on the *orbit set* G/H. The corresponding *homomorphism*

$$\Phi^{\mathsf{H}} : \mathsf{G} \to \mathfrak{S}_{\mathsf{G}/\mathsf{H}}, \quad g \mapsto \lambda_g^{\mathsf{H}} \ (: \mathsf{G}/\mathsf{H} \to \mathsf{G}/\mathsf{H})$$

is the so-called (left) **regular representation** of G. (Here λ_g^{H} takes the left coset $a\,\mathsf{H}$ to $(ga)\,\mathsf{H}$.)

◇ **Exercise 23.** Show that the map $\Phi^{\mathsf{H}} : \mathsf{G} \to \mathfrak{S}_{\mathsf{G}/\mathsf{H}}, \quad g \mapsto \lambda_g^{\mathsf{H}}$ is a homomorphism.

The regular representation of the group G by permutations of cosets of the subgroup H in G is much more efficient than the one obtained using CAYLEY'S THEOREM.

EXAMPLE 21. (The **conjugation action**) Another way to make G into a G-set is to use *conjugation*: the function

$$\mathsf{G} \times \mathsf{G} \ni (g, x) \mapsto gxg^{-1} \in \mathsf{G}$$

is also an action of G on itself. Clearly, $\mathbf{1}x = x$ and

$$g_2(g_1x) = g_2\left(g_1xg_1^{-1}\right)g_2^{-1} = (g_2g_1)x(g_2g_1)^{-1} = (g_2g_1)x$$

for all $x \in \mathsf{G}$. The conjugation action carries over to subsets and subgroups of G. Two subsets $\mathsf{S}, \mathsf{T} \subseteq \mathsf{G}$ are **conjugate** if $\mathsf{T} = g\,\mathsf{S}\,g^{-1}$ for some $g \in \mathsf{G}$. Let $\mathsf{H} \leq \mathsf{G}$. It is customary to call (the group) $\mathsf{N}\,(\mathsf{H}) := \{g \in \mathsf{G} \ : \ g\,\mathsf{H}\,g^{-1} = \mathsf{H}\}$ the **normalizer** of H in G. (The subgroup H is normal in G precisely when $\mathsf{N}\,(\mathsf{H}) = \mathsf{H}$.)

NOTE: The first mathematicians who studied group-theoretic problems (e.g., J.L. LAGRANGE) were concerned with the question: What happens to the polynomial $f(X_1, \ldots, X_m)$ if one permutes the variables ? More precisely, if $\pi \in \mathfrak{S}_m$, define

$$\pi \cdot f(X_1, \ldots, X_m) := f(X_{\pi(1)}, \ldots, X_{\pi(m)});$$

given $f \in \mathbb{R}[X_1, \ldots, X_m]$, how many distinct polynomials $\pi \cdot f$ are there ? (Here $\mathbb{R}[X_1, \ldots, X_m]$ denotes the set – in fact, ring – of polynomials in m variables $X_1, \ldots, X_m$ with real coefficients.)

If $\pi \cdot f = f$ for all $\pi \in \mathfrak{S}_m$, then (the polynomial) f is called a **symmetric function**. If a polynomial $f(X) = \sum_{i=0}^m a_iX^i \in \mathbb{R}[X]$ has roots $r_1, \ldots, r_m$, then each of the coefficients a_i of $f(X) = a_m\Pi_{i=0}^m(X - r_i)$ is a symmetric function of $r_1, \ldots, r_m$. Other interesting functions of the roots may not be symmetric. For example, the **discriminant** of $f(X)$ is defined to be the number d^2, where $d := \Pi_{i<j}(r_i - r_j)$. If $D(X_1, \ldots, X_m) = \Pi_{i<j}(X_i - X_j)$, then it is easy to see that $\pi \cdot D = \pm D$ for every $\pi \in \mathfrak{S}_m$. Indeed, D is an **alternating function** of the roots: $\pi \cdot D = D$ if and only if $\pi \in \mathfrak{A}_m$. This suggests a slight change in viewpoint. Given $f(X_1, \ldots, X_m)$, find $\mathcal{S}(f) := \{\pi \in \mathfrak{S}_m : \pi \cdot f = f\}$; this is precisely what Lagrange did. It is easy to see that $\mathcal{S}(f) \leq \mathfrak{S}_m$; moreover, f is symmetric if and only if $\mathcal{S}(f) = \mathfrak{S}_m$, while $\mathcal{S}(D) = \mathfrak{A}_m$.

Modern mathematicians are concerned with the same type of problem. If M is a G-set, then the set of all $\alpha : \mathsf{M} \to \mathsf{M}$ such that $\alpha(g \cdot x) = \alpha(x)$ for all $x \in \mathsf{M}$ and all $g \in \mathsf{G}$ is usually valuable in analyzing M.

Example 22. Let $\Bbbk$ be a *field* (think of either $\mathbb{R}$ or $\mathbb{C}$). The symmetric group $\mathfrak{S}_m$ acts on the set $\mathsf{M} = \Bbbk[X_1, \ldots, X_m]$ by

$$\mathfrak{S}_m \times \mathsf{M} \ni (\pi, f) \mapsto \pi \cdot f$$

where $\pi \cdot f(X_1, \ldots, X_m) = f(X_{\pi(1)}, \ldots, X_{\pi(m)})$.

$\diamond$ **Exercise 24.** For any permutation $\pi \in \mathfrak{S}_m$ and $x = (x_1, \ldots, x_m) \in \mathbb{R}^m$, define

$$\pi \cdot x := \left(x_{\pi^{-1}(1)}, \ldots, x_{\pi^{-1}(m)}\right).$$

Show that $\alpha \cdot (\beta \cdot x) = (\alpha\beta) \cdot x$ for $\alpha, \beta \in \mathfrak{S}_m$ and $x \in \mathbb{R}^m$. (This means that the map $(\pi, x) \mapsto \pi \cdot x$ is an action of the symmetric group $\mathfrak{S}_m$ on the set $\mathbb{R}^m$. What is the induced action on the function set $\mathfrak{F}_{\mathbb{R}^m}$? But on the ring of polynomial functions $\mathbb{R}[X_1, \ldots, X_m]$?)

Any homomorphism $\Phi : \mathsf{G} \to \mathfrak{S}_{\mathsf{M}}$ gives rise to an action θ of G on M defined by

$$\theta(g, x) = g \cdot x := \Phi(g)(x)$$

for all $g \in \mathsf{G}$ and all $x \in \mathsf{M}$. This really is an action because

$$g_2 \cdot (g_1 \cdot x) = \Phi(g_2)\left(\Phi(g_1)(x)\right) = \left(\Phi(g_2)\Phi(g_1)\right)(x) = \Phi(g_2 g_1)(x) = (g_2 g_1) \cdot x$$

for all $g_1, g_2 \in \mathsf{G}$ and all $x \in \mathsf{M}$, and

$$\mathbf{1} \cdot x = \Phi(\mathbf{1})(x) = \mathbf{1}_{\mathsf{M}}(x) = x$$

for all $x \in \mathsf{M}$.

Conversely, suppose that θ is an action of G on M. For a fixed element $g \in \mathsf{G}$ consider the mapping

$$\theta(g, \cdot) := \theta_g : \mathsf{M} \to \mathsf{M}, \quad x \mapsto g \cdot x.$$

This is invertible: it has an inverse, namely $\theta_{g^{-1}}$ because (for all $x \in \mathsf{M}$)

$$\theta_g \theta_{g^{-1}}(x) = \theta_g\left(\theta_{g^{-1}}(x)\right) = g \cdot (g^{-1} \cdot x) = (gg^{-1}) \cdot x = \mathbf{1} \cdot x = x$$

and similarly $\theta_{g^{-1}}\theta_g(x) = x$, which shows that

$$\theta_g \theta_{g^{-1}} = \theta_{g^{-1}} \theta_g = \mathbf{1}_{\mathsf{M}}.$$

In this way each element of G acts as a permutation of M. Furthermore, the map

$$\Phi : \mathsf{G} \to \mathfrak{S}_{\mathsf{M}}, \quad g \mapsto \theta_g$$

is a homomorphism. Indeed, for all $x \in \mathsf{M}$, we have

$$\Phi(g_2g_1)(x) = (g_2g_1) \cdot x = g_2 \cdot (g_1 \cdot x) = \Phi(g_2)\left(\Phi(g_1)(x)\right) = \left(\Phi(g_2)\Phi(g_1)\right)(x)$$

and so

$$\Phi(g_2g_1) = \Phi(g_2)\Phi(g_1).$$

We call a homomorphism $\Phi : \mathsf{G} \to \mathfrak{S}_{\mathsf{M}}$ a **permutation representation** of G on M or a *representation* of G as a group of transformations (permutations) of M. What we have just shown is that every such representation gives rise to an action of G on M and that, conversely, every action gives rise to a permutation representation.

To summarize, we have

Proposition 15. *There is a one-to-one correspondence between actions of the group G on the set M and the representations of G by permutations of M.*

In view of this result we shall use the language of group actions and of permutation representations interchangeably.

Note: All this can be done with right actions, but a little care must be exercised. If $(x, g) \mapsto x \cdot g$ is the right action of G on M, the corresponding permutation representation of G on M is given by $g \mapsto \tau(\cdot, g^{-1})$. (Without this inverse we would not obtain a homomorphism from G to $\mathfrak{S}_{\mathsf{M}}$ but an "anti-homomorphism" or, if one prefers, a homomorphism from the *opposite* group G^{op} to $\mathfrak{S}_{\mathsf{M}}$.)

One can use *right translations* $R_a : \mathsf{G} \to \mathsf{G}, \quad g \mapsto ga$ to define natural actions of G on itself and on the set of *right* cosets (denoted by $\mathsf{G}\backslash\mathsf{H}$). The action ρ^{H} of G on the set $\mathsf{G}\backslash\mathsf{H}$ has a corresponding *right regular representation*

$$\mathsf{G} \to \mathfrak{S}_{\mathsf{G}\backslash\mathsf{H}}, \quad g \mapsto \rho^{\mathsf{H}}_{g^{-1}}.$$

A major theme of mathematical endeavour is to understand groups in terms of their actions. An interesting (and important) case is when the set M on which the group acts carries some extra structure—which will generally have a "geometric" flavour—and we will require the group action to respect this structure.

Consider the case when $\mathsf{M} = \mathbb{R}^m$, which is a *vector space*, and require G to act on $\mathbb{R}^m$ by *linear* transformations. That is, we replace the symmetric group $\mathfrak{S}_{\mathbb{R}^m}$ with $\mathsf{GL}\left(\mathbb{R}^m\right)$, the group of invertible linear maps $\mathbb{R}^m \to \mathbb{R}^m$; this group is (isomorphic to) the general linear group $\mathsf{GL}\left(m, \mathbb{R}\right)$. A homomorphism $\rho : \mathsf{G} \to \mathsf{GL}\left(m, \mathbb{R}\right)$ is called a **linear representation** of G. To put it another way, a linear representation of G is a concrete realization of the group G as a collection of invertible matrices.

A **faithful** linear representation of a group G is an *embedding* of G into a matrix group; that is, the homomorphism $\rho : \mathsf{G} \to \mathsf{GL}\,(m, \mathbb{R})$ is injective: distinct elements of the group correspond to distinct matrices. In this case we refer to G as a **linear group**.

Example 23. For every $m \in \mathbb{N}$, *the symmetric group* $\mathfrak{S}_m$ *can be embedded into* $\mathsf{GL}\,(m, \mathbb{R})$. In order to embed $\mathfrak{S}_m$ into (the general linear group) $\mathsf{GL}\,(m, \mathbb{R})$, we must find an injective homomorphism $\rho : \mathfrak{S}_m \to \mathsf{GL}\,(m, \mathbb{R})$; this involves assigning to each permutation $\pi \in \mathfrak{S}_m$ an invertible linear map $\rho_\pi : \mathbb{R}^m \to \mathbb{R}^m$.

Given an ordered basis $(e_i)_{1 \leq i \leq m}$ of $\mathbb{R}^m$, let $\rho_\pi : \mathbb{R}^m \to \mathbb{R}^m$ be the $\underline{\text{unique}}$ linear map that permutes these elements according to the permutation $\pi \in \mathfrak{S}_m$. That is, each π corresponds to a bijection $\pi : [m] \to [m]$; define now the action of ρ_π on the vectors e_i by

$$\rho_\pi(e_i) := e_{\pi(i)}, \quad i = 1, \dots, m.$$

Since the vectors e_i are linearly independent, ρ_π extends $\underline{\text{uniquely}}$ to a linear map on their span, which is $\mathbb{R}^m$: it sends the vector $x = x_1 e_1 + \cdots + x_m e_m \in \mathbb{R}^m$ to the vector

$$\begin{aligned} \rho_\pi(x) &= x_1 \rho_\pi(e_1) + \cdots + x_m \rho_\pi(e_m) \\ &= x_1 e_{\pi(1)} + \cdots + x_m e_{\pi(m)} \\ &= x_{\pi^{-1}(1)} e_1 + \cdots + x_{\pi^{-1}(m)} e_m \in \mathbb{R}^m. \end{aligned}$$

(The action $(\pi, x) \mapsto \pi \cdot x := \rho_\pi(x)$ of $\mathfrak{S}_m$ on $\mathbb{R}^m$ is given by $\pi \cdot (x_1, \dots, x_m) = (x_{\pi^{-1}(1)}, \dots, x_{\pi^{-1}(m)})$.) The map $\rho : \pi \mapsto \rho_\pi$ is a homomorphism since

$$\rho_\pi \rho_\sigma(e_i) = \rho_\pi\left(e_{\sigma(i)}\right) = e_{\pi(\sigma(i))} = \rho_{\pi\sigma}(e_i).$$

(It is easy to see that this homomorphism is injective.)

Choosing the standard basis for the vector space $\mathbb{R}^m$, the linear transformations ρ_π are represented by *permutation matrices* $P = \Phi(\pi) \in \mathsf{GL}\,(m, \mathbb{R})$ (see Example 15).

◇ **Exercise 25.** Verify that the natural homomorphism $\rho : \mathfrak{S}_m \to \mathsf{GL}\,(m, \mathbb{R})$ is injective.

One wants to think of two G-sets M and $\overline{\mathsf{M}}$ as being "essentially the same" if M can be identified with $\overline{\mathsf{M}}$ in such a way that the actions θ and $\bar{\theta}$ become the same. Formally, we say that M and $\overline{\mathsf{M}}$ are **equivalent** if there exist an automorphism $\phi \in \mathsf{Aut}\,(\mathsf{G})$ and a bijection $\beta : \mathsf{M} \to \overline{\mathsf{M}}$ such that for all $g \in \mathsf{G}$ the following diagram commutes:

$$\begin{tikzcd} \mathsf{M} \arrow[r, "\beta"] \arrow[d, "\theta_g"'] & \overline{\mathsf{M}} \arrow[d, "\bar{\theta}_{\phi(g)}"] \\ \mathsf{M} \arrow[r, "\beta"] & \overline{\mathsf{M}} \end{tikzcd}$$

In other words, $\beta(g \cdot x) = \phi(g) \cdot \beta(x)$ for all $x \in \mathsf{M}$ and all $g \in \mathsf{G}$. (We say that β is ϕ-**equivariant**.)

◇ **Exercise 26.** Show that two G-sets M and $\overline{\mathsf{M}}$ are equivalent if and only if there exist an automorphism $\phi \in \mathsf{Aut}(\mathsf{G})$ and a bijection $\beta : \mathsf{M} \to \overline{\mathsf{M}}$ such that the following diagram commutes:

$$\begin{tikzcd}
\mathsf{G} \times \mathsf{M} \arrow[r, "\phi\times\beta"] \arrow[d, "\theta"'] & \mathsf{G} \times \overline{\mathsf{M}} \arrow[d, "\overline{\theta}"] \\
\mathsf{M} \arrow[r, "\beta"] & \overline{\mathsf{M}}
\end{tikzcd}$$

2.2. **Orbits and stabilizers.** There are two fundamental aspects of G-sets: orbits and stabilizers. Let M be a G-set (with respect to $\Phi : \mathsf{G} \to \mathfrak{S}_{\mathsf{M}}$). Two points $x, y \in \mathsf{M}$ are said to be G-**equivalent** if $y = g \cdot x$ for some $g \in \mathsf{G}$.

◇ **Exercise 27.** Verify that the G-equivalence relation is a genuine equivalence relation on M (which divides M into disjoint equivalence classes).

Each equivalence class is called a G-**orbit**. (Usually, we will say simply *orbit* instead of G-orbit.) The orbit containing $x \in \mathsf{M}$ is denoted $\mathsf{Orb}\,(x)$ or $\mathsf{G}x$; thus

$$\mathsf{Orb}\,(x) = \mathsf{G}x := \{g \cdot x \,:\, g \in \mathsf{G}\}.$$

NOTE: The notion of an orbit arose from geometry. For example, if $\mathsf{G} = \mathsf{SO}\,(2)$ is the group of rotations of the (Euclidean) plane about the origin, then the orbit of a point P is the circle centered at the origin passing through P, and the set $\mathsf{M} = \mathbb{R}^2$ is the union of all the concentric circles, including the one with zero radius (consisting of a single point, the origin).

Let $x \in \mathsf{M}$. Consider the set

$$\mathsf{St}\,(x) = \mathsf{G}_x := \{g \in \mathsf{G} \,:\, g \cdot x = x\}.$$

It is called the **stabilizer** (or the **isotropy group**) of x.

◇ **Exercise 28.** Show that for any G-set M and any element $x \in \mathsf{M}$, the stabilizer $\mathsf{St}\,(x)$ is a subgroup of G.

EXAMPLE 24. If G acts on itself by left translations and $x \in \mathsf{G}$, then $\mathsf{Orb}\,(x) = \mathsf{G}$ (there is only one orbit) and $\mathsf{St}\,(x)$ is the trivial group $\{\mathbf{1}\}$.

EXAMPLE 25. If G acts on itself by conjugation and $x \in \mathsf{G}$, then $\mathsf{Orb}\,(x)$ is the *conjugacy class* $C(x)$ of x (i.e., the set of all group elements of the form gxg^{-1} as g varies over G), and

$$\mathsf{St}\,(x) = \left\{g \in \mathsf{G} \,:\, gxg^{-1} = x\right\} = \{g \in \mathsf{G} \,:\, gx = xg\}$$

(the *centralizer* $\mathsf{Z}\,(x)$ of x).

Example 26. If G acts by conjugation on the set of all its subgroups and $\mathsf{H} \leq \mathsf{G}$, then $\mathsf{Orb}\,(\mathsf{H}) = \{g\,\mathsf{H}\,g^{-1} \,:\, g \in \mathsf{G}\}$ (all the conjugates of H) and $\mathsf{St}\,(\mathsf{H})$ is the *normalizer* $\mathsf{N}\,(\mathsf{H})$ of H.

Example 27. Let $\mathsf{M} = \Bbbk[X_1, \dots, X_m]$ and $\mathsf{G} = \mathfrak{S}_m$. If $f \in \mathsf{M}$, then $\mathsf{Orb}\,(f)$ is the set of distinct polynomials of the form $\pi \cdot f$, $\pi \in \mathfrak{S}_m$, and

$$\mathsf{St}\,(f) = \mathcal{S}\,(f) = \{\pi \in \mathfrak{S}_m \,:\, \pi \cdot f = f\}\,.$$

◇ **Exercise 29.** Show that $\mathsf{St}\,(g \cdot x) = g\,\mathsf{St}\,(x)\,g^{-1}$ for all $x \in \mathsf{M}$ and all $g \in \mathsf{G}$. (This means that points in the same orbit have conjugate stabilizers.)

Theorem 16. (Orbit-Stabilizer Theorem) *For each $x \in \mathsf{M}$, the correspondence $g \cdot x \mapsto g\,\mathsf{St}\,(x)$ is a bijection between the orbit $\mathsf{Orb}\,(x)$ and the set $\mathsf{G}/\mathsf{St}\,(x)$ of left cosets of the stabilizer $\mathsf{St}\,(x)$ in G.*

Proof. The correspondence is clearly surjective. It is injective because if $g\,\mathsf{St}\,(x) = g'\,\mathsf{St}\,(x)$, then $g = g'h$ for some element $h \in \mathsf{St}\,(x)$, and therefore

$$g \cdot x = (g'h) \cdot x = g' \cdot (h \cdot x) = g' \cdot x.$$

□

Corollary 17. *If G is finite, the size of each orbit is a divisor of the order of G.*

Proof. By the Orbit-Stabilizer Theorem, the size of the orbit $\mathsf{Orb}\,(x)$ is $|\mathsf{G}/\mathsf{St}\,(x)| = |\mathsf{G}|/|\mathsf{St}\,(x)|$, therefore

$$|\mathsf{Orb}\,(x)| \cdot |\mathsf{St}\,(x)| = |\mathsf{G}|.$$

□

◇ **Exercise 30.** Let M and $\overline{\mathsf{M}}$ be two G-sets such that there exists a bijection $\beta : \mathsf{M} \to \overline{\mathsf{M}}$ such that $\beta(g \cdot x) = g \cdot \beta(x)$ for all $g \in \mathsf{G}$ and all $x \in \mathsf{M}$. (This means that M and $\overline{\mathsf{M}}$ are equivalent.) Let $x \in \mathsf{M}$ and $\overline{x} \in \overline{\mathsf{M}}$ such that $\overline{x} = \beta(x)$. Show that $\mathsf{St}\,(x) = \mathsf{St}\,(\overline{x})$.

2.3. **Particular G-sets.** Let M be a G-set (with respect to the homomorphism $\Phi : \mathsf{G} \to \mathfrak{S}_{\mathsf{M}}$). We define some particular properties the action $\theta : \mathsf{G} \times \mathsf{M} \to \mathsf{M}$ can have.

Definition 28. θ is an **effective action** if Φ is injective (i.e., $\mathsf{Ker}\,\Phi = \{\mathbf{1}\}$).

This always happens when $\mathsf{G} \leq \mathfrak{S}_{\mathsf{M}}$. We observe that $\mathsf{Ker}\,\Phi = \bigcap_{x \in \mathsf{M}} \mathsf{St}\,(x)$, an element of $\mathsf{Ker}\,\Phi$ being exactly an element of G contained in every isotropy group. If θ is not effective, then there exists a factorization $\widehat{\Phi}$ through $\mathsf{G}/\mathsf{Ker}\,\Phi$

$$\begin{tikzcd}
\mathsf{G} \arrow[r] \arrow[d, "\Phi"'] & \mathsf{G}/\mathsf{Ker}\,\Phi \arrow[d, "\widehat{\Phi}"] \\
\mathfrak{S}_{\mathsf{M}} \arrow[r, equal] & \mathfrak{S}_{\mathsf{M}}
\end{tikzcd}$$

and $\mathsf{G}/\mathsf{Ker}\,\Phi$ acts effectively on M.

Example 29. The action of G on itself by conjugation (inner automorphisms) has the center $\mathsf{Z}\left(\mathsf{G}\right)$ as kernel.

In case of an effective action, we think of the group G as being identified with its image under (the associated homomorphism) Φ, a subgroup of the symmetric group $\mathfrak{S}_{\mathsf{M}}$ and we are essentially back with the important special case of permutation (transformation) groups.

Definition 30. θ is a **free action** if $g \cdot x = x$ for some $x \in \mathsf{M}$ implies $g = \mathbf{1}$.

This means that the transformation $\theta_g : x \mapsto g \cdot x$ for $g \neq \mathbf{1}$ has no fix points (free means "free of fix points"). The isotropy group is reduced to trivial subgroup: $\mathsf{St}\left(x\right) = \{\mathbf{1}\}$ for every $x \in \mathsf{M}$. Clearly, *every free action is effective.* A G-set with a free action is also called a **principal** G-set.

Example 31. The action of G on itself by left translations is free.

Definition 32. θ is a **transitive action** if for $x_1, x_2 \in \mathsf{M}$ there exists a $g \in \mathsf{G}$ such that $g \cdot x_1 = x_2$, and **simply transitive** if, moreover, the element g is unique.

A simply transitive action is free. Conversely, *a free action is simply transitive on each orbit.* Indeed, let $x = g_1 \cdot x_0 = g_2 \cdot x_0 \in \mathsf{Orb}\left(x_0\right)$. Then

$$x_0 = g_2^{-1} \cdot x = g_2^{-1} \cdot \left(g_1 \cdot x_0\right) = \left(g_2^{-1} g_1\right) \cdot x_0$$

and therefore $g_2^{-1} g_1 \in \mathsf{St}\left(x_0\right) = \{\mathbf{1}\}$, hence $g_1 = g_2$.

Note: Stabilizers, in some sense, tell us how far a group is from acting simply transitively: just notice that $g \cdot x = h \cdot x \iff h^{-1} g \in \mathsf{St}\left(x\right)$.

Proposition 18. *If G is an Abelian group, any effective and transitive action is simply transitive.*

Proof. Let M be a G-set and let $x, y \in \mathsf{M}$. Since our action is transitive, there is *at least* some element $g \in \mathsf{G}$ such that $g \cdot x = y$. Assume that we have $g_1, g_2 \in \mathsf{G}$ with $g_1 \cdot x = g_2 \cdot x = y$. We shall prove that, actually, $g_1 \cdot z = g_2 \cdot z$ for all $z \in \mathsf{M}$. As our action is effective, we must have $g_1 = g_2$, and this proves our statement.

Let $z \in \mathsf{M}$. There is some $g' \in \mathsf{G}$ such that $z = g' \cdot x$. Then we have

$$\begin{aligned} g_1 \cdot z &= g_1 \cdot (g' \cdot x) \\ &= (g_1 g') \cdot x \\ &= (g' g_1) \cdot x \quad \text{(since } \mathsf{G} \text{ is Abelian)} \\ &= g' \cdot (g_1 \cdot x) \\ &= g' \cdot (g_2 \cdot x) \\ &= (g' g_2) \cdot x \\ &= (g_2 g') \cdot x \quad \text{(since } \mathsf{G} \text{ is Abelian)} \\ &= g_2 \cdot (g' \cdot x) \\ &= g_2 \cdot z. \end{aligned}$$

Therefore, $g_1 \cdot z = g_2 \cdot z$ for all $z \in \mathsf{M}$, as claimed. □

Definition 33. A G-set M is called **homogeneous** if G acts transitively on M.

Example 34. The action of G on itself by left translations is transitive. For, if $x, y \in \mathsf{M} = \mathsf{G}$, and if we take $g := yx^{-1}$, then $gx = y$.

Example 35. The action of the general linear group $\mathsf{GL}\,(n, \Bbbk)$ on $\Bbbk^n \backslash \{0\}$ is transitive. For, given any non-zero vector x in $\Bbbk^n$, there certainly exists an invertible $n \times n$ matrix A over $\Bbbk$, whose first column is x and then

$$A \begin{bmatrix} 1 \\ \vdots \\ 0 \end{bmatrix} = x.$$

The transitivity of the action follows. Why ?

Example 36. The orthogonal group $\mathsf{O}\,(n)$ acts transitively on the unit sphere $\mathbb{S}^{n-1} \subset \mathbb{R}^n$. More generally, *an action of* G *on* M *induces a transitive action on each orbit.*

It is not difficult to see that every G-set is expressible in a unique way as a disjoint union of orbits (this is Problem 7). So many questions about actions of groups (on sets) can be reduced to the study of homogeneous G-sets.

There is a simple method for constructing a homogeneous set (this is Example 20). Let G be a group and consider a subgroup $\mathsf{H} \leq \mathsf{G}$. Then we can define an action of G on the orbit set G/H. The left translation $L_g : \mathsf{G} \to \mathsf{G}$ satisfies $L_g\,(a\,\mathsf{H}) = (ga)\,\mathsf{H}$ and therefore defines a transformation (of G/H)

$$\lambda_g^{\mathsf{H}} : \mathsf{G}/\mathsf{H} \to \mathsf{G}/\mathsf{H}, \quad a\,\mathsf{H} \mapsto (ga)\,\mathsf{H}.$$

This makes G/H a G-set, which is *homogeneous.*

◇ **Exercise 31.** Show that the action (of G on G/H)

$$\mathsf{G} \times \mathsf{G}/\mathsf{H} \ni (g, a\,\mathsf{H}) \mapsto (ga)\,\mathsf{H} \in \mathsf{G}/\mathsf{H}$$

is transitive.

The method of constructing a homogeneous set (as described above) is, in a certain sense, universal: *every homogeneous* G*-set is equivalent to a (homogeneous)* G*-set of the form* G/H *for a suitable* $\mathsf{H} \leq \mathsf{G}$. (This orbit set provides a sort of "canonical form" for homogeneous G-sets, under equivalence). Indeed, let M be an arbitrary homogeneous G-set and let $x_0 \in \mathsf{M}$. Put $\mathsf{H} = \mathsf{St}\,(x_0) \leq \mathsf{G}$. Then the map

$$\beta : \mathsf{M} \ni x = g \cdot x_0 \mapsto g\,\mathsf{H} \in \mathsf{G}/\mathsf{H}$$

is an equivalence of G. (β is well defined, bijective and equivariant.)

Note: If M is a homogeneous G-set, then we have (for each $x \in \mathsf{M}$)

$$\mathsf{M} = \mathsf{Orb}\,(x) \approx \mathsf{G}/\mathsf{St}\,(x) \qquad \text{(one-to-one correspondence).}$$

This equation is invaluable, for it reduces the study of a set M (usually endowed with some "structure") to an algebraic problem, namely the study of the pair $(\mathsf{G}, \mathsf{St}\,(x))$.

In the context of group actions (on sets), the homogeneous G-sets play a role somewhat similar to that played by the vector spaces $\Bbbk^n$ in the context of linear algebra. The result, stated above, regarding the homogeneous G-sets corresponds to the classical result which states that every finite-dimensional vector space V over $\Bbbk$ is isomorphic to some (vector space) $\Bbbk^n$. In the same way an isomorphism between V and $\Bbbk^n$ assumes the choice of a basis for V, an equivalence (G-isomorphism) between a homogeneous G-set M and G/H assumes the choice of a point in G. Also, in the same way a vector space $\Bbbk^n$ admits a preferred basis, a homogeneous set admits a preferred point. Finally, the statement that two vector spaces $\Bbbk^m$ and $\Bbbk^n$ are isomorphic if and only if $m = n$, corresponds to the statement that *two homogeneous* G*-sets* G/H_1 *and* G/H_2 *are equivalent if and only if* H_1 *and* H_2 *are conjugate in* G (this is Problem 10).

2.4. **Examples of group actions.** We now give some further examples of group actions on sets.

Example 37. Let $\mathsf{M} = \mathsf{G}$ be a group. Then G acts on itself in several important ways:

(a) $\theta_g(x) = gx$ (left translation);
(b) $\theta_g(x) = xg^{-1}$ (right translation);
(c) $\theta_g(x) = gxg^{-1}$ (inner automorphism).

Example 38. If $\mathsf{M} = \mathsf{V}$ is a *vector space* (over the field $\Bbbk$), its linear group

$$\mathsf{G} = \mathsf{GL}\,(\mathsf{V}) := \big\{\alpha \in \mathsf{V}^{\mathsf{V}} \,:\, \alpha \text{ is linear and bijective}\big\} \leq \mathfrak{S}_{\mathsf{V}}$$

acts on M. When $\mathsf{V} = \Bbbk^n$, the group $\mathsf{GL}\,(\Bbbk^n)$ is (isomorphic to) the general linear group $\mathsf{GL}\,(n, \Bbbk)$. So the group $\mathsf{GL}\,(n, \Bbbk)$ acts on $\Bbbk^n$ by left multiplication:

$$\mathsf{GL}\,(n, \Bbbk) \times \Bbbk^n \ni (A, x) \mapsto Ax \in \Bbbk^n$$

(here the elements of $\Bbbk^n$ are viewed as $n \times 1$ matrices over $\Bbbk$).

Example 39. Let $\mathsf{M} = \mathsf{E}$ be an *Euclidean vector space*, and put

$$\mathsf{G} = \mathsf{O}\,(\mathsf{E}) := \{\alpha \in \mathsf{GL}\,(\mathsf{E}) \,:\, \alpha \text{ is an isometry}\}\,.$$

Then there is a natural action of G on M. When $\mathsf{E} = \mathbb{R}^n$, the group $\mathsf{O}\,(\mathbb{R}^n)$ is the orthogonal group $\mathsf{O}\,(n) \leq \mathsf{GL}\,(n, \mathbb{R})$. In particular, the rotation group $\mathsf{SO}\,(n)$ acts naturally on $\mathbb{R}^n$. For $n \geq 1$, let

$$\mathbb{S}^{n-1} := \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n \,:\, x_1^2 + \cdots + x_n^2 = 1\right\} \subset \mathbb{R}^n$$

be the *unit sphere*. In particular, $\mathbb{S}^2$ is the usual sphere in $\mathbb{R}^3$. Thus, we have an action

$$\mathsf{SO}\,(3) \times \mathbb{S}^2 \to \mathbb{S}^2, \quad (R, x) \mapsto Rx.$$

This action is *transitive*. This is so because, for any two points $x, y \in \mathbb{S}^2$, there is a rotation whose axis is perpendicular to the plane containing x, y, and the center of the sphere (this plane is not unique when x and y are antipodal, i.e., on a diameter). Similarly, for $n \geq 1$, we get an action of $\mathsf{SO}\,(n)$ on $\mathbb{S}^{n-1}$.

Note: An **Euclidean vector space** E is a finite-dimensional vector space over $\mathbb{R}$, together with a positive definite symmetric bilinear form ϕ (i.e., $\phi : \mathsf{E} \times \mathsf{E} \to \mathbb{R}$ is symmetric and bilinear, and $\phi(x, x) > 0$ for all $x \neq 0$). We write $\phi(x, y) = (x \,|\, y)$ and call this number the *scalar product* of x and y. The *norm* of x is $\|x\| := \sqrt{\phi(x, x)} = \sqrt{(x \,|\, x)}$. If $(x \,|\, y) = 0$, we say that x and y are *orthogonal.*

The standard example of an Euclidean vector space is $\mathsf{E} = \mathbb{R}^n$, with

$$\phi((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = x_1 y_1 + \cdots + x_n y_n.$$

An orthogonal set of non-zero vectors is linearly independent. If $(e_i)_{1 \leq i \leq n}$ is an *orthonormal* basis for E, the coefficients of the decomposition $x = x_1 e_1 + \cdots + x_n e_n$ are given by $x_i = (x \,|\, e_i)$. Moreover,

$$(x_1 e_1 + \cdots + x_n e_n \,|\, y_1 e_1 + \cdots + y_n e_n) = x_1 y_1 + \cdots + x_n y_n.$$

Let $\mathsf{E}, \bar{\mathsf{E}}$ be two Euclidean vector spaces of the same dimension, and let $\alpha : \mathsf{E} \to \bar{\mathsf{E}}$ be a map. The following conditions are logically equivalent:

(a) α is linear, and $\|\alpha(x)\| = \|x\|$ for all $x \in \mathsf{E}$;

(b) $(\alpha(x) \,|\, \alpha(y)) = (x \,|\, y)$ for all $x, y \in \mathsf{E}$.

Such a map is necessarily bijective and is called an **isometry**. The set of all such isometries is denoted by $\mathsf{O}\,(\mathsf{E}; \bar{\mathsf{E}})$. *Every n-dimensional Euclidean vector space is isometric to $\mathbb{R}^n$.*

The group $\mathsf{O}\,(\mathsf{E}) := \mathsf{O}\,(\mathsf{E}; \mathsf{E})$ is called the **orthogonal group** of E; we write $\mathsf{O}\,(n) = \mathsf{O}\,(\mathbb{R}^n)$. The condition $\alpha \in \mathsf{O}\,(\mathsf{E})$ is equivalent to $A^\top A = \mathbf{1}$, where A is the matrix of α in some (or any) orthonormal basis and $\mathbf{1}$ is the identity matrix. In particular, $\det\alpha = \pm 1$. We set $\mathsf{SO}(\mathsf{E}) := \{\alpha \in \mathsf{O}(\mathsf{E}) \,:\, \det\alpha = 1\}$. The elements of (the group) $\mathsf{SO}(\mathsf{E})$ are called *rotations*.

EXAMPLE 40. Let $\Bbbk$ be a field (think of $\mathbb{R}$ or $\mathbb{C}$). Given $A \in \mathsf{GL}(n, \Bbbk)$ and $b \in \Bbbk^n$, we define a map

$$\tau_{A,b} : \Bbbk^n \to \Bbbk^n, \quad x \mapsto Ax + b.$$

These maps are known as **affine transformations** of $\Bbbk^n$; they constitute the **affine group**

$$\mathsf{Aff}(n, \Bbbk) := \{\tau_{A,b} \,:\, A \in \mathsf{GL}(n, \Bbbk),\ b \in \Bbbk^n\} \leq \mathfrak{S}_{\Bbbk^n}.$$

The group $\mathsf{Aff}(n, \Bbbk)$ acts naturally on $\Bbbk^n$: this makes $\Bbbk^n$ an $\mathsf{Aff}(n, \Bbbk)$-set (known as the n-dimensional *affine space* over $\Bbbk$). Notice that $\mathsf{GL}(n, \Bbbk)$ is a subgroup of $\mathsf{Aff}(n, \Bbbk)$; that is, invertible linear transformations are some of the affine transformations of $\Bbbk^n$. Notice also that the group of translations $x \mapsto x + b$ (which is a normal subgroup of $\mathsf{Aff}(n, \Bbbk)$) acts on the affine space $\Bbbk^n$ regularly (in the sense of EXAMPLE 20).

NOTE: Let M be a (non-empty) set and let V be vector space over the field $\Bbbk$ (think again of $\mathbb{R}$ or $\mathbb{C}$) considered with its additive group structure. An **affine space** over $\Bbbk$ is a structure $(\mathsf{M}, \mathsf{V}, \theta)$, where $\theta : \mathsf{V} \times \mathsf{M} \to \mathsf{M}$ is an effective and transitive action on M. The vector space V is said to underlie the affine space M. We put

$$\theta(v, x) = x + v.$$

The map $\theta_v = \theta(v, \cdot)$ is called the *translation* of M by the vector v. The action θ is simply transitive (see PROPOSITION 18), so there exists a function $\Theta : \mathsf{M} \times \mathsf{M} \to \mathsf{V}$ such that $y = \theta(\Theta(x, y), x)$ for all $x, y \in \mathsf{M}$. We set $\overrightarrow{xy} := \Theta(x, y)$, and sometimes say that $\overrightarrow{xy}$ is the *free vector* associated with the pair (x, y). We also write $\overrightarrow{xy} = y - x$. The fact that θ is a V-action can be translated as follows:

$$(x + v) + w = x + (v + w).$$

In particular, Θ satisfies the following conditions:

(AS1) $\Theta_x : \mathsf{M} \ni y \mapsto \Theta(x, y) \in \mathsf{V}$ is a bijection for all $x \in \mathsf{M}$;
(AS2) $\Theta(x, y) + \Theta(y, z) = \Theta(x, z)$ for all $x, y, z \in \mathsf{M}$

since we have $\Theta_x^{-1}(v) = x + v$. (The identity $\Theta(x, y) + \Theta(y, z) = \Theta(x, z)$ is known as CHASLES'S RELATION.)

Alternative definition. Given a (non-empty) set M and a vector space V over the field $\Bbbk$, assume that $\Theta : \mathsf{M} \times \mathsf{M} \to \mathsf{V}$ is a function satisfying conditions (AS1) and (AS2). Then M is an affine space under the action $\theta(v, x) = \Theta_x^{-1}(v)$. This indeed is an equivalent definition, for we have $\Theta(x, x) = 0$, $\Theta(y, x) = -\Theta(x, y)$, $\theta(-v) \circ \theta(v) = \mathbf{1}_{\mathsf{M}}$, and thus

$$\theta(v) \circ \theta(w) = \theta(v + w).$$

Affine maps (morphisms) can be defined between two affine spaces (over the same field $\Bbbk$). Heuristically, such a map consists of a translation and a linear transformation. (If $\mathsf{M} = \overline{\mathsf{M}} = \mathbb{R}$, we recover the well-known maps $x \mapsto ax + b$ for $a, b \in \mathbb{R}$, $a \neq 0$.) The set

$$\mathsf{AGL}\,(\mathsf{M}) := \{\alpha \in \mathfrak{S}_{\mathsf{M}} \,:\, \alpha \text{ is an affine map}\}$$

is a group, called the **affine group** of M. It turns out that *(the affine space)* M *is a homogeneous* $\mathsf{AGL}\,(\mathsf{M})$*-set.* The vector space $\Bbbk^n$ has a natural affine structure: when $\mathsf{M} = \mathsf{V} = \Bbbk^n$, the function

$$\Bbbk^n \times \Bbbk^n \ni (x, y) \mapsto y - x \in \Bbbk^n$$

induces an action of $\Bbbk^n$ on itself. We write $\mathsf{AGL}\,(\Bbbk^n) = \mathsf{Aff}\,(n, \Bbbk) \leq \mathfrak{S}_{\Bbbk^n}$.

Example 41. Let $\mathsf{G} = \mathbb{R}$ and $\mathsf{M} = \mathbb{S}^3 := \{x \in \mathbb{R}^4 \,:\, \|x\| = 1\} \subset \mathbb{R}^4$. Identifying $\mathbb{R}^4$ with $\mathbb{C}^2$, we can define the action

$$\mathbb{R} \times \mathbb{S}^3 \to \mathbb{S}^3, \quad (t, z, z') \mapsto \left(e^{it}z, e^{it}z'\right).$$

This example is of great importance in geometry.

Problems (6–10)

(6) (a) The **upper half-plane** $\mathbb{H}^2$ is the (open) subset of $\mathbb{R}^2$ consisting of all points $(x, y) \in \mathbb{R}^2$ with $y > 0$. It is convenient to identify $\mathbb{R}^2$ with the set of complex numbers. So

$$\mathbb{H}^2 := \{z = x + iy \in \mathbb{C} \,:\, y > 0\}.$$

Define the map

$$\theta : \mathsf{SL}\,(2, \mathbb{R}) \times \mathbb{H}^2 \to \mathbb{H}^2, \qquad \left(A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, z\right) \mapsto \frac{az + b}{cz + d}.$$

Show that θ is an action of the special linear group $\mathsf{SL}\,(2, \mathbb{R})$ on the upper half-plane $\mathbb{H}^2$. Is this action transitive ?

(b) Consider the set of all Möbius transformations $\mu_{abcd} : \mathbb{C}_\infty \to \mathbb{C}_\infty$ corresponding to the case $a, b, c, d \in \mathbb{R}$ with $ad - bc = 1$. This set is denoted by $\mathsf{M\ddot{o}b}^+_{\mathbb{R}}$. Show that $\mathsf{M\ddot{o}b}^+_{\mathbb{R}}$ is a subgroup of the Möbius group $\mathsf{M\ddot{o}b}$.

(c) Define the function

$$\Phi : \mathsf{SL}\,(2, \mathbb{R}) \to \mathsf{M\ddot{o}b}^+_{\mathbb{R}}, \quad A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \mu_{abcd}.$$

Show that Φ is a surjective homomorphism (epimorphism) whose kernel is $\mathsf{Ker}\,(\Phi) = \{\mathbf{1}, -\mathbf{1}\}$. Hence deduce that the group $\mathsf{M\ddot{o}b}^+_{\mathbb{R}}$ is isomorphic to the quotient group $\mathsf{SL}\,(2, \mathbb{R})/\{\mathbf{1}, -\mathbf{1}\}$, denoted by $\mathsf{PSL}\,(2, \mathbb{R})$. (This latter group turns out to be the group of *projective transformations* of the real projective line $\mathbb{RP}^1$.)

(7) Show that every G-set can be expressed in just one way as the disjoint union of a family of orbits.

(8) For which n is the special linear group $\mathsf{SL}\,(n, \mathbb{R})$ acting transitively on $\mathbb{R}^n \setminus \{0\}$?

(9) Show that two homogeneous G-spaces M and $\overline{\mathsf{M}}$ are equivalent if and only if there exist a automorphism $\phi \in \mathsf{Aut}\,(\mathsf{G})$ and elements $x_0 \in \mathsf{M}$ and $\overline{x}_0 \in \overline{\mathsf{M}}$ such that

$$\phi\,(\mathsf{St}\,(x_0)) = \mathsf{St}\,(\overline{x}_0).$$

(10) Show that two homogeneous G-sets $\mathsf{M} = \mathsf{G}/\mathsf{H}_1$ and $\overline{\mathsf{M}} = \mathsf{G}/\mathsf{H}_2$ are equivalent if and only if the subgroups H_1 and H_2 are conjugate in G.

3. Euclidean Spaces

Inner product and norm • Open and closed sets • Continuity • Differentiation.

3.1. **Inner product and norm.** Let $\mathbb{R}$ be the *set* of real numbers and let $\mathbb{R}^m$ $(m \geq 1)$ denote the Cartesian product of m copies of $\mathbb{R}$. The elements of $\mathbb{R}^m$ are ordered m-tuples of real numbers. Thus

$$\mathbb{R}^m := \{x = (x_1, \ldots, x_m) \, : \, x_i \in \mathbb{R}\} \, .$$

An element of $\mathbb{R}^m$ is often called a *point*. Under the usual operations

$$x + y := (x_1 + y_1, \ldots, x_m + y_m) \quad \text{and} \quad \lambda x := (\lambda x_1, \ldots, \lambda x_m) \qquad (x, y \in \mathbb{R}^m, \, \lambda \in \mathbb{R})$$

$\mathbb{R}^m$ is a *vector space* over $\mathbb{R}$. Hence the elements of $\mathbb{R}^m$ can also be referred to as *vectors*.

Note : The set $\mathbb{R}^m$ may be equipped with various natural structures (e.g., group structure, vector space structure, topological structure, etc.) thus yielding various *spaces*, each such space having the same underlying *set* $\mathbb{R}^m$. We must usually decide from the context which structure is intended.

Many geometric concepts require an extra structure on $\mathbb{R}^m$ that we now define.

Definition 42. The **Euclidean space** $\mathbb{R}^m$ is the above mentioned vector space $\mathbb{R}^m$ together with the **standard inner product** (or dot product)

$$x \bullet y := x_1 y_1 + \cdots + x_m y_m \qquad (x, y \in \mathbb{R}^m).$$

We say that $x, y \in \mathbb{R}^m$ are **orthogonal** if $x \bullet y = 0$. The most important properties of the standard inner product are the following.

Proposition 19. *If x, y, z are vectors in $\mathbb{R}^m$ and $\lambda \in \mathbb{R}$, then*

(IP1) $x \bullet y = y \bullet x$ *(symmetry).*

(IP2) $(\lambda x + y) \bullet z = \lambda x \bullet z + y \bullet z$ *(linearity).*

(IP3) $x \bullet x \geq 0$, *and* $x \bullet x = 0$ *if and only if* $x = 0$ *(positive definiteness).*

Proof. Straightforward computation. □

Definition 43. The **Euclidean norm** $\|x\|$ of $x \in \mathbb{R}^m$ is defined as

$$\|x\| := \sqrt{x \bullet x}.$$

If $m = 1$, then $\|x\|$ is the usual *absolute value* $|x|$ of x. The relationship between the norm and the vector structure of $\mathbb{R}^m$ is very important.

◇ **Exercise 32.** Show that if $x, y \in \mathbb{R}^m$ and $\lambda \in \mathbb{R}$, then

(a) $\|x\| \geq 0$, and $\|x\| = 0$ if and only if $x = 0$ (positivity).
(b) $\|\lambda x\| = |\lambda| \, \|x\|$ (homogeneity).
(c) $x \bullet y = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right)$ (polarization identity).
(d) $\|x \pm y\|^2 = \|x\|^2 + \|y\|^2$ if and only if $x \bullet y = 0$ (Pythagorean property).

Theorem 20. (Cauchy-Schwarz Inequality) *If $x, y \in \mathbb{R}^m$, then*

$$|x \bullet y| \leq \|x\| \, \|y\|.$$

Equality holds if and only if x and y are linearly dependent.

Proof. If x and y are linearly dependent, equality clearly holds. Why ? If not, then $\lambda x - y \neq 0$ for all $\lambda \in \mathbb{R}$, so

$$\begin{aligned} 0 < \|\lambda x - y\|^2 &= (\lambda x_1 - y_1)^2 + \cdots + (\lambda x_m - y_m)^2 \\ &= (x_1^2 + \cdots + x_m^2)\,\lambda^2 - 2(x_1 y_1 + \cdots + x_m y_m)\,\lambda + y_1^2 + \cdots + y_m^2. \end{aligned}$$

Therefore the right hand side is a quadratic equation in λ with no real solution, and its discriminant must be negative. Thus

$$4\left(x_1 y_1 + \cdots + x_m y_m\right)^2 - 4\left(x_1^2 + \cdots + x_m^2\right)\left(y_1^2 + \cdots + y_m^2\right) < 0$$
$$\left(x \bullet y\right)^2 < \|x\|^2 \, \|y\|^2$$

which implies $|x \bullet y| < \|x\| \, \|y\|$. □

The Cauchy-Schwarz Inequality serves in proving several other inequalities (this is Problem 11).

Definition 44. The **standard basis** for $\mathbb{R}^m$ consists of the vectors

$$e_j = (\delta_{1j}, \ldots, \delta_{mj}), \quad j = \overline{1,m}$$

where δ_{ij} equals 1 if $i = j$ and equals 0 if $i \neq j$.

Thus we write

$$x = x_1 e_1 + \cdots + x_m e_m \qquad (x \in \mathbb{R}^m).$$

With respect to the standard inner product on $\mathbb{R}^m$, the standard basis is **orthonormal**, i.e., $e_i \bullet e_j = \delta_{ij}$ for $i, j = \overline{1,m}$. (Thus $\|e_j\| = 1$, while e_i and e_j for distinct i and j are orthogonal vectors.)

Definition 45. For $x, y \in \mathbb{R}^m$ we define the **Euclidean distance** $d(x,y)$ by

$$d(x,y) := \|x - y\|.$$

From **Exercise 32** and PROBLEM 11 we immediately obtain (for $x, y, z \in \mathbb{R}^m$)

(M1) $d(x,y) \geq 0$, and $d(x,y) = 0$ if and only if $x = y$.
(M2) $d(x,y) = d(y,x)$.
(M3) $d(x,z) \leq d(x,y) + d(y,z)$.

NOTE : (1) More generally, a **metric space** is defined as a set M equipped with a *distance* between its elements satisfying the properties (M1) – (M3). So *the Euclidean space* $\mathbb{R}^m$ *is a metric space*. The notation $d(x,y) = \|x - y\|$ is frequently useful even when we are dealing with the Euclidean space $\mathbb{R}^m$ as a metric space and not using its vector space structure. In particular, $\|x\| = d(x,0)$.

(2) An *abstract* concept of **Euclidean space** (i.e., a space satisfying the *axioms* of Euclidean geometry) can be introduced. It is defined as a structure $(\mathsf{M}, \mathsf{E}, \Phi)$, consisting of a (non-empty) *set* M, an associated *standard vector space* E (which is a real Euclidean vector space, i.e., a real vector space equipped with a scalar product $(\cdot|\cdot) : \mathsf{E} \times \mathsf{E} \to \mathbb{R}$), and a *structure map*

$$\Phi : \mathsf{M} \times \mathsf{M} \to \mathsf{E}, \quad (x,y) \mapsto \overrightarrow{xy}$$

such that

(ES1) $\overrightarrow{xy} + \overrightarrow{yz} = \overrightarrow{xz}$ for every $x, y, z \in \mathsf{M}$;
(ES2) for every $o \in \mathsf{M}$ and every $v \in \mathsf{E}$, there is a unique $x \in \mathsf{M}$ such that $\overrightarrow{ox} = v$.

Elements of M are called **points**, whereas elements of E are called **vectors**. ($\overrightarrow{ox}$ is the *position vector* of x with the initial point o.) The *dimension* of the space M is the dimension of the vector space E. It turns out that

(i) if we fix an arbitrary point $o \in \mathsf{M}$, there is a one-to-one correspondence between (the space) M and (the associated vector space) E (the mapping $x \mapsto \overrightarrow{ox}$ is a bijection);

(ii) in addition, if we fix an arbitrary (ordered) orthonormal basis $(e_1, e_2, \dots, e_m)$ of E, the (inner product) spaces E and $\mathbb{R}^m$ are *isomorphic*. In other words, the scalar product on E "is" the dot product: for $v, w \in \mathsf{E}$,

$$\begin{aligned} (v\,|\,w) &= (\,v_1e_1 + \cdots + v_me_m \,|\, w_1e_1 + \cdots + w_me_m\,) \\ &= v_1w_1 + \cdots + v_mw_m. \end{aligned}$$

In this sense, we *identify* the abstract m-dimensional Euclidean space M with the (concrete) Euclidean space $\mathbb{R}^m$.

We conclude this section with some important remarks (about notation). The element (vector) $(0, \dots, 0) \in \mathbb{R}^m$ will usually be denoted simply 0.

If $\tau : \mathbb{R}^m \to \mathbb{R}^m$ is a linear transformation, the matrix of τ with respect to the standard basis of $\mathbb{R}^m$ is the $m \times m$ matrix $T = [t_{ij}]$, where $T(e_j) = \sum_{i=1}^m t_{ij}e_i$ (the coefficients of $T(e_j)$ appear in the j^{th} column of the matrix). If the linear transformation $\sigma : \mathbb{R}^m \to \mathbb{R}^m$ has the matrix S, then (the composite) transformation $\sigma\tau$ has the matrix ST (matrix multiplication).

3.2. **Open and closed sets.** The analog in $\mathbb{R}^m$ of an *open interval* in $\mathbb{R}$ is introduced in the following

Definition 46. For $p \in \mathbb{R}^m$ and $\delta > 0$, we denote the **open ball** of center p and radius δ by

$$\mathcal{B}(p,\delta) := \{x \in \mathbb{R}^m \,:\, \|x-p\| < \delta\}.$$

A point p in a set $\mathsf{A} \subseteq \mathbb{R}^m$ is said to be an **interior point** of A if there exists $\delta > 0$ such that $\mathcal{B}(p,\delta) \subseteq \mathsf{A}$. The set of interior points of A is called the **interior** of A and is denoted by $\operatorname{int}(\mathsf{A})$. Note that $\operatorname{int}(\mathsf{A}) \subseteq \mathsf{A}$.

Definition 47. A set $\mathsf{A} \subseteq \mathbb{R}^m$ is said to be **open** (in $\mathbb{R}^m$) if $\mathsf{A} = \operatorname{int}(\mathsf{A})$ (i.e., if every point of A is an interior point of A).

Note that the *empty set* $\emptyset$ satisfies every definition involving conditions on its elements, therefore $\emptyset$ is open. Furthermore, the whole space $\mathbb{R}^m$ is open.

Proposition 21. *The set $\mathcal{B}(p,\delta)$ is open in $\mathbb{R}^m$, for every $p \in \mathbb{R}^m$ and $\delta > 0$.*

Proof. For arbitrary $q \in \mathcal{B}(p,\delta)$ set $\beta = \|q-p\|$, then $\delta - \beta > 0$. Hence $\mathcal{B}(q,\delta-\beta) \subseteq \mathcal{B}(p,\delta)$, because for every $x \in \mathcal{B}(q,\delta-\beta)$

$$\|x-p\| \leq \|x-q\| + \|q-p\| < (\delta-\beta)+\beta = \delta.$$

□

Proposition 22. *For any $\mathsf{A} \subseteq \mathbb{R}^m$, the interior $\operatorname{int}(\mathsf{A})$ is the largest open set contained in A.*

Proof. First, we show that $\operatorname{int}(\mathsf{A})$ is open. If $p \in \operatorname{int}(\mathsf{A})$, there is $\delta > 0$ such that $\mathcal{B}(p,\delta) \subseteq \mathsf{A}$. As in the proof of Proposition 21, we find for any $q \in \mathcal{B}(p,\delta)$ a $\beta > 0$ such that $\mathcal{B}(q,\beta) \subseteq \mathsf{A}$. But this implies $\mathcal{B}(p,\delta) \subseteq \operatorname{int}(\mathsf{A})$, and hence $\operatorname{int}(\mathsf{A})$ is an open set.

Furthermore, if $\mathsf{U} \subseteq \mathsf{A}$ is open, it is clear by definition that $\mathsf{U} \subseteq \operatorname{int}(\mathsf{A})$, thus $\operatorname{int}(\mathsf{A})$ is the largest open set contained in A. □

◇ **Exercise 33.** Show that

(a) the union of any collection of open subsets of $\mathbb{R}^m$ is again open in $\mathbb{R}^m$;
(b) the intersection of finitely many open subsets of $\mathbb{R}^m$ is open in $\mathbb{R}^m$.

Let $\emptyset \neq \mathsf{A} \subseteq \mathbb{R}^m$. An **open neighborhood** of A is an open set containing A, and a **neighborhood** of A is any set containing an open neighborhood of A. A neighborhood of a set $\{p\}$ is also called a neighborhood of the point $p \in \mathbb{R}^m$. (Note that $p \in \mathsf{A} \subseteq \mathbb{R}^m$ is an interior point of A if and only if A is a neighborhood of p.)

Definition 48. A set F is said to be **closed** if its *complement* $\mathsf{F}^c := \mathbb{R}^m \setminus \mathsf{F}$ is open.

The empty set is closed, and so is the entire space $\mathbb{R}^m$.

Proposition 23. *For every $p \in \mathbb{R}^m$ and $\delta > 0$, the set $\overline{\mathcal{B}}(p,\delta) := \{x \in \mathbb{R}^m : \|x - p\| \leq \delta\}$ is closed. ($\overline{\mathcal{B}}(p,\delta)$ is the **closed ball** of center p and radius δ.)*

Proof. For arbitrary $q \in \overline{\mathcal{B}}(p,\delta)^c$ set $\beta = \|p-q\|$, then $\beta - \delta > 0$. So $\mathcal{B}(q, \beta-\delta) \subseteq \overline{\mathcal{B}}(p,\delta)^c$, because by the reverse triangle inequality (this is Problem 11), for every $x \in \mathcal{B}(q,\beta-\delta)$

$$\|p - x\| \geq \|p - q\| - \|x - q\| > \beta - (\beta - \delta) = \delta.$$

This proves that $\overline{\mathcal{B}}(p,\delta)^c$ is open. □

Definition 49. A point $p \in \mathbb{R}^m$ is said to be a **cluster point** of a set $\mathsf{A} \subseteq \mathbb{R}^m$ if for every $\delta > 0$ we have $\mathcal{B}(p,\delta) \cap \mathsf{A} \neq \emptyset$. The set of cluster points of A is called the **closure** of A and is denoted by $\mathrm{cl}(\mathsf{A})$.

Proposition 24. *Let $\mathsf{A} \subseteq \mathbb{R}^m$. Then $\mathrm{cl}(\mathsf{A})^c = \mathrm{int}(A^c)$; in particular, the closure of A is a closed set. Moreover, $\mathrm{int}(\mathsf{A})^c = \mathrm{cl}(\mathsf{A}^c)$.*

Proof. Note that $\mathsf{A} \subseteq \mathrm{cl}(\mathsf{A})$. To say that x is *not* a cluster point of A means that it is an interior point of A^c. Thus $\mathrm{cl}(\mathsf{A})^c = \mathrm{int}(\mathsf{A}^c)$, or $\mathrm{cl}(\mathsf{A}) = \mathrm{int}(\mathsf{A}^c)^c$, which implies that $\mathrm{cl}(\mathsf{A})$ is closed in $\mathbb{R}^m$.

Furthermore, by applying this identity to A^c we obtain that $\mathrm{int}(\mathsf{A})^c = \mathrm{cl}(\mathsf{A}^c)$. □

By taking complements of sets we immediately obtain the following result.

Proposition 25. *For any $\mathsf{A} \subseteq \mathbb{R}^m$, the closure $\mathrm{cl}(\mathsf{A})$ is the smallest closed set containing A.*

From set theory we recall De Morgan's Laws, which state, for arbitrary collections $(\mathsf{A}_i)_{i\in I}$ of sets $\mathsf{A}_i \subseteq \mathbb{R}^m$, that

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \quad \text{and} \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

In view of these laws and **Exercise 33** we find, by taking complements of sets,

Proposition 26.

(a) *The intersection of any collection of closed subsets of $\mathbb{R}^m$ is again closed in $\mathbb{R}^m$.*

(b) *The union of finitely many closed subsets of $\mathbb{R}^m$ is closed in $\mathbb{R}^m$.*

3.3. **Continuity.** Let $\mathsf{U} \subseteq \mathbb{R}^m$ be an *open* set. A mapping $F : \mathsf{U} \to \mathbb{R}^n$ is **continuous** at $p \in \mathsf{U}$ if given $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$F\left(\mathcal{B}(p,\delta)\right) \subseteq \mathcal{B}(F(p),\varepsilon).$$

In other words, F is continuous at p if points arbitrarily close to $F(p)$ are images of points sufficiently close to p. We say that F is **continuous** provided it is continuous at each $p \in \mathsf{U}$.

NOTE : Equivalently, F is continuous at $p \in \mathsf{U}$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $\|F(x) - F(p)\| < \varepsilon$ for $\|x - p\| < \delta$. This simply means that $\lim_{x\to p} F(x) = F(p)$.

A mapping $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ determines n $\mathbb{R}$-valued functions (of m variables) as follows. Let $x = (x_1, \dots, x_m) \in \mathsf{U}$ and $F(x) = (y_1, \dots, y_n)$. Then we can write

$$y_1 = F_1(x_1, \dots, x_m), \quad y_2 = F_2(x_1, \dots, x_m), \quad \dots, \quad y_n = F_n(x_1, \dots, x_m).$$

The functions $F_i : \mathsf{U} \to \mathbb{R}$, $i = \overline{1,n}$ are the **component functions** of F. The continuity of the mapping F is equivalent to the continuity of its component functions.

◇ **Exercise 34.** Prove that a mapping $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is continuous if and only if each component function $F_i : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}$ is continuous.

The following results are standard (and easy to prove).

PROPOSITION 27. *Let $F, G : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be continuous mappings and let $\lambda \in \mathbb{R}$. Then $F + G$, λF, and $F \bullet G$ are each continuous. If $n = 1$ and $G(x) \neq 0$ for all $x \in \mathsf{U}$, then the quotient $\frac{F}{G}$ is also continuous.*

PROPOSITION 28. *Let $F : \mathsf{U} \subseteq \mathbb{R}^\ell \to \mathbb{R}^m$ and $G : \mathsf{V} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be continuous mappings, where U and V are open sets such that $F(\mathsf{U}) \subseteq \mathsf{V}$. Then $G \circ F$ is a continuous mapping.*

◇ **Exercise 35.** Show that the following mappings are continuous.

(a) The *identity mapping* $\mathbf{1}_{\mathbb{R}^m} : \mathbb{R}^m \to \mathbb{R}^m, \quad x \mapsto x.$
(b) The *norm function* $\nu : \mathbb{R}^m \to \mathbb{R}, \quad x \mapsto \|x\|.$
(c) The i^{th} *natural projection* $\text{pr}_i : \mathbb{R}^m \to \mathbb{R}, \quad x \mapsto x_i.$

Hence derive that every *polynomial function* (in several variables)

$$p_k : \mathbb{R}^m \to \mathbb{R}, \quad x = (x_1, \dots, x_m) \mapsto \sum_{\substack{i_1,\dots,i_m=0 \\ i_1+\cdots+i_m \leq k}}^{k} a_{i_1\dots i_m} x_1^{i_1} \dots x_m^{i_m}$$

is continuous.

Note : More generally, every *rational function* (i.e., a quotient of two polynomial functions) is continuous. It can be shown that *elementary* functions like exp, log, sin, and cos are also continuous.

Linear mappings $L : \mathbb{R}^m \to \mathbb{R}^n$ play an important role in differentiation. Such mappings are continuous.

$\diamond$ **Exercise 36.** Show that every linear mapping $L : \mathbb{R}^m \to \mathbb{R}^n$ is continuous.

In most applications it is convenient to express continuity in terms of neighborhoods instead of open balls.

$\diamond$ **Exercise 37.** Prove that a mapping $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is continuous at $p \in \mathsf{U}$ if and only if given a neighborhood $\mathcal{N}$ of $F(p)$ in $\mathbb{R}^n$ there exists a neighborhood $\mathcal{M}$ of p in $\mathbb{R}^m$ such that $F(\mathcal{M}) \subseteq \mathcal{N}$.

It is often necessary to deal with mappings (functions) defined on arbitrary (i.e., not necessarily open) sets. To extend the previous ideas to this situation, we shall proceed as follows.

Let $F : \mathsf{A} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be a mapping, where A is an *arbitrary* set. We say that F is **continuous** on A provided there exists an open set $\mathcal{U} \subseteq \mathbb{R}^m$ containing A, and a continuous mapping $\overline{F} : \mathcal{U} \to \mathbb{R}^n$ such that (the restriction) $\overline{F}\big|_{\mathsf{A}} = F$. In other words, F is continuous on A if it is the restriction of a continuous mapping defined on an open neighborhood of A.

Note : It is clear that if $F : \mathsf{A} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is continuous and $p \in \mathsf{A}$, then given a neighborhood $\mathcal{N}$ of $F(p)$ in $\mathbb{R}^n$, there exists a neighborhood $\mathcal{M}$ of p in $\mathbb{R}^m$ such that $F(\mathcal{M} \cap \mathsf{A}) \subseteq \mathcal{N}$. For this reason, it is convenient to call the set $\mathcal{M} \cap \mathsf{A}$ a *neighborhood* of p in A.

Example 50. An important class of continuous mappings is formed by the mappings $F : \mathsf{A} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ that are **Lipschitz continuous**, i.e., for which there exists $k > 0$ such that

$$\|F(x) - F(y)\| \leq k\,\|x - y\| \qquad (x, y \in \mathsf{A}).$$

Such a number k is called a *Lipschitz constant* for F. For example, the norm function $\nu : x \mapsto \|x\|$ is a Lipschitz continuous on $\mathbb{R}^m$ with Lipschitz constant 1.

$\diamond$ **Exercise 38.** Consider a mapping $F : \mathsf{A} \to \mathbb{R}^n$, where $\mathsf{A} \subseteq \mathbb{R}^m$ is an *arbitrary* set. Show that the following statements are logically equivalent.

(a) F is continuous.

(b) $F^{-1}(\mathsf{O})$ is open in A for every open set O in $\mathbb{R}^n$. (In particular, if A is open in $\mathbb{R}^m$ then: $F^{-1}(\mathsf{O})$ is open in $\mathbb{R}^m$ for every open set O in $\mathbb{R}^n$.)

(c) $F^{-1}(\mathsf{F})$ is closed in A for every closed set F in $\mathbb{R}^n$. (In particular, if A is closed in $\mathbb{R}^m$ then: $F^{-1}(\mathsf{F})$ is closed in $\mathbb{R}^m$ for every closed set F in $\mathbb{R}^n$.)

(A subset $\mathsf{U} \subseteq \mathsf{A}$ is said to be *open in* A if there is an open set W such that $\mathsf{U} = \mathsf{A} \cap \mathsf{W}$. Likewise, a subset V is said to be *closed in* A if there exists a closed set W such that $\mathsf{V} = \mathsf{A} \cap \mathsf{W}$.)

Definition 51. A set $\mathsf{A} \subseteq \mathbb{R}^m$ is said to be **disconnected** if there exist open sets U and V in $\mathbb{R}^m$ such that

$$\mathsf{A} \cap \mathsf{U} \neq \emptyset, \quad \mathsf{A} \cap \mathsf{V} \neq \emptyset, \quad (\mathsf{A} \cap \mathsf{U}) \cap (\mathsf{A} \cap \mathsf{V}) = \emptyset, \quad (\mathsf{A} \cap \mathsf{U}) \cup (\mathsf{A} \cap \mathsf{V}) = \mathsf{A}.$$

(In other words, A is the union of two disjoint non-empty subsets that are open in A.) The set A is said to be **connected** if A is *not* disconnected.

It is not difficult to prove that *the only connected subsets of* $\mathbb{R}$ *are the intervals: open, closed or half-open* (these include the singletons and the set $\mathbb{R}$ itself). The following result then follows (this is Problem 14):

Theorem 29 (Intermediate Value Theorem). *Let* $\mathsf{A} \subseteq \mathbb{R}^m$ *be connected and let* $F : \mathsf{A} \to \mathbb{R}$ *be a continuous function. Then* $F(\mathsf{A})$ *is an interval in* $\mathbb{R}$*; in particular,* F *takes all values between any two that it assumes.*

Definition 52. We say that a continuous mapping $F : \mathsf{A} \subseteq \mathbb{R}^m \to \mathbb{R}^m$ is a **homeomorphism** onto $F(\mathsf{A})$ if F is one-to-one and the inverse $F^{-1} : F(\mathsf{A}) \subseteq \mathbb{R}^m \to \mathbb{R}^m$ is continuous. In this case A and $F(\mathsf{A})$ are *homeomorphic* sets.

Example 53. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be given by

$$F(x_1, x_2, x_3) = (ax_1, bx_2, cx_3), \qquad a, b, c \in \mathbb{R} \setminus \{0\}.$$

F is clearly continuous, and the restriction of F to the (unit) *sphere*

$$\mathbb{S}^2 = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \;:\; x_1^2 + x_2^2 + x_3^2 = 1\right\}$$

is a continuous mapping $\widetilde{F} : \mathbb{S}^2 \to \mathbb{R}^3$. Observe that $\widetilde{F}(\mathbb{S}^2) = \mathbb{E}$, where $\mathbb{E}$ is the *ellipsoid*

$$\mathbb{E} = \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 \;:\; \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} + \frac{x_3^2}{c^2} = 1\right\}.$$

It is also clear that F is one-to-one and that

$$F^{-1}(x_1, x_2, x_3) = \left(\frac{x_1}{a}, \frac{x_2}{b}, \frac{x_3}{c}\right)\cdot$$

Thus $\widetilde{F}^{-1} = F^{-1}|_{\mathbb{E}}$ is continuous. Therefore, $\widetilde{F}$ is a *homeomorphism* of the sphere $\mathbb{S}^2$ onto the ellipsoid $\mathbb{E}$.

Note : There is a class of infinite sets, called *compact sets*, that in certain limited aspects behave very much like finite sets. A set $\mathsf{K} \subseteq \mathbb{R}^m$ is said to be **sequentially compact** if every sequence of elements in K contains a subsequence which *converges* to a point in K. (A sequence $(x_k)_{k\in\mathbb{N}}$ of elements $x_k \in \mathbb{R}^m$ is said to be **convergent**, with *limit* $p \in \mathbb{R}^m$, if $\lim_{k\to\infty} \|x_k - p\| = 0$, which is a limit of numbers in $\mathbb{R}$. Recall that this limit means: for every $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that $\|x_k - p\| < \varepsilon$ for $k \geq N$. In this case we write $\lim_{k\to\infty} x_k = p$.) It follows immediately that *a subset of a sequentially compact set* $\mathsf{K} \subseteq \mathbb{R}^m$ *is sequentially compact* <u>*if and only if*</u> *it is closed in* K.

Continuous mapping do not necessarily preserve closed sets; on the other hand, they do preserve (sequentially) compact sets. (In this sense compact and finite sets behave similarly: the image of a finite set under a mapping is a finite set too.) More precisely, *if* $\mathsf{K} \subset \mathbb{R}^m$ *is a sequentially compact set and* $F : \mathbb{R}^m \to \mathbb{R}^n$ *is continuous, then* $F(\mathsf{K}) \subset \mathbb{R}^n$ *is sequentially compact.* The following characterization is very useful: *A set* $\mathsf{K} \subset \mathbb{R}^m$ *is sequentially compact* <u>*if and only if*</u> *it is bounded and closed.* (A set $\mathsf{A} \subset \mathbb{R}^m$ is **bounded** if there exists a number $k > 0$ such that $\|x\| \leq k$ for all $x \in \mathsf{A}$; equivalently, if there exists a number $k > 0$ such that $\mathsf{A} \subseteq \overline{\mathcal{B}}(0,k)$.)

There is an alternative, more general, definition of compactness for sets. A subset $\mathsf{K} \subset \mathbb{R}^m$ is said to be **compact** if every open covering of K contains a finite subcovering of K. (A collection $(\mathsf{O}_i)_{i\in I}$ of open sets in $\mathbb{R}^m$ is said to be an *open covering* of a set $\mathsf{K} \subseteq \mathbb{R}^m$ if $\mathsf{K} \subseteq \bigcup_{i\in I} \mathsf{O}_i$.)

In spaces like $\mathbb{R}^m$, however, the two definitions of compactness coincide; this is a consequence of the following result.

(Heine-Borel Theorem) *A set* $\mathsf{K} \subset \mathbb{R}^m$ *is compact* <u>*if and only if*</u> *it is bounded and closed.*

3.4. **Differentiation.** Let U be an open subset of $\mathbb{R}^m$ and let $p \in \mathsf{U}$. A function $F : \mathsf{U} \to \mathbb{R}$ is **differentiable** at p if there exists a linear functional $L_p : \mathbb{R}^m \to \mathbb{R}$ such that

$$\lim_{x\to p} \frac{F(x) - F(p) - L_p(x-p)}{\|x-p\|} = 0$$

or, equivalently, if there exist a linear functional $L_p : \mathbb{R}^m \to \mathbb{R}$ and a function $R(\cdot, p)$, defined on an open neighborhood $\mathcal{V}$ of p, such that

$$F(x) = F(p) + L_p(x-p) + \|x-p\| \cdot R(x,p), \qquad x \in \mathcal{V}$$

and

$$\lim_{x\to p} R(x,p) = 0.$$

Then L_p is called a **derivative** (or differential) of F at p. We say that F is **differentiable** provided it is differentiable at each $p \in \mathsf{U}$.

Note : We think of a derivative L_p as a "linear" *approximation* of F near p. By the definition, the error involved in replacing $F(x)$ by $F(p) + L_p(x-p)$ (this is an *affine* map) is negligible compared to the distance from x to p, provided that this distance is sufficiently small.

If $L_p(x) = b_1x_1 + \cdots + b_mx_m$ is a derivative of F at p, then

$$b_i = \frac{\partial F}{\partial x_i}(p) := \lim_{t\to 0} \frac{1}{t}\left(F(p+te_i) - F(p)\right), \quad i = \overline{1,m}.$$

In particular, if F is differentiable at p, these *partial derivatives* exist and the derivative L_p is *unique.* We denote by $DF(p)$ (or sometimes $F'(p)$) *the* derivative of F at p, and write (by a slight abuse of notation)

$$DF(p) = \frac{\partial F}{\partial x_1}(p)(x_1 - p_1) + \frac{\partial F}{\partial x_2}(p)(x_2 - p_2) + \cdots + \frac{\partial F}{\partial x_m}(p)(x_m - p_m).$$

◇ **Exercise 39.** Show that any linear functional $F : \mathbb{R}^m \to \mathbb{R}$ is differentiable and $DF(p) = F$ for all $p \in \mathbb{R}^m$.

◇ **Exercise 40.** Prove that any differentiable function $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}$ is continuous.

Note : Mere existence of partial derivatives is *not* sufficient for differentiability (of the function F). For example, the function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$F(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2} \quad \text{and} \quad F(0,0) = 0$$

is *not* continuous at $(0,0)$, yet both partial derivatives are defined there. However, *if all partial derivatives* $\frac{\partial F}{\partial x_i}$, $i = \overline{1,m}$ *are defined and continuous in a neighborhood of* p, *then* F *is differentiable at* p.

If the function $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}$ has all partial derivatives continuous (on U) we say that F is **continuously differentiable** (or of *class* C^1) on U. We denote this class of functions by $C^1(\mathsf{U})$. (The class of continuous functions on U is denoted by $C^0(\mathsf{U})$.)

Note : We have seen that

$$F \in C^1(\mathsf{U}) \Rightarrow F \text{ is differentiable (on } \mathsf{U}) \Rightarrow \text{all partial derivatives } \frac{\partial F}{\partial x_i} \text{ exist (on } \mathsf{U})$$

but the converse implications may fail. Many results actually need F to be of class C^1 rather than differentiable.

If $r \geq 1$, the class $C^r(\mathsf{U})$ of functions $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}$ that are *r-fold continuously differentiable* (or C^r functions) is specified inductively by requiring that the partial derivatives of F exist and belong to $C^{r-1}(\mathsf{U})$. If F is of class C^r for all r, then we say that F is of *class* C^∞ or simply **smooth**. The class of smooth functions on U is denoted by $C^\infty(\mathsf{U})$.

Note : If $F \in C^r(\mathsf{U})$, then (at any point of U) the value of the partial derivatives of order k, $1 < k \leq r$ is independent of the order of differentiation; that is, if $(j_1, \ldots, j_k)$ is a permutation of $(i_1, \ldots, i_k)$, then

$$\frac{\partial^k F}{\partial x_{i_1} \ldots \partial x_{i_k}} = \frac{\partial^k F}{\partial x_{j_1} \ldots \partial x_{j_k}}.$$

We are now interested in extending the notion of differentiability to mappings $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$. We say that F is **differentiable** at $p \in \mathsf{U}$ if its component functions are differentiable at p; that is, by writing

$$F(x_1,\dots,x_m) = (F_1(x_1,\dots,x_m),\dots,F_n(x_1,\dots,x_m))$$

the functions $F_i : \mathsf{U} \to \mathbb{R}$, $i = \overline{1,n}$ have partial derivatives at $p \in \mathsf{U}$. F is **differentiable** provided it is differentiable at each $p \in \mathsf{U}$. (For the case $m = 1$, we obtain the notion of a differentiable parametrized curve in Euclidean space $\mathbb{R}^n$.)

The class $C^r(\mathsf{U},\mathbb{R}^n)$, $1 \leq r \leq \infty$ of C^r-mappings $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is defined in the obvious way. We will be concerned primarily with *smooth* (i.e., of class C^∞) mappings. So if F is a smooth mapping, then its component functions F_i have continuous partial derivatives of all orders and each such derivative is independent of the order of differentiation.

NOTE : Let us define a (geometric) **tangent vector** at $p \in \mathbb{R}^m$ as an ordered pair (p,v). As a matter of notation, we will abbreviate (p,v) as v_p. We think of v_p as the *vector* v with its *initial point* at p. (In other words, $p+v$ is considered as the "position vector" of a point; we shall always picture v_p as the "arrow" from the point p to the point $p+v$.) Clearly, two tangent vectors v_p and w_q are *equal* if $v = w$ and $p = q$. (It is essential to recognize that v_p and v_q are *different* tangent vectors if $p \neq q$.)

The set $\{p\} \times \mathbb{R}^m$ of all tangent vectors at p is denoted by $T_p\,\mathbb{R}^m$, and is called the **tangent space** of $\mathbb{R}^m$ at p. Thus

$$T_p\,\mathbb{R}^m := \{v_p = (p,v) \,:\, p,v \in \mathbb{R}^m\}\,.$$

This set is a *vector space* over $\mathbb{R}$ (obviously isomorphic to $\mathbb{R}^m$ itself) under the natural operations: $v_p + w_p := (v+w)_p$ and $\lambda\, v_p := (\lambda v)_p$. The tangent vectors $(e_i)_p$, $i = \overline{1,m}$ form a basis for $T_p\,\mathbb{R}^m$. (In fact, as a vector space, $T_p\,\mathbb{R}^m$ is essentially the same as $\mathbb{R}^m$ itself; the only reason we add T_p is so that the geometric tangent spaces $T_p\,\mathbb{R}^m$ and $T_q\,\mathbb{R}^m$ at distinct points p and q be disjoint sets.)

Let v_p be a tangent vector in $\mathbb{R}^m$. One can associate with it the function (parametrized line)

$$\mathbb{R} \ni t \mapsto p + tv \in \mathbb{R}^m.$$

If $F : \mathbb{R}^m \to \mathbb{R}$ is a differentiable function, then $t \mapsto F(p+tv)$ is an ordinary function $\mathbb{R} \to \mathbb{R}$. (The derivative of this function at $t = 0$ tells the *initial rate of change* of F as p moves in the v direction.) The number

$$v_p[F] := \left.\frac{d}{dt}F(p+tv)\right|_{t=0}$$

is called the **directional derivative** of F with respect to v_p. We have

$$v_p[F] = v_1\frac{\partial F}{\partial x_1}(p) + \cdots + v_m\frac{\partial F}{\partial x_m}(p) \qquad (v = (v_1,\dots,v_m) \in \mathbb{R}^m).$$

The map $v_p[\cdot] : C^\infty(\mathbb{R}^m) \to \mathbb{R}$, $F \mapsto v_p[F]$ is linear and satisfies the *Leibniz rule* (i.e., $v_p[FG] = v_p[F]G(p) + F(p)v_p[G]$ for $F,G \in C^\infty(\mathbb{R}^m)$); such a mapping is called a **derivation** at p. So *any geometric tangent vector* v_p *defines a derivation* $v_p[\cdot]$ *at* p. In fact, each derivation at p is defined by a *unique* geometric tangent vector (at p). Moreover, for any $p \in \mathbb{R}^m$, the correspondence $v_p \mapsto v_p[\cdot]$ is an isomorphism from the tangent space $T_p\,\mathbb{R}^m$ to the vector space of all derivations on p. It is customary (and convenient) to denote the derivation $(e_i)_p[\cdot]$ by $\frac{\partial}{\partial x_i}\Big|_p$; thus, $\frac{\partial}{\partial x_i}\Big|_p [F] = \frac{\partial F}{\partial x_i}(p)$.

Let $T_p\,\mathbb{R}^m$ be the tangent space to $\mathbb{R}^m$ at p; this vector space can be *identified* with $\mathbb{R}^m$ via

$$v_1\,\frac{\partial}{\partial x_1}\bigg|_p + \cdots + v_m\,\frac{\partial}{\partial x_m}\bigg|_p \mapsto (v_1, \cdots, v_m).$$

Let $\alpha : \mathsf{U} \subseteq \mathbb{R} \to \mathbb{R}^m$ be a smooth (parametrized) curve with component functions $\alpha_1, \ldots, \alpha_m$. The **velocity vector** (or tangent vector) to α at $t \in \mathsf{U}$ is the element

$$\dot{\alpha}(t) := \left(\frac{d\alpha_1}{dt}(t), \cdots, \frac{d\alpha_m}{dt}(t)\right) \in T_{\alpha(t)}\,\mathbb{R}^m.$$

Example 54. Given a point $p \in \mathsf{U} \subseteq \mathbb{R}^m$ and a tangent vector $v \in T_p\,\mathbb{R}^m$, we can always find a smooth curve $\alpha : (-\varepsilon, \varepsilon) \to \mathsf{U}$ with $\alpha(0) = p$ and $\dot{\alpha}(0) = v$. Simply define $\alpha(t) = p + tv$, $t \in (-\varepsilon, \varepsilon)$. By writing $p = (p_1, \ldots, p_m)$ and $v = (v_1, \ldots, v_m)$, the component functions of α are $\alpha_i(t) = p_i + tv_i$, $i = \overline{1, m}$. Thus α is smooth, $\alpha(0) = p$ and

$$\dot{\alpha}(0) = \left(\frac{d\alpha_1}{dt}(0), \cdots, \frac{d\alpha_m}{dt}(0)\right) = (v_1, \ldots, v_m) = v.$$

We shall now introduce the concept of *derivative* (or differential) of a differentiable mapping. Let $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be a differentiable mapping. To each $p \in \mathsf{U}$ we associate a linear mapping

$$DF(p) : \mathbb{R}^m = T_p\,\mathbb{R}^m \to \mathbb{R}^n = T_{F(p)}\,\mathbb{R}^n$$

which is called the **derivative** (or *differential*) of F at p and is defined as follows. Let $v \in T_p\,\mathbb{R}^m$ and let $\alpha : (-\varepsilon, \varepsilon) \to \mathsf{U}$ be a differentiable curve such that $\alpha(0) = p$ and $\dot{\alpha}(0) = v$. By the chain rule (for functions), the curve $\beta = F \circ \alpha : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ is also differentiable. Then

$$DF(p) \cdot v := \dot{\beta}(0).$$

Note : The above definition of $DF(p)$ does not depend on the choice of the curve which passes through p with tangent vector v, and $DF(p)$ is, in fact, linear. So

$$DF(p) \cdot v = \frac{d}{dt} F(\alpha(t))\bigg|_{t=0} \in T_{F(p)}\,\mathbb{R}^n = \mathbb{R}^n.$$

The derivative $DF(p)$ is also denoted by $T_p F$ and called the *tangent mapping* of F at p.

The matrix of the linear mapping $DF(p)$ (relative to bases $\left(\left.\frac{\partial}{\partial x_1}\right|_p, \ldots, \left.\frac{\partial}{\partial x_m}\right|_p\right)$ of $T_p\,\mathbb{R}^m$ and $\left(\left.\frac{\partial}{\partial y_1}\right|_{F(p)}, \ldots, \left.\frac{\partial}{\partial y_n}\right|_{F(p)}\right)$ of $T_{F(p)}\,\mathbb{R}^n$) is the **Jacobian matrix**

$$\frac{\partial F}{\partial x}(p) = \frac{\partial(F_1, \ldots, F_n)}{\partial(x_1, \ldots, x_m)}(p) := \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(p) & \cdots & \frac{\partial F_1}{\partial x_m}(p) \\ \vdots & & \vdots \\ \frac{\partial F_n}{\partial x_1}(p) & \cdots & \frac{\partial F_n}{\partial x_m}(p) \end{bmatrix} \in \mathbb{R}^{n \times m}$$

of F at p. When $m = n$ this is a square matrix and its determinant is then defined. This determinant is called the **Jacobian** of F at p and is denoted by $J_F(p)$. Thus

$$J_F(p) = \left|\frac{\partial F}{\partial x}(p)\right| := \det \frac{\partial F}{\partial x}(p)\cdot$$

◇ **Exercise 41.** Let $f : \mathsf{I} \to \mathbb{R}$ and $g : \mathsf{J} \to \mathbb{R}$ be differentiable functions, where I and J are open intervals such that $f(\mathsf{I}) \subseteq \mathsf{J}$. Show that the function $g \circ f$ is differentiable and (for $t \in \mathsf{I}$)

$$(g \circ f)'(t) = g'(f(t)) \cdot f'(t).$$

The standard *chain rule* (for scalar-valued) functions extends to (vector-valued) mappings.

Proposition 30 (General Chain Rule). *Let $F : \mathsf{U} \subseteq \mathbb{R}^\ell \to \mathbb{R}^m$ and $G : \mathsf{V} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ be differentiable mappings, where U and V are open sets such that $F(\mathsf{U}) \subseteq \mathsf{V}$. Then $G \circ F$ is a differentiable mapping and (for $p \in \mathsf{U}$)*

$$D(G \circ F)(p) = DG(F(p)) \circ DF(p).$$

Proof. The fact that $G \circ F$ is differentiable is a consequence of the chain rule for functions. Now, let $v \in T_p\,\mathbb{R}^\ell$ be given and let us consider a (differentiable) curve $\alpha : (-\varepsilon, \varepsilon) \to \mathsf{U}$ with $\alpha(0) = p$ and $\dot{\alpha}(0) = v$. Set $DF(p) \cdot v = w$ and observe that

$$DG(F(p)) \cdot w = \left.\frac{d}{dt}(G \circ F \circ \alpha)\right|_{t=0}.$$

Then

$$\begin{aligned} D(G \circ F)(p) \cdot v &= \left.\frac{d}{dt}(G \circ F \circ \alpha)\right|_{t=0} \\ &= DG(F(p)) \cdot w \\ &= DG(F(p)) \circ DF(p) \cdot v. \end{aligned}$$

□

Note : In terms of Jacobian matrices, the general chain rule can be written

$$\frac{\partial(G \circ F)}{\partial x}(p) = \frac{\partial G}{\partial y}(F(p)) \cdot \frac{\partial F}{\partial x}(p)\cdot$$

Thus if $H = G \circ F$ and $y = F(x)$, then

$$\frac{\partial H}{\partial x} = \begin{bmatrix} \frac{\partial G_1}{\partial y_1} & \cdots & \frac{\partial G_1}{\partial y_m} \\ \vdots & & \vdots \\ \frac{\partial G_n}{\partial y_1} & \cdots & \frac{\partial G_n}{\partial y_m} \end{bmatrix} \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \cdots & \frac{\partial F_1}{\partial x_\ell} \\ \vdots & & \vdots \\ \frac{\partial F_m}{\partial x_1} & \cdots & \frac{\partial F_m}{\partial x_\ell} \end{bmatrix}$$

where $\dfrac{\partial G_1}{\partial y_1}, \dots, \dfrac{\partial G_n}{\partial y_m}$ are evaluated at $y = F(x)$ and $\dfrac{\partial F_1}{\partial x_1}, \cdots, \dfrac{\partial F_m}{\partial x_\ell}$ at x. Writing this out, we obtain

$$\frac{\partial H_i}{\partial x_j} = \frac{\partial G_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial G_i}{\partial y_m}\frac{\partial y_m}{\partial x_j} \qquad (i = \overline{1,n}\, ;\, j = \overline{1,\ell}).$$

◇ **Exercise 42.** Let

$$F(x_1, x_2) = (x_1^2 - x_2^2 + x_1x_2, x_2^2 - 1) \quad \text{and} \quad G(y_1, y_2) = (y_1 + y_2, 2y_1, y_2^2).$$

(a) Show that F and G are differentiable, and that $G \circ F$ exists.

(b) Compute $D(G \circ F)(1, 1)$

(i) directly

(ii) using the chain rule.

Note : The precise sense in which the derivative $DF(p)$ of the (differentiable) mapping F at p is an (affine) approximation of F near p is given by the following result (in which $DF(p)$ is interpreted as a linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$) : *If the mapping $F : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is differentiable, then for each $p \in \mathsf{U}$,*

$$\lim_{x \to p} \frac{\|F(x) - F(p) - DF(p) \cdot (x - p)\|}{\|x - p\|} = 0$$

or, equivalently, there exists a (local) map $\epsilon_p : \mathbb{R}^m \to \mathbb{R}^n$ satisfying, for all h with $p + h \in \mathsf{U}$,

$$(5) \qquad F(p + h) = F(p) + DF(p) \cdot h + \epsilon_p(h) \quad \textit{with} \quad \lim_{h \to 0} \frac{\|\epsilon_p(h)\|}{\|h\|} = 0.$$

The mapping $\mathbb{R}^m \to \mathbb{R}^n$, $x \mapsto F(p) + DF(p) \cdot (x - p)$ is the *best affine approximation* to F at p. (It is the unique affine approximation for which the difference mapping ϵ_p satisfies the estimate (5).)

If $\mathsf{A} \subseteq \mathbb{R}^m$ is an *arbitrary* set, then $C^\infty(\mathsf{A})$ denotes the set of all functions $F : \mathsf{A} \to \mathbb{R}$ such that $F = \overline{F}\big|_{\mathsf{A}}$, where $\overline{F} : \mathcal{U} \to \mathbb{R}$ is a smooth function on some open neighborhood $\mathcal{U}$ of A.

Problems (11–15)

(11) Let $x, y \in \mathbb{R}^m$. Prove the following inequalities.

(a) $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality).

(b) $|\,\|x\| - \|y\|\,| \leq \|x - y\|$ (reverse triangle inequality).

(c) $|x_i| \leq \|x\| \leq |x_1| + \cdots + |x_m| \leq \sqrt{m}\,\|x\|, \quad i = \overline{1,m}.$

(12) Let $\tau : \mathbb{R}^m \to \mathbb{R}^m$ be a linear transformation, and let $A \in \mathbb{R}^{m\times m}$ denote its matrix with respect to the standard basis of $\mathbb{R}^m$. Show that the following statements are logically equivalent.

(a) $\|\tau(x)\| = \|x\|$ for all $x \in \mathbb{R}^m$.

(b) $\tau(x) \bullet \tau(y) = x \bullet y$ for all $x, y \in \mathbb{R}^m$.

(c) $A^\top A = \mathbf{1}$ (i.e., the matrix A is orthogonal).

(Such a linear transformation is called an **orthogonal transformation**.) Hence deduce that such a linear transformation τ is invertible. Is τ^{-1} of the same sort?

(13) Let F be a subset of $\mathbb{R}^m$. Show that the following statements are logically equivalent.

(a) F is closed.

(b) $\mathsf{F} = \mathrm{cl}\,(\mathsf{F})$.

(c) For every sequence $(x_k)_{k\in\mathbb{N}}$ of points $x_k \in \mathbb{R}^m$ that is convergent to a limit, say p, we have $p \in \mathsf{F}$.

(14) Let $\mathsf{A} \subseteq \mathbb{R}^m$ be an arbitrary set.

(a) Show that the following statements are logically equivalent.

(i) A is disconnected.

(ii) There exists a surjective continuous function $\mathsf{A} \to \{0, 1\}$.

(Recall the definition of the **characteristic function** χ_{A} of a set A: $\chi_{\mathsf{A}}(x) = 1$ if $x \in \mathsf{A}$ and $\chi_{\mathsf{A}}(x) = 0$ if $x \notin \mathsf{A}$.)

(b) Assume that A is connected and let $F : \mathsf{A} \to \mathbb{R}^n$ be a continuous mapping. Show that $F(\mathsf{A})$ is connected in $\mathbb{R}^n$.

(c) Let A be connected and let $F : \mathsf{A} \to \mathbb{R}$ be a continuous function. Show that $F(\mathsf{A})$ is an interval in $\mathbb{R}$; in particular, F takes all the values between any two that it assumes.

(15) Show that

(i) if $\sigma : \mathbb{R}^2 \to \mathbb{R}$ is defined by $\sigma(x, y) = x + y$, then $D\sigma(a, b) = \sigma$.

(ii) if $\pi : \mathbb{R}^2 \to \mathbb{R}$ is defined by $\pi(x, y) = x \cdot y$, then $D\pi(a, b) \cdot (x, y) = bx + ay$.

Hence deduce that if the functions $F, G : \mathsf{U} \subseteq \mathbb{R}^m \to \mathbb{R}$ are differentiable at $p \in \mathsf{U}$, then

$$\begin{aligned} D(F + G)(p) &= DF(p) + DG(p) \\ D(F \cdot G)(p) &= G(p)DF(p) + F(p)DG(p). \end{aligned}$$

If moreover $G(p) \neq 0$, then

$$D\left(\frac{F}{G}\right)(p) = \frac{G(p)DF(p) - F(p)DG(p)}{(G(p))^2}.$$

4. Matrix Groups

Matrix algebra • Matrix groups • Linear Lie groups: examples • Complex matrix groups as real matrix groups.

4.1. **Matrix algebra.** Throughout, we shall denote by $\Bbbk$ either the *field* $\mathbb{R}$ of real numbers or the *field* $\mathbb{C}$ of complex numbers. Let $\Bbbk^m$ be the set of all m-tuples of elements of $\Bbbk$. Under the usual addition and scalar multiplication, $\Bbbk^m$ is a *vector space* over $\Bbbk$. The set $\mathsf{Hom}\,(\Bbbk^n, \Bbbk^m)$ of all linear mappings from $\Bbbk^n$ to $\Bbbk^m$ (i.e., mappings $L : \Bbbk^n \to \Bbbk^m$ such that $L(\lambda x + \mu y) = \lambda L(x) + \mu L(y)$ for every $x, y \in \Bbbk^n$ and $\lambda, \mu \in \Bbbk$) is also a *vector space* over $\Bbbk$.

◇ **Exercise 43.** Determine the *dimension* of the vector space $\mathsf{Hom}\,(\Bbbk^n, \Bbbk^m)$.

Let $\Bbbk^{m\times n}$ be the set of all $m \times n$ matrices with elements (entries) from $\Bbbk$. Under the usual matrix addition and multiplication, $\Bbbk^{m\times n}$ is a *vector space* over $\Bbbk$. There is a natural one-to-one correspondence $A \mapsto L_A\,(: x \mapsto Ax)$ between the $m \times n$ matrices with elements from $\Bbbk$ and the linear mappings from $\Bbbk^n$ to $\Bbbk^m$.

◇ **Exercise 44.** Show that the vector spaces $\Bbbk^{m\times n}$ and $\mathsf{Hom}\,(\Bbbk^n, \Bbbk^m)$ are *isomorphic.*

In particular, the (n-dimensional) vector spaces $\Bbbk^{1\times n}$ and $\mathsf{Hom}\,(\Bbbk^n, \Bbbk) = (\Bbbk^n)^*$ (the *dual* of $\Bbbk^n$) are isomorphic. Any matrix $A \in \Bbbk^{m\times n}$ can be *interpreted* as a linear mapping $L_A \in \mathsf{Hom}\,(\Bbbk^n, \Bbbk^m)$, whereas any linear mapping $L \in \mathsf{Hom}\,(\Bbbk^n, \Bbbk^m)$ can be *realized* as a matrix $A \in \Bbbk^{m\times n}$. Henceforth we shall not distinguish notationwise between a matrix A and its corresponding linear mapping $x \mapsto Ax$.

Note : A matrix (or linear mapping, if one prefers) $A \in \Bbbk^{n\times n}$ can be viewed as a **vector field** (on $\Bbbk^n$) : A associates to each point p in $\Bbbk^n$ the tangent vector $A(p) = Ap \in \Bbbk^n$. We may think of a *fluid* in motion, so that the velocity of the fluid particles passing through p is always $A(p)$. The vector field is then the current of the *flow* and the paths of the fluid particles are the trajectories. This kind of flow is, of course, very special : $A(p)$ is independent of time, and depends linearly on p.

The (structured) set $\Bbbk^{n\times n}$ is not just a vector space. It also has a multiplication which is associative and distributes over addition (on either side). In other words, under the usual addition and multiplication, $\Bbbk^{n\times n}$ is a *ring* (in general not commutative), with identity $\mathbf{1}$. Moreover, for all $A, B \in \Bbbk^{n\times n}$ and $\lambda \in \Bbbk$,

$$\lambda(AB) = (\lambda A)B = A(\lambda B).$$

Such a structure is called an (associative) **algebra** over $\mathbb{k}$.

For $x \in \mathbb{k}^n$ $(= \mathbb{k}^{n\times 1})$, let

$$\|x\|_2 := \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

be the 2-norm (or *Euclidean norm*) on $\mathbb{k}^n$.

NOTE : For $r \geq 1$, the *r-norm* of $x \in \mathbb{k}^n$ is defined as

$$\|x\|_r := \left(|x_1|^r + |x_2|^r + \cdots + |x_n|^r\right)^{1/r}.$$

The following properties hold (for $x, y \in \mathbb{k}^n$ and $\lambda \in \mathbb{k}$) :

$$\begin{aligned}
&\|x\|_r \geq 0, \quad \text{and} \quad \|x\|_r = 0 \iff x = 0 \; ; \\
&\|\lambda x\|_r = |\lambda| \, \|x\|_r \; ; \\
&\|x + y\|_r \leq \|x\|_r + \|y\|_r.
\end{aligned}$$

In practice, only three of the r-norms are used, and they are :

$$\begin{aligned}
\|x\|_1 &= |x_1| + |x_2| + \cdots + |x_n| \quad \text{(the grid norm);} \\
\|x\|_2 &= \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} \quad \text{(the Euclidean norm);} \\
\|x\|_\infty = \lim_{r\to\infty} \|x\|_r &= \max\{|x_1|, |x_2|, \ldots, |x_n|\} \quad \text{(the max norm).}
\end{aligned}$$

For $x \in \mathbb{k}^n$, we have

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \cdot \|x\|_2 \leq n \cdot \|x\|_\infty$$

and so any two of these norms are *equivalent* (i.e., the associated metric topologies are identical). In fact, *all norms on a finite-dimensional vector space (over* $\mathbb{k}$ *) are equivalent.*

The metric topology *induced* by (the Euclidean distance) $(x, y) \mapsto \|x - y\|_2$ is the *natural topology* on the set (vector space) $\mathbb{k}^n$.

◇ **Exercise 45.** Show that, for $x, y \in \mathbb{k}^n$,

$$\big|\, \|x\|_2 - \|y\|_2 \,\big| \leq \|x - y\|_2.$$

Hence deduce that the function $\|\cdot\|_2 : \mathbb{k}^n \to \mathbb{R}, \quad x \mapsto \|x\|_2$ is *continuous* (with respect to the natural topologies on $\mathbb{k}^n$ and $\mathbb{R}$).

◇ **Exercise 46.** Given $A \in \mathbb{k}^{n\times n}$, show that the linear mapping (on $\mathbb{k}^n$) $x \mapsto Ax$ is *continuous* (with respect to the natural topology on $\mathbb{k}^n$).

Let $A \in \mathbb{k}^{n\times n}$. The 2-norm $\|\cdot\|_2$ on $\mathbb{k}^{n\times 1}$ induces a (matrix) norm on $\mathbb{k}^{n\times n}$ by setting

$$\|A\| := \max_{\|x\|_2 = 1} \|Ax\|_2.$$

The subset $K = \{x \in \mathbb{k}^n \, : \, \|x\|_2 = 1\} \subset \mathbb{k}^n$ is closed and bounded, and so is *compact.* [A subset of the metric space $\mathbb{k}^n$ is compact if and only if it is closed and bounded.] On the other hand, the function $f : K \to \mathbb{R}, \quad x \mapsto \|Ax\|_2$ is *continuous.* [The composition of two continuous maps is a continuous map.] Hence the maximum value $\max_{x\in K} \|Ax\|_2$ *must* exist.

NOTE : The following topological result holds : *If $K \subset \Bbbk^n$ is a (non-empty) compact set, then any continuous function $f : K \to \mathbb{R}$ is bounded; that is, the image set $f(K) = \{f(x) : x \in K\} \subseteq \mathbb{R}$ is bounded.* Moreover, f has a global maximum (and a global minimum).

$\diamond$ **Exercise 47.** Show that the induced norm $\|\cdot\|$ is *compatible* with its underlying norm $\|\cdot\|_2$; that is (for $A \in \Bbbk^{n\times n}$ and $x \in \Bbbk^n$),

$$\|Ax\|_2 \leq \|A\|\,\|x\|_2.$$

$\|\cdot\|$ is a *matrix norm* on $\Bbbk^{n\times n}$, called the **operator norm**; that is, it has the following four properties (for $A, B \in \Bbbk^{n\times n}$ and $\lambda \in \Bbbk$) :

(MN1) $\|A\| \geq 0$, and $\|A\| = 0 \iff A = 0$;
(MN2) $\|\lambda A\| = |\lambda|\,\|A\|$;
(MN3) $\|A + B\| \leq \|A\| + \|B\|$;
(MN4) $\|AB\| \leq \|A\|\,\|B\|$.

NOTE : There is a simple procedure (well known in numerical linear algebra) for calculating the operator norm of an $n \times n$ matrix A. This is $\|A\| = \sqrt{\lambda_{\max}}$, where $\lambda_{\max}$ is the largest eigenvalue of the matrix A^*A. Here A^* denotes the *Hermitian conjugate* (i.e., the conjugate transpose) matrix of A; in the case $\Bbbk = \mathbb{R}$, $A^* = A^\top$.

We define a *metric* ρ on (the algebra) $\Bbbk^{n\times n}$ by

$$\rho(A, B) := \|A - B\|.$$

Associated to this metric is a natural *topology* on $\Bbbk^{n\times n}$. Hence fundamental topological concepts, like *open sets, closed sets, compactness, connectedness,* as well as *continuity*, can be introduced. In particular, we can speak of continuous functions $\Bbbk^{n\times n} \to \Bbbk$.

$\diamond$ **Exercise 48.** For $1 \leq i, j \leq n$, show that the coordinate function

$$\mathrm{coord}_{ij} : \Bbbk^{n\times n} \to \Bbbk, \quad A \mapsto a_{ij}$$

is *continuous.* [HINT : Show first that $|a_{ij}| \leq \|A\|$ and then verify the defining condition for continuity.]

It follows immediately that *if $f : \Bbbk^{n^2} \to \Bbbk$ is continuous, then the associated function*

$$\widetilde{f} = f \circ (\mathrm{coord}_{ij}) : \Bbbk^{n\times n} \to \Bbbk, \quad A \mapsto f((a_{ij}))$$

is also continuous. Here $(a_{ij}) = (a_{11}, \ldots, a_{n1}, \ldots, a_{1n}, \ldots, a_{nn}) \in \Bbbk^{n^2}$.

$\diamond$ **Exercise 49.** Show that the *determinant function*

$$\det : \Bbbk^{n\times n} \to \Bbbk, \quad A \mapsto \det A := \sum_{\sigma\in\mathfrak{S}_n} (-1)^{|\sigma|} a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

and the *trace function*

$$\mathrm{tr} : \Bbbk^{n\times n} \to \Bbbk, \quad A \mapsto \mathrm{tr}\, A := \sum_{i=1}^{n} a_{ii}$$

are *continuous.*

The metric space $(\Bbbk^{n\times n}, \rho)$ is *complete.* This means that *every Cauchy sequence* $(A_r)_{r\geq 0}$ *in* $\Bbbk^{n\times n}$ *has a unique limit* $\lim\limits_{r\to\infty} A_r$. Furthermore,

$$\left(\lim_{r\to\infty} A_r\right)_{ij} = \lim_{r\to\infty} (A_r)_{ij}.$$

Indeed, the limit on the RHS exists, so it is sufficient to check that the required matrix limit is the matrix A with $a_{ij} = \lim\limits_{r\to\infty} (A_r)_{ij}$. The sequence $\left(A_r - A\right)_{r\geq 0}$ satisfies

$$\|A_r - A\| \leq \sum_{i,j=1}^{n} |(A_r)_{ij} - a_{ij}| \to 0 \quad \text{as} \;\; r \to \infty$$

and so $A_r \to A$.

4.2. **Matrix groups.** Let $\mathsf{GL}\,(n, \Bbbk)$ be the set of all invertible $n \times n$ matrices over $\Bbbk$. So

$$\mathsf{GL}\,(n, \Bbbk) := \{A \in \Bbbk^{n\times n} \; : \; \det A \neq 0\}.$$

$\diamond$ **Exercise 50.** Verify that the set $\mathsf{GL}\,(n, \Bbbk)$ is a *group* under matrix multiplication.

$\mathsf{GL}\,(n, \Bbbk)$ is called the **general linear group** over $\Bbbk$. We will refer to $\mathsf{GL}\,(n, \mathbb{R})$ and $\mathsf{GL}\,(n, \mathbb{C})$ as the *real* and *complex* general linear group, respectively. A 1×1 matrix over $\Bbbk$ is just an element of $\Bbbk$ and matrix multiplication of two such elements is just multiplication in $\Bbbk$. So we see that

$$\mathsf{GL}\,(1, \Bbbk) = \Bbbk^{\times} \quad \text{(the multiplicative group of } \Bbbk \setminus \{0\}) .$$

Any subgroup of $\mathsf{GL}\,(n, \Bbbk)$ is customarily referred to as a **linear group** or sometimes as a matrix group.

Proposition 31. *$\mathsf{GL}\,(n, \Bbbk)$ is an open subset of $\Bbbk^{n\times n}$.*

Proof. We have seen that the function $\det : \Bbbk^{n\times n} \to \Bbbk$ is continuous (see **Exercise 49**). Then observe that

$$\mathsf{GL}\,(n, \Bbbk) = \Bbbk^{n\times n} \setminus \det{}^{-1}(0).$$

Since the set $\{0\}$ is closed (in $\Bbbk$), it follows that $\det^{-1}(0) = \det^{-1}(\{0\}) \subset \Bbbk^{n\times n}$ is also closed. [The preimage of a closed set under a continuous map is a closed set.] Hence $\mathsf{GL}\,(n, \Bbbk)$ is open. [The complement of a closed set is an open set.] $\square$

We observe that the general linear group $\mathsf{GL}\,(n, \Bbbk)$ has more than just an *algebraic* structure: it has a *topological* structure as well (PROPOSITION 31). Thus we may naturally consider subsets which are not only closed in the algebraic sense (that is, subgroups), but in the topological sense as well.

DEFINITION 55. A **linear Lie group** is a closed subgroup of $\mathsf{GL}\,(n, \Bbbk)$.

Linear Lie groups are also known as *matrix* Lie groups. This terminology emphasizes the remarkable fact that *every closed linear group is a Lie group.*

NOTE : The condition that a set (group) of matrices $\mathsf{G} \subseteq \mathsf{GL}\,(n, \Bbbk)$ is a closed subset of (the metric space) $\mathsf{GL}\,(n, \Bbbk)$ means that the following condition is satisfied : *if* $(A_r)_{r\geq 0}$ *is any sequence of matrices in* G *and* $A_r \to A$*, then either* $A \in \mathsf{G}$ *or* A *is not invertible (i.e.* $A \notin \mathsf{GL}\,(n, \Bbbk)$*).* The condition that G be a *closed* subgroup, as opposed to merely a subgroup, should be regarded as a "technicality" since most of the *interesting* subgroups of $\mathsf{GL}\,(n, \Bbbk)$ have this property. Almost all of the matrix groups we will consider have the stronger property that if $(A_r)_{r\geq 0}$ is any sequence of matrices in G converging to some matrix A, then $A \in \mathsf{G}$.

We shall use the customary notation $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ to indicate that G is a subgroup of $\mathsf{GL}\,(n, \Bbbk)$.

EXAMPLE 56. The general linear group $\mathsf{GL}\,(n, \Bbbk)$ is a linear Lie group.

EXAMPLE 57. An example of a group of matrices which is <u>not</u> a linear Lie group is the set $\mathsf{GL}\,(n, \mathbb{Q})$ of all $n \times n$ invertible matrices all of whose entries are rational numbers. This is in fact a subgroup of $\mathsf{GL}\,(n, \mathbb{C})$ but not a closed subgroup; that is, one can (easily) have a sequence of invertible matrices with rational entries converging to an invertible matrix with some irrational entries.

NOTE : The *closure* of $\mathsf{GL}\,(2, \mathbb{Q})$ (in $\mathsf{GL}\,(2, \mathbb{C})$) can be thought of as (the direct product) $\mathbb{S}^1 \times \mathbb{S}^1$ and so <u>is</u> a linear Lie group (see **Exercise 61**).

PROPOSITION 32. *Let* G *be a linear Lie group and* H *a closed subgroup of* G*. Then* H *is a linear Lie group.*

Proof. Every sequence $(A_r)_{r\geq 0}$ in H with a limit in $\mathsf{GL}\,(n, \Bbbk)$ actually has its limit in G since each $A_r \in \mathsf{H} \subseteq \mathsf{G}$ and G is closed in $\mathsf{GL}\,(n, \Bbbk)$. Since H is closed in G, this means that $(A_r)_{r\geq 0}$ has a limit in H. So H is closed in $\mathsf{GL}\,(n, \Bbbk)$, showing it is a linear Lie group. □

◇ **Exercise 51.** Prove that any *intersection* of linear Lie groups is a linear Lie group.

Example 58. Denote by $\mathsf{SL}\,(n, \Bbbk)$ the set of all $n \times n$ matrices over $\Bbbk$, having determinant one. So

$$\mathsf{SL}\,(n, \Bbbk) : = \{A \in \Bbbk^{n\times n} \,:\, \det A = 1\} \subset \mathsf{GL}\,(n, \Bbbk).$$

◇ **Exercise 52.** Show that $\mathsf{SL}\,(n, \Bbbk)$ is a closed subgroup of $\mathsf{GL}\,(n, \Bbbk)$ and hence is a linear Lie group.

$\mathsf{SL}\,(n, \Bbbk)$ is called the **special linear group** over $\Bbbk$. We will refer to $\mathsf{SL}\,(n, \mathbb{R})$ and $\mathsf{SL}\,(n, \mathbb{C})$ as the *real* and *complex* special linear group, respectively.

Definition 59. A closed subgroup of a linear Lie group G is called a **linear Lie subgroup**.

Example 60. We can consider $\mathsf{GL}\,(n, \Bbbk)$ as a subgroup of $\mathsf{GL}\,(n+1, \Bbbk)$ by *identifying* the $n \times n$ matrix $A = [\,a_{ij}\,]$ with

$$\begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & a_{11} & \dots & a_{1n} \\ 0 & a_{21} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & a_{n1} & \dots & a_{nn} \end{bmatrix}.$$

It is easy to verify that $\mathsf{GL}\,(n, \Bbbk)$ is closed in $\mathsf{GL}\,(n+1, \Bbbk)$ and hence $\mathsf{GL}\,(n, \Bbbk)$ is a linear Lie subgroup of $\mathsf{GL}\,(n+1, \Bbbk)$.

◇ **Exercise 53.** Show that $\mathsf{SL}\,(n, \Bbbk)$ is a linear Lie subgroup of $\mathsf{SL}\,(n+1, \Bbbk)$.

4.3. **Linear Lie groups: examples.** The vector space $\Bbbk^{n\times n}$ over $\Bbbk$ can be considered to be a *real* vector space, of dimension n^2 or $2n^2$, respectively. Explicitly, $\mathbb{R}^{n\times n}$ is (isomorphic to) $\mathbb{R}^{n^2}$, and $\mathbb{C}^{n\times n}$ is (isomorphic to) $\mathbb{C}^{n^2} \cong \mathbb{R}^{2n^2}$. Hence we may assume, without any loss of generality, that $\Bbbk^{n\times n}$ is some Euclidean space $\mathbb{R}^m$.

4.3.1. *The real general linear group* $\mathsf{GL}\,(n, \mathbb{R})$. We have seen that $\mathsf{GL}\,(n, \mathbb{R})$ is a linear Lie group and that it is an *open* subset of the vector space $\mathbb{R}^{n\times n}\ \big(= \mathbb{R}^{n^2}\big)$. Since the set $\mathsf{GL}\,(n, \mathbb{R})$ is not closed, it is *not* compact. [Any compact set is a closed set.] The determinant function $\det : \mathsf{GL}\,(n, \mathbb{R}) \to \mathbb{R}$ is continuous (in fact, *smooth*) and maps $\mathsf{GL}\,(n, \mathbb{R})$ onto the two *components* of $\mathbb{R}^\times$. Thus $\mathsf{GL}\,(n, \mathbb{R})$ is *not* connected. [The image of a connected set under a continuous map is a connected set.]

Note : A linear Lie group G is said to be **connected** if given any two matrices $A, B \in \mathsf{G}$, there exists a continuous *path* $\gamma : [a, b] \to \mathsf{G}$ with $\gamma(a) = A$ and $\gamma(b) = B$. This property is what is called **path-connectedness** in topology, which is not (in general) the same as connectedness. However, it is a fact (not particularly obvious at the moment) that *a linear Lie group is connected* *if and only if it is path-connected*. So in a slight abuse of terminology we shall continue to refer to the above property as *connectedness*.

A linear Lie group G which is not connected can be decomposed (uniquely) as a union of several pieces, called *components*, such that two elements of the same component can be joined by a continuous path, but two elements of different components cannot. The component of G containing the identity is a closed subgroup of G and hence a connected linear Lie group.

Consider the sets

$$\begin{aligned} \mathsf{GL}^+(n,\mathbb{R}) &:= \{A \in \mathsf{GL}(n,\mathbb{R}) \,:\, \det A > 0\} \\ \mathsf{GL}^-(n,\mathbb{R}) &:= \{B \in \mathsf{GL}(n,\mathbb{R}) \,:\, \det B < 0\}. \end{aligned}$$

These two disjoint subsets of $\mathsf{GL}(n,\mathbb{R})$ are *open* and satisfy

$$\mathsf{GL}^+(n,\mathbb{R}) \cup \mathsf{GL}^-(n,\mathbb{R}) = \mathsf{GL}(n,\mathbb{R}).$$

[The preimage of an open set under a continuous map is an open set.]

◇ **Exercise 54.** Show that $\mathsf{GL}^+(n,\mathbb{R})$ is a linear Lie subgroup of $\mathsf{GL}(n,\mathbb{R})$ but $\mathsf{GL}^-(n,\mathbb{R})$ is not.

The mapping

$$A \in \mathsf{GL}^+(n,\mathbb{R}) \mapsto SA \in \mathsf{GL}^-(n,\mathbb{R})$$

where $S = \operatorname{diag}(1,\dots,1,-1)$, is a bijection (in fact, a *diffeomorphism*). The transformation $x \mapsto Sx$ may be thought of as a *reflection* in the hyperplane $\mathbb{R}^{n-1} = \mathbb{R}^{n-1} \times \{0\} \subset \mathbb{R}^n$.

Note : The group $\mathsf{GL}^+(n,\mathbb{R})$ is *connected*, which proves that $\mathsf{GL}^+(n,\mathbb{R})$ *is the connected component of the identity in* $\mathsf{GL}(n,\mathbb{R})$ and that $\mathsf{GL}(n,\mathbb{R})$ has two (connected) components.

4.3.2. *The real special linear group* $\mathsf{SL}(n,\mathbb{R})$. Recall that

$$\mathsf{SL}(n,\mathbb{R}) := \{A \in \mathsf{GL}(n,\mathbb{R}) \,:\, \det A = 1\} = \det^{-1}(1).$$

It follows that $\mathsf{SL}(n,\mathbb{R})$ is a closed subgroup of $\mathsf{GL}(n,\mathbb{R})$ and hence is a linear Lie group. [The preimage of a closed set under a continuous map is a closed set.] We introduce a new matrix norm on $\mathbb{R}^{n\times n}$, called the **Frobenius norm**, as follows :

$$\|A\|_F := \sqrt{\operatorname{tr}(A^\top A)} = \sqrt{\sum_{i,j=1}^n a_{ij}^2}.$$

Note : The Frobenius norm coincides with the Euclidean norm on $\mathbb{R}^{n^2}$, and is much easier to compute than the operator norm. However, *all* matrix norms on $\mathbb{R}^{n\times n}$ are *equivalent* (i.e., they generate the same metric topology).

We shall use this (matrix) norm to show that $\mathsf{SL}\,(n,\mathbb{R})$ is *not* compact. Indeed, all matrices of the form

$$\begin{bmatrix} 1 & 0 & \dots & t \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

are elements of $\mathsf{SL}\,(n,\mathbb{R})$ whose norm equals $\sqrt{n+t^2}$ for any $t \in \mathbb{R}$. Thus $\mathsf{SL}\,(n,\mathbb{R})$ is *not* a bounded subset of $\mathbb{R}^{n\times n}$ and hence is *not* compact. [In a metric space, any compact set is bounded.]

Note : The special linear group $\mathsf{SL}\,(n,\mathbb{R})$ is *connected.*

4.3.3. *The orthogonal and special orthogonal groups* $\mathsf{O}\,(n)$ *and* $\mathsf{SO}\,(n)$. The set

$$\mathsf{O}\,(n) := \{A \in \mathbb{R}^{n\times n} \,:\, A^\top A = \mathbf{1}\}$$

is the **orthogonal group**. Clearly, every *orthogonal matrix* $A \in \mathsf{O}\,(n)$ has an inverse, namely $A^\top$. Hence $\mathsf{O}\,(n) \subset \mathsf{GL}\,(n,\mathbb{R})$.

◇ **Exercise 55.** Verify that $\mathsf{O}\,(n)$ is a *subgroup* of the general linear group $\mathsf{GL}\,(n,\mathbb{R})$.

The single matrix equation $A^\top A = \mathbf{1}$ is equivalent to n^2 equations for the n^2 real numbers $a_{ij},\ i,j = \overline{1,n}$:

$$\sum_{k=1}^n a_{ki}a_{kj} = \delta_{ij}.$$

This means that $\mathsf{O}\,(n)$ is a *closed* subset of $\mathbb{R}^{n\times n}$ and hence of $\mathsf{GL}\,(n,\mathbb{R})$.

◇ **Exercise 56.** Prove that $\mathsf{O}\,(n)$ is a closed subset of $\mathbb{R}^{n^2}$.

Thus $\mathsf{O}\,(n)$ is a linear Lie group. The group $\mathsf{O}\,(n)$ is also *bounded* in $\mathbb{R}^{n\times n}$. Indeed, the (Frobenius) norm of $A \in \mathsf{O}\,(n)$ is

$$\|A\|_F = \sqrt{\operatorname{tr}\,(A^\top A)} = \sqrt{\operatorname{tr}\mathbf{1}} = \sqrt{n}.$$

Hence the group $\mathsf{O}\,(n)$ is *compact.* [A subset of $\mathbb{R}^{n\times n}$ is compact if and only if it is closed and bounded.] Let us consider the determinant function (restricted to $\mathsf{O}\,(n)$), $\det : \mathsf{O}\,(n) \to \mathbb{R}^\times$. Then for $A \in \mathsf{O}\,(n)$

$$\det \mathbf{1} = \det\,(A^\top A) = \det A^\top \cdot \det A = (\det A)^2.$$

Hence $\det A = \pm 1$. So we have

$$\mathsf{O}\,(n) = \mathsf{O}^+\,(n) \cup \mathsf{O}^-\,(n)$$

where

$$\mathsf{O}^+\,(n) := \{A \in \mathsf{O}\,(n) \,:\, \det A = 1\} \quad \text{and} \quad \mathsf{O}^-\,(n) := \{A \in \mathsf{O}\,(n) \,:\, \det A = -1\}.$$

Note : The group $\mathsf{O}^+(n)$ is *connected*, which proves that $\mathsf{O}^+(n)$ *is the connected component of the identity in* $\mathsf{O}(n)$.

The **special orthogonal group** is defined as

$$\mathsf{SO}(n) := \mathsf{O}(n) \cap \mathsf{SL}(n, \mathbb{R}).$$

That is,

$$\mathsf{SO}(n) = \{A \in \mathsf{O}(n) \; : \; \det A = 1\} = \mathsf{O}^+(n).$$

It follows that $\mathsf{SO}(n)$ is a closed subset of $\mathsf{O}(n)$ and hence is *compact*. [A closed subset of a compact set is compact.]

Note : One of the main reasons for the study of these groups $\mathsf{O}(n)$, $\mathsf{SO}(n)$ is their relationship with *isometries* (i.e., distance-preserving transformations on the Euclidean space $\mathbb{R}^n$). If such an isometry fixes the origin, then it is actually a linear transformation and so – with respect to the standard basis – corresponds to a matrix A. The isometry condition is equivalent to the fact that (for all $x, y \in \mathbb{R}^n$)

$$Ax \bullet Ay = x \bullet y$$

which in turn is equivalent to the condition that $A^\top A = \mathbf{1}$ (i.e., A is orthogonal). Elements of $\mathsf{SO}(n)$ are (identified with) *rotations* (or direct isometries); elements of $\mathsf{O}^-(n)$ are sometimes referred to as indirect isometries.

4.3.4. *The Lorentz group* $\mathsf{Lor}(1, n)$. Consider the inner product (i.e., non-degenerate symmetric bilinear form) $\odot$ on the vector space $\mathbb{R}^{n+1}$ given by (for $x, y \in \mathbb{R}^{n+1}$)

$$x \odot y := -x_1 y_1 + \sum_{i=2}^{n+1} x_i y_i$$

(the so-called **Minkowski product**). It is standard to denote this inner product space by $\mathbb{R}^{1,n}$.

◇ **Exercise 57.** Show that the group of all linear isometries (i.e., linear transformations on $\mathbb{R}^{1,n}$ that preserve the Minkowski product) is *isomorphic* to the matrix group

$$\mathsf{O}(1, n) := \left\{ A \in \mathsf{GL}(n+1, \mathbb{R}) \; : \; A^\top S A = S \right\}$$

where

$$S = \operatorname{diag}(-1, 1, \dots, 1) = \begin{bmatrix} -1 & 0 \\ 0 & \mathbf{1} \end{bmatrix} \in \mathsf{GL}(n+1, \mathbb{R}).$$

In a similar fashion, one can define more general matrix groups

$$\mathsf{O}(k, \ell) \leq \mathsf{GL}(k+\ell, \mathbb{R}) \quad \text{and} \quad \mathsf{SO}(k, \ell) \leq \mathsf{SL}(k+\ell, \mathbb{R})$$

usually called "pseudo-orthogonal" groups (this is Problem 21).

Note : Since $\mathsf{O}(k, \ell)$ and $\mathsf{O}(\ell, k)$ are essentially the same group, we may assume (without any loss of generality) that $1 \leq k \leq \ell$. *The pseudo-orthogonal groups are neither compact nor* *connected*. The groups $\mathsf{O}\,(k,\ell)$ have four (connected) components, whereas the groups $\mathsf{SO}\,(k,\ell)$ have two components.

For each positive number $\rho > 0$, the *hyperboloid*

$$\mathcal{H}_{1,n}(\rho) := \left\{x \in \mathbb{R}^{1,n} \; : \; x \odot x = -\rho\right\}$$

has two (connected) components

$$\mathcal{H}^{+}_{1,n}(\rho) = \{x \in \mathcal{H}_{1,n}(\rho) \; : \; x_1 > 0\} \quad \text{and} \quad \mathcal{H}^{-}_{1,n}(\rho) = \{x \in \mathcal{H}_{1,n}(\rho) \; : \; x_1 < 0\}\,.$$

We define the **Lorentz group** $\mathsf{Lor}\,(1,n)$ to be the (closed) subgroup of $\mathsf{SO}\,(1,n)$ preserving each of the connected sets $\mathcal{H}^{\pm}_{1,n}(1)$. Thus

$$\mathsf{Lor}\,(1,n) := \left\{A \in \mathsf{SO}\,(1,n) \; : \; A\mathcal{H}^{\pm}_{1,n}(1) = \mathcal{H}^{\pm}_{1,n}(1)\right\} \leq \mathsf{SO}\,(1,n).$$

It turns out that $A \in \mathsf{Lor}\,(1,n)$ <u>if and only if</u> it preserves the hyperboloids $\mathcal{H}^{\pm}_{1,n}(\rho)$, $\rho > 0$ and the "light cones" $\mathcal{H}^{\pm}_{1,n}(0)$.

NOTE : The Lorentz group $\mathsf{Lor}\,(1,n)$ is *connected*.

Of particular interest in physics is the Lorentz group $\mathsf{Lor} = \mathsf{Lor}\,(1,3)$. That is,

$$\mathsf{Lor} = \left\{L \in \mathsf{SO}\,(1,3) \; : \; L\mathcal{H}^{\pm}_{1,3}(\rho) = \mathcal{H}^{\pm}_{1,3}(\rho), \;\; \rho \geq 0\right\} \leq \mathsf{SO}\,(1,3).$$

NOTE : We can write

$$\begin{aligned} \mathsf{SO}\,(1,1) &= \mathsf{Lor}\,(1,1) \cup \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \mathsf{Lor}\,(1,1) \\ \mathsf{O}\,(1,1) &= \mathsf{SO}\,(1,1) \cup \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \mathsf{SO}\,(1,1). \end{aligned}$$

(See also PROBLEM 22.)

4.3.5. *The real symplectic group* $\mathsf{Sp}\,(2n,\mathbb{R})$. Let

$$\mathbb{J} := \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{bmatrix} \in \mathsf{SL}\,(2n,\mathbb{R}).$$

A matrix $A \in \mathbb{R}^{2n\times 2n}$ is called *symplectic* if

$$A^\top \mathbb{J} A = \mathbb{J}.$$

NOTE : The word *symplectic* was invented by HERMANN WEYL (1885-1955), who substituted Greek for Latin roots in the word *complex* to obtain a term which would describe a group (related to "line complexes" but which would not be confused with complex numbers).

Let $\mathsf{Sp}\,(2n,\mathbb{R})$ be the set of all $2n \times 2n$ symplectic matrices. Taking determinants of the condition $A^\top \mathbb{J} A = \mathbb{J}$ gives

$$1 = \det \mathbb{J} = (\det A^\top) \cdot (\det \mathbb{J}) \cdot (\det A) = (\det A)^2.$$

Hence $\det A = \pm 1$, and so $A \in \mathsf{GL}\,(2n, \mathbb{R})$. Furthermore, if $A, B \in \mathsf{Sp}\,(2n, \mathbb{R})$, then

$$(AB)^\top \mathbb{J}(AB) = B^\top A^\top \mathbb{J} AB = \mathbb{J}.$$

Hence $AB \in \mathsf{Sp}\,(2n, \mathbb{R})$. Now, if $A^\top \mathbb{J} A = \mathbb{J}$, then

$$\mathbb{J}A = (A^\top)^{-1}\mathbb{J} = (A^{-1})^\top \mathbb{J}$$

so

$$\mathbb{J} = (A^{-1})^\top \mathbb{J} A^{-1}.$$

It follows that $A^{-1} \in \mathsf{Sp}\,(2n, \mathbb{R})$ and hence $\mathsf{Sp}\,(2n, \mathbb{R})$ is a group. In fact, it is a *closed* subgroup of $\mathsf{GL}\,(2n, \mathbb{R})$, and thus a linear Lie group.

NOTE : The symplectic group $\mathsf{Sp}\,(2n, \mathbb{R})$ is *connected.* (It turns out that the determinant of a symplectic matrix must be positive; this fact is by no means obvious.)

◇ **Exercise 58.** Check that $\mathsf{Sp}\,(2, \mathbb{R}) = \mathsf{SL}\,(2, \mathbb{R})$. (In general, it is not true that $\mathsf{Sp}\,(2n, \mathbb{R}) = \mathsf{SL}\,(2n, \mathbb{R})$.)

All matrices of the form

$$\begin{bmatrix} \mathbf{1} & 0 \\ t\mathbf{1} & \mathbf{1} \end{bmatrix} \in \mathsf{SL}\,(2n, \mathbb{R})$$

are symplectic. However, the (Frobenius) norm of such a matrix is equal to $\sqrt{2n + t^2 n}$, which is *unbounded* if $t \in \mathbb{R}$. Therefore, $\mathsf{Sp}\,(2n, \mathbb{R})$ is not a bounded subset of $\mathbb{R}^{2n \times 2n}$ and hence is *not* compact.

◇ **Exercise 59.** Consider the skew-symmetric bilinear form on (the vector space) $\mathbb{R}^{2n}$ defined by

$$\Omega(x, y) := \sum_{i=1}^{n} (x_i y_{n+i} - x_{n+i} y_i)$$

(the standard *symplectic form* or the "canonical" symplectic structure). Show that a linear transformation (on $\mathbb{R}^{2n}$) $x \mapsto Ax$ preserves the symplectic form Ω if and only if $A^\top \mathbb{J} A = \mathbb{J}$ (i.e., the matrix A is symplectic). Such a structure-preserving transformation is called a *symplectic transformation.*

The group of all symplectic transformations on $\mathbb{R}^{2n}$ (equipped with the symplectic form Ω) is *isomorphic* to the linear Lie group $\mathsf{Sp}\,(2n, \mathbb{R})$.

NOTE : The symplectic group is related to *classical mechanics.* Consider a particle of mass m moving in a *potential field* V. Newton's second law states that the particle moves along a curve $t \mapsto x(t)$ in Euclidean space $\mathbb{R}^3$ in such a way that $m\ddot{x} = -\operatorname{grad} V(x)$. Introduce the conjugate *momenta* $p_i = m\dot{x}_i$, $i = 1, 2, 3$ and the *energy* (Hamiltonian)

$$H(x, p) := \frac{1}{2m} \sum_{i=1}^{3} p_i^2 + V(x).$$

Then

$$\frac{\partial H}{\partial x_i} = \frac{\partial V}{\partial x_i} = -m\ddot{x}_i = -\dot{p}_i \qquad \text{and} \qquad \frac{\partial H}{\partial p_i} = \frac{1}{m} p_i = \dot{x}_i$$

and hence *Newton's law* $\mathbf{F} = m\,a$ is equivalent to *Hamilton's equations*

$$\dot{x}_i = \frac{\partial H}{\partial p_i} \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial x_i} \qquad (i = 1,2,3).$$

Writing $z = (x,p)$,

$$\mathbb{J} \cdot \operatorname{grad} H(z) = \begin{bmatrix} 0 & I_3 \\ -I_3 & 0 \end{bmatrix} \begin{bmatrix} \frac{\partial H}{\partial x} \\ \\ \frac{\partial H}{\partial p} \end{bmatrix} = (\dot{x}, \dot{p}) = \dot{z}$$

so Hamilton equations read $\dot{z} = \mathbb{J} \cdot \operatorname{grad} H(z)$. Now let

$$F : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3 \times \mathbb{R}^3$$

and write $w(t) = F(z(t))$. If $z(t)$ satisfies Hamilton's equations

$$\dot{z} = \mathbb{J} \cdot \operatorname{grad} H(z)$$

then $w(t) = F(z(t))$ satisfies $\dot{w} = A^\top \dot{z}$, where $A^\top = [\partial w^i / \partial z^j]$ is the Jacobian matrix of F. By the chain rule,

$$\dot{w} = A^\top \mathbb{J} \operatorname{grad}_z H(z) = A^\top \mathbb{J} A \operatorname{grad}_w H(z(w)).$$

Thus, the equations for $w(t)$ have the form of Hamilton's equations with energy $K(w) = H(z(w))$ if and only if $A^\top \mathbb{J} A = \mathbb{J}$; that is, if and only if A is symplectic. A nonlinear transformation F is *canonical* <u>if and only if</u> its Jacobian matrix is symplectic (or, if one prefers, its tangent mapping is a symplectic transformation).

As a special case, consider a (linear transformation) $A \in \mathsf{Sp}\,(2n, \mathbb{R})$ and let $w = Az$. Suppose H is quadratic (i.e., of the form $H(z) = \frac{1}{2} z^\top B z$ where B is a symmetric matrix). Then $\operatorname{grad} H(z) = Bz$ and thus the equations of motion become the linear equations $\dot{z} = \mathbb{J} B z$. Now

$$\dot{w} = A\dot{z} = A\mathbb{J}Bz = \mathbb{J}(A^\top)^{-1}Bz = \mathbb{J}(A^\top)^{-1}BA^{-1}Az = \mathbb{J}B'w$$

where $B' = (A^\top)^{-1}BA^{-1}$ is symmetric. For the new Hamiltonian we get

$$\begin{aligned} H'(w) &= \frac{1}{2} w^\top (A^\top)^{-1} B A^{-1} w = \frac{1}{2} (A^{-1}w)^\top B A^{-1} w \\ &= H(A^{-1}w) = H(z). \end{aligned}$$

Thus $\mathsf{Sp}\,(2n, \mathbb{R})$ *is the linear invariance group of classical mechanics.*

4.3.6. *The complex general linear group* $\mathsf{GL}\,(n, \mathbb{C})$. Many important matrix groups involve *complex* matrices. As in the real case,

$$\mathsf{GL}\,(n, \mathbb{C}) := \{A \in \mathbb{C}^{n \times n} \,:\, \det A \neq 0\}$$

is an *open* subset of $\mathbb{C}^{n \times n}$, and hence is *not* compact. Clearly $\mathsf{GL}\,(n, \mathbb{C})$ is a group under matrix multiplication.

Note : The general linear group $\mathsf{GL}\,(n, \mathbb{C})$ is *connected*. This is in contrast with the fact that $\mathsf{GL}\,(n, \mathbb{R})$ has two components.

4.3.7. *The complex special linear group* $\mathsf{SL}\,(n,\mathbb{C})$. This group is defined by

$$\mathsf{SL}\,(n,\mathbb{C}) := \{A \in \mathsf{GL}\,(n,\mathbb{C}) \,:\, \det A = 1\}$$

and is treated as in the real case. The matrix group $\mathsf{SL}\,(n,\mathbb{C})$ is *not* compact but *connected.*

4.3.8. *The unitary and special unitary groups* $\mathsf{U}\,(n)$ *and* $\mathsf{SU}\,(n)$. For $A = [a_{ij}] \in \mathbb{C}^{n\times n}$,

$$A^* := \bar{A}^\top = \overline{A^\top}$$

is the *Hermitian conjugate* (i.e., the conjugate transpose) matrix of A; thus, $(A^*)_{ij} = \bar{a}_{ji}$. The **unitary group** is defined as

$$\mathsf{U}\,(n) := \{A \in \mathsf{GL}\,(n,\mathbb{C}) \,:\, A^*A = \mathbf{1}\}.$$

◇ **Exercise 60.** Verify that $\mathsf{U}\,(n)$ is a *subgroup* of the general linear group $\mathsf{GL}\,(n,\mathbb{C})$.

The unitary condition amounts to n^2 equations for the n^2 complex numbers a_{ij}, $i,j = \overline{1,n}$

$$\sum_{k=1}^{n} \bar{a}_{ki} a_{kj} = \delta_{ij}.$$

By taking real and imaginary parts, these equations actually give $2n^2$ equations in the $2n^2$ real and imaginary parts of the a_{ij} (although there is some redundancy). This means that $\mathsf{U}\,(n)$ is a *closed* subset of $\mathbb{C}^{n\times n} = \mathbb{R}^{2n^2}$ and hence of $\mathsf{GL}\,(n,\mathbb{C})$. Thus $\mathsf{U}\,(n)$ is a complex linear Lie group.

Note : The unitary group $\mathsf{U}\,(n)$ is *compact* and *connected.*

Let $A \in \mathsf{U}\,(n)$. From $|\det A| = 1$, we see that the determinant function $\det : \mathsf{GL}\,(n,\mathbb{C}) \to \mathbb{C}$ maps $\mathsf{U}\,(n)$ onto the unit circle $\mathbb{S}^1 = \{z \in \mathbb{C} \,:\, |z| = 1\}$.

Note : In the special case $n = 1$, a complex linear mapping $\phi : \mathbb{C} \to \mathbb{C}$ is multiplication by some complex number z, and ϕ is an *isometry* if and only if $|z| = 1$. In this way, the unitary group $\mathsf{U}\,(1)$ is *identified* with the unit circle $\mathbb{S}^1$. The group $\mathsf{U}\,(1)$ is more commonly known as the *circle group* or the one-dimensional *torus*, and is also denoted by $\mathbb{T}^1$.

The dot product on $\mathbb{R}^n$ can be extended to $\mathbb{C}^n$ by setting (for $x, y \in \mathbb{C}^{n\times 1}$)

$$x \bullet y := x^* y = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n.$$

Note : This is *not* $\mathbb{C}$-linear but satisfies (for $x, y \in \mathbb{C}^{n\times 1}$ and $u, v \in \mathbb{C}$)

$$(ux) \bullet (vy) = \bar{u}v\,(x \bullet y).$$

This dot product allows us to define the *norm* of a complex vector $x \in \mathbb{C}^{n\times 1}$ by

$$\|x\| := \sqrt{x \bullet x}.$$

Then a matrix $A \in \mathbb{C}^{n\times n}$ is *unitary* <u>if and only if</u>

$$Ax \bullet Ay = x \bullet y \qquad (x, y \in \mathbb{C}^n).$$

◇ **Exercise 61.** If $\mathsf{G}_i \leq \mathsf{GL}\,(n_i, \Bbbk)$, $i = 1, 2$ are linear Lie groups, show that their (direct) *product* $\mathsf{G}_1 \times \mathsf{G}_2$ is also a linear Lie group (in $\mathsf{GL}\,(n_1 + n_2, \Bbbk)$). Observe, in particular, that the k-dimensional *torus*

$$\mathbb{T}^k : = \mathbb{T}^1 \times \mathbb{T}^1 \times \cdots \times \mathbb{T}^1$$

is a linear Lie group (in $\mathsf{GL}\,(k, \mathbb{C})$). These groups are compact connected Abelian linear Lie groups. (In fact, they are the *only* linear Lie groups with these properties.)

The **special unitary group**

$$\mathsf{SU}\,(n) := \{A \in \mathsf{U}\,(n) \,:\, \det A = 1\}$$

is a closed subgroup of $\mathsf{U}\,(n)$ and hence a complex matrix group.

Note : The matrix group $\mathsf{SU}\,(n)$ is *compact* and *connected.* In the special case $n = 2$, $\mathsf{SU}\,(2)$ is *diffeomorphic* to the unit sphere $\mathbb{S}^3$ in $\mathbb{C}^2$ (or $\mathbb{R}^4$). The group $\mathsf{SU}\,(2)$ is used in the construction of the gauge group for the Yang-Mills equations in *elementary particle physics.* Also, there is a two-to-one surjection (in fact, a surjective *submersion*)

$$\pi : \mathsf{SU}\,(2) \to \mathsf{SO}\,(3)$$

which is of crucial importance in *computational mechanics* (it is related to the quaternionic representation of rotations in Euclidean space $\mathbb{R}^3$).

4.3.9. *The complex orthogonal groups* $\mathsf{O}\,(n, \mathbb{C})$ *and* $\mathsf{SO}\,(n, \mathbb{C})$. Consider the bilinear form on the vector space $\mathbb{C}$ defined by

$$(x, y) := x_1y_1 + x_2y_2 + \cdots + x_ny_n \qquad (x, y \in \mathbb{C}^n).$$

This form is *not* an inner product because of the lack of complex conjugation in the definition. The set of all complex $n \times n$ matrices which preserve this form (i.e., such that $(Ax, Ay) = (x, y)$ for all $x, y \in \mathbb{C}^n$) is the **complex orthogonal group** $\mathsf{O}\,(n, \mathbb{C})$. Thus

$$\mathsf{O}\,(n, \mathbb{C}) := \left\{A \in \mathsf{GL}\,(n, \mathbb{C}) \,:\, A^\top A = \mathbf{1}\right\} \subset \mathsf{GL}\,(n, \mathbb{C}).$$

It is easy to show that $\mathsf{O}\,(n, \mathbb{C})$ is a liner Lie group, and that $\det A = \pm 1$ for all $\mathsf{O}\,(n, \mathbb{C})$.

Note : The linear Lie group $\mathsf{O}\,(n, \mathbb{C})$ is <u>not</u> the same as the unitary group $\mathsf{U}\,(n)$.

The **complex special orthogonal group**

$$\mathsf{SO}\,(n, \mathbb{C}) := \{A \in \mathsf{O}\,(n, \mathbb{C}) \,:\, \det A = 1\}$$

is also a linear Lie group.

4.3.10. *The unipotent group* $\mathsf{UT}\,(n,\Bbbk)$. A matrix $A = [a_{ij}] \in \Bbbk^{n\times n}$ is *upper triangular* if all the entries below the main diagonal are equal to 0. Let $\mathsf{T}\,(n,\Bbbk)$ denote the set of all $n \times n$ *invertible* upper triangular matrices (over $\Bbbk$). Thus

$$\mathsf{T}\,(n,\Bbbk) := \{A \in \mathsf{GL}\,(n,\Bbbk) \,:\, a_{ij} = 0 \;\text{ for } i > j\}.$$

◇ **Exercise 62.** Show that $\mathsf{T}\,(n,\Bbbk)$ is a *closed* subgroup of the general linear group $\mathsf{GL}\,(n,\Bbbk)$ and hence a linear Lie group.

The group $\mathsf{T}\,(n,\Bbbk)$ is called the (upper) **triangular group**. This group is *not* compact.

Note : Likewise, one can define the *lower* triangular group

$$\widetilde{\mathsf{T}}\,(n,\Bbbk) := \{A \in \mathsf{GL}\,(n,\Bbbk) \,:\, a_{ij} = 0 \;\text{ for } i < j\}.$$

Clearly, $A \in \widetilde{\mathsf{T}}\,(n,\Bbbk)$ $\underline{\text{if and only if}}$ $A^\top \in \mathsf{T}\,(n,\Bbbk)$. The matrix groups $\mathsf{T}\,(n,\Bbbk)$ and $\widetilde{\mathsf{T}}\,(n,\Bbbk)$ are *isomorphic* and there is no need to distinguish between them.

◇ **Exercise 63.** Show that the **diagonal group**

$$\mathsf{D}\,(n,\Bbbk) := \{A \in \mathsf{GL}\,(n,\Bbbk) \,:\, a_{ij} = 0 \;\text{ for } i \neq j\}$$

is a closed subgroup of $\mathsf{T}\,(n,\Bbbk)$ and hence a *linear Lie group.*

◇ **Exercise 64.** For $k \leq n$, let $\mathsf{P}\,(k)$ denote the group of all linear transformations (i.e., invertible linear mappings) on $\mathbb{R}^n$ that preserve the subspace $\mathbb{R}^k = \mathbb{R}^k \times \{0\} \subseteq \mathbb{R}^n$. Show that $\mathsf{P}\,(k)$ is (*isomorphic* to) the matrix group

$$\left\{ \begin{bmatrix} A & X \\ 0 & B \end{bmatrix} \,:\, A \in \mathsf{GL}\,(k,\mathbb{R}),\, B \in \mathsf{GL}\,(n-k,\mathbb{R}),\, X \in \mathbb{R}^{k\times(n-k)} \right\}.$$

An upper triangular matrix $A = [a_{ij}]$ is *unipotent* if it has all diagonal entries equal to 1. The (real or complex) **unipotent group** is (the subgroup of $\mathsf{GL}\,(n,\Bbbk)$)

$$\mathsf{UT}\,(n,\Bbbk) := \{A \in \mathsf{GL}\,(n,\Bbbk) \,:\, a_{ij} = 0 \;\text{ for } i > j \;\text{ and } a_{ii} = 1\}.$$

It is easy to see that the unipotent group $\mathsf{UT}\,(n,\Bbbk)$ is a closed subgroup of $\mathsf{GL}\,(n,\Bbbk)$ and hence a *liner Lie group.*

Note : $\mathsf{UT}\,(n,\Bbbk)$ is a closed subgroup of $\mathsf{T}\,(n,\Bbbk)$.

For the case

$$\mathsf{UT}\,(2,\Bbbk) = \left\{ \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \in \mathsf{GL}\,(n,\Bbbk) \,:\, t \in \Bbbk \right\}$$

the mapping

$$\theta : \Bbbk \to \mathsf{UT}\,(2,\Bbbk), \quad t \mapsto \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}$$

is a *continuous* group homomorphism which is an isomorphism with continuous inverse. This allows us to *view* $\Bbbk$ *as a linear Lie group.*

Note : Given two linear Lie groups G and $\overline{\mathsf{G}}$, a group homomorphism $\theta : \mathsf{G} \to \overline{\mathsf{G}}$ is a *continuous homomorphism* if it is continuous and its image $\theta(\mathsf{G}) \leq \overline{\mathsf{G}}$ is a closed subset of $\overline{\mathsf{G}}$. For instance,

$$\theta : \mathsf{UT}\,(2, \mathbb{R}) \to \mathsf{U}\,(1), \qquad \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix} \mapsto e^{2\pi t i}$$

is a continuous homomorphism of matrix groups, but (for $a \in \mathbb{R} \setminus \mathbb{Q}$)

$$\theta' : \mathsf{G} = \left\{ \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \in \mathsf{UT}\,(2, \mathbb{R}) \,:\, k \in \mathbb{Z} \right\} \to \mathsf{U}\,(1), \qquad \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \mapsto e^{2\pi k a i}$$

is *not* (since its image is a *dense* proper subset of $\mathsf{U}\,(1)$). Whenever we have a continuous homomorphism of linear Lie groups $\theta : \mathsf{G} \to \overline{\mathsf{G}}$ which is a *homeomorphism* (i.e., a continuous bijection with continuous inverse) we say that θ is a *continuous isomorphism* and regard G and $\overline{\mathsf{G}}$ as "identical" (as linear Lie groups).

The unipotent group $\mathsf{UT}\,(3, \mathbb{R})$ is the **Heisenberg group**

$$\mathsf{H}_3 := \left\{ \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix} \,:\, a, b, c \in \mathbb{R} \right\}$$

which is particularly important in *quantum physics*; the *Lie algebra* of H_3 gives a realization of the *Heisenberg commutation relations* of quantum mechanics.

◇ **Exercise 65.** Verify that the 4×4 unipotent matrices A of the form

$$A = \begin{bmatrix} 1 & a_2 & a_3 & a_4 \\ 0 & 1 & a_1 & \frac{a_1^2}{2} \\ 0 & 0 & 1 & a_1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

form a closed subgroup of $\mathsf{UT}\,(4, \mathbb{R})$ and hence a linear Lie group. Generalize (to $n \times n$ matrices).

Several other matrix groups are of great interest. We describe briefly some of them.

4.3.11. *The general affine group* $\mathsf{AGL}\,(n, \Bbbk)$. The **general affine group** (over $\Bbbk$) is the group

$$\mathsf{AGL}\,(n, \Bbbk) := \left\{ \begin{bmatrix} 1 & 0 \\ c & A \end{bmatrix} \in \mathsf{GL}\,(n+1, \Bbbk) \,:\, c \in \Bbbk^{n \times 1} \text{ and } A \in \mathsf{GL}\,(n, \Bbbk) \right\}.$$

This is clearly a closed subgroup of the general linear group $\mathsf{GL}\,(n+1, \Bbbk)$ and hence a *linear Lie group.* The general affine group $\mathsf{AGL}\,(n, \Bbbk)$ is *not* compact. Likewise the case of the general linear group, the linear Lie group $\mathsf{AGL}\,(n, \mathbb{C})$ is connected but $\mathsf{AGL}\,(n, \mathbb{R})$ is not.

Note : If we identify the element $x \in \Bbbk^n$ with $\begin{bmatrix} 1 \\ x \end{bmatrix} \in \Bbbk^{(n+1)\times 1}$, then since

$$\begin{bmatrix} 1 & 0 \\ c & A \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = \begin{bmatrix} 1 \\ Ax + c \end{bmatrix}$$

we obtain an *action* of the group $\mathsf{AGL}\,(n, \Bbbk)$ on the vector space $\Bbbk^n$. Transformations on $\Bbbk^n$ having the form $x \mapsto Ax + c$ (with A invertible) are called *affine transformations* and they preserve *lines* (i.e., one-dimensional linear submanifolds of $\Bbbk^n$). The associated geometry is **affine geometry** that has $\mathsf{AGL}\,(n, \Bbbk)$ as its symmetry group.

The (additive group of the) vector space $\Bbbk^n$ (in fact, $\Bbbk^{n\times 1}$) can be viewed as (and identified with) the *translation subgroup* of $\mathsf{AGL}\,(n, \Bbbk)$

$$\left\{ \begin{bmatrix} 1 & 0 \\ c & \mathbf{1} \end{bmatrix} \in \mathsf{GL}\,(n+1, \Bbbk) \;:\; c \in \Bbbk^{n\times 1} \right\} \leq \mathsf{AGL}\,(n, \Bbbk)$$

and this is a closed subgroup.

The identity component of the general affine group $\mathsf{AGL}\,(n, \mathbb{R})$ is (the linear Lie group)

$$\mathsf{AGL}^+\,(n, \mathbb{R}) = \left\{ \begin{bmatrix} 1 & 0 \\ c & A \end{bmatrix} \;:\; c \in \mathbb{R}^{n\times 1} \text{ and } A \in \mathsf{GL}^+\,(n, \mathbb{R}) \right\}.$$

In particular,

$$\mathsf{AGL}^+\,(1, \mathbb{R}) = \left\{ \begin{bmatrix} 1 & 0 \\ c & e^a \end{bmatrix} \;:\; a, c \in \mathbb{R} \right\}$$

is a *connected* linear Lie group (of "dimension" 2). Its elements are (in fact, can be identified with) transformations of the real line $\mathbb{R}$ having the form $x \mapsto bx + c$ (with $b, c \in \mathbb{R}$ and $b > 0$).

4.3.12. *The Euclidean group* $\mathsf{E}\,(n)$. This is the matrix group

$$\mathsf{E}\,(n) := \left\{ \begin{bmatrix} 1 & 0 \\ c & A \end{bmatrix} \in \mathsf{GL}\,(n+1, \mathbb{R}) \;:\; c \in \mathbb{R}^{n\times 1} \text{ and } A \in \mathsf{O}\,(n) \right\}.$$

The Euclidean group $\mathsf{E}\,(n)$ is a closed subgroup of the general affine group $\mathsf{AGL}\,(n, \mathbb{R})$ and also is neither compact nor connected. It can be viewed as (and thus identified with) the group of all *isometries* (i.e., *rigid motions*) of the Euclidean space $\mathbb{R}^n$.

4.3.13. *The special Euclidean group* $\mathsf{SE}\,(n)$. The **special Euclidean group** $\mathsf{SE}\,(n)$ is (the linear Lie group) defined by

$$\mathsf{SE}\,(n) := \left\{ \begin{bmatrix} 1 & 0 \\ c & R \end{bmatrix} \in \mathsf{GL}\,(n+1, \mathbb{R}) \;:\; c \in \mathbb{R}^{n\times 1} \text{ and } R \in \mathsf{SO}\,(n) \right\}.$$

This group is *isomorphic* to the group of all *orientation-preserving isometries* (i.e., *proper rigid motions*) on the Euclidean space $\mathbb{R}^n$. It is *not* compact but *connected.*

4.3.14. *Further examples.* Several important groups which are not naturally groups of matrices can be viewed as linear Lie groups. We have seen that the multiplicative groups $\mathbb{R}^\times$ and $\mathbb{C}^\times$ (of non-zero real numbers and complex numbers, respectively) are *isomorphic* to the linear Lie groups $\mathsf{GL}\,(1,\mathbb{R})$ and $\mathsf{GL}\,(1,\mathbb{C})$, respectively. Also, the *circle group* $\mathbb{S}^1$ (of complex numbers with absolute value one) is *isomorphic* to $\mathsf{U}\,(1)$. The *n-torus* (the direct product of n copies of $\mathbb{S}^1$)

$$\mathbb{T}^n = \mathbb{S}^1 \times \cdots \times \mathbb{S}^1 \leq \mathsf{GL}\,(n,\mathbb{C})$$

is *isomorphic* to the linear Lie group of $n \times n$ diagonal matrices with complex entries of modulus one. ($\mathbb{T}^n$ can also be realized as the *quotient group* $\mathbb{R}^n/\mathbb{Z}^n$: an element $(\theta_1,\dots,\theta_n) \mod \mathbb{Z}^n$ of $\mathbb{R}^n/\mathbb{Z}^n$ can be identified with the diagonal matrix
$\operatorname{diag}\left(e^{2\pi i\theta_1},\dots,e^{2\pi i\theta_n}\right)$.)

Note : If $\theta : \mathsf{G} \to \overline{\mathsf{G}}$ is a continuous homomorphism of linear Lie groups, then its *kernel* $\operatorname{Ker}\theta \leq \mathsf{G}$ is a linear Lie group. Moreover, the *quotient group* $\mathsf{G}/\operatorname{Ker}\theta$ can be identified with the linear Lie group $\operatorname{Im}\theta$ by the usual quotient isomorphism $\widetilde{\theta} : \mathsf{G}/\operatorname{Ker}\theta \to \operatorname{Im}\theta$. However, it is important to realize that *not every normal matrix subgroup* H *of the linear Lie group* G *gives rise to a linear Lie group* G/H; there are examples for which G/H is a *Lie group* but not a linear Lie group. (It is true, but by no means obvious, that every linear Lie group is in fact a Lie group.)

Recall that the (additive) groups $\mathbb{R}$ and $\mathbb{C}$ are *isomorphic* to the unipotent groups $\mathsf{UT}\,(2,\mathbb{R})$ and $\mathsf{UT}\,(2,\mathbb{C})$, respectively.

◇ **Exercise 66.** Verify that the map

$$x \in \mathbb{R} \mapsto [e^x] \in \mathsf{GL}^+\,(1,\mathbb{R})$$

is a *continuous isomorphism* of linear Lie groups, and then show that the additive group $\mathbb{R}^n$ is *isomorphic* to the linear Lie group of all $n \times n$ diagonal matrices with positive entries.

The *symmetric group* $\mathfrak{S}_n$ of permutations on n elements may be considered as well as a linear Lie group. Indeed, we can make $\mathfrak{S}_n$ to *act* on $\Bbbk^n$ by linear transformations :

$$\sigma \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{\sigma^{-1}(1)} \\ x_{\sigma^{-1}(2)} \\ \vdots \\ x_{\sigma^{-1}(n)} \end{bmatrix}.$$

Thus (for the standard unit vectors $e_1, e_2, \dots, e_n$) $\sigma \cdot e_i = e_{\sigma(i)}, \quad i = \overline{1,n}$.

The matrix $[\sigma]$ of the linear transformation induced by $\sigma \in \mathfrak{S}_n$ (with respect to the standard basis) has all its entries 0 or 1, with exactly one 1 in each row and column. Such a matrix is usually called a *permutation matrix.*

$\diamond$ **Exercise 67.** Write down the permutations matrices induces by the elements (permutations) of $\mathfrak{S}_3$.

When $\Bbbk = \mathbb{R}$ each of these permutation matrices is orthogonal, while when $\Bbbk = \mathbb{C}$ it is unitary. So, for a given $n \geq 1$, *the symmetric group* $\mathfrak{S}_n$ *is (isomorphic to) a closed subgroup of* $\mathsf{O}(n)$ *or* $\mathsf{U}(n)$.

NOTE : Any *finite* group is (isomorphic to) a linear Lie subgroup of some orthogonal group $\mathsf{O}(n)$.

4.4. **Complex matrix groups as real matrix groups.** Recall that the (complex) vector space $\mathbb{C}$ can be viewed as a *real* two-dimensional vector space (with basis $\{1, i\}$, for example).

$\diamond$ **Exercise 68.** Show that the mapping

$$\rho : \mathbb{C} \to \mathbb{R}^{2\times 2}, \quad z = x + iy \mapsto \begin{bmatrix} x & -y \\ y & x \end{bmatrix}$$

is an injective *ring homomorphism* (i.e., a one-to-one mapping such that, for $z, z' \in \mathbb{C}$,

$$\rho(z + z') = \rho(z) + \rho(z') \quad \text{and} \quad \rho(zz') = \rho(z)\rho(z').)$$

We can view $\mathbb{C}$ as a *subring* of $\mathbb{R}^{2\times 2}$. In other words, we can *identify* the complex number $z = x + iy$ with the 2×2 real matrix $\rho(z)$.

NOTE : This can also be expressed as

$$\rho(x + iy) = xI_2 - yJ_2, \quad \text{where} \quad J_2 := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Also, for $z \in \mathbb{C}$,

$$\rho(\bar{z}) = \rho(z)^\top$$

(complex conjugation corresponds to transposition).

More generally, given $Z = [z_{rs}] \in \mathbb{C}^{n\times n}$ with $z_{rs} = x_{rs} + iy_{rs}$, we can write

$$Z = X + iY,$$

where $X = [x_{rs}],\ Y = [y_{rs}] \in \mathbb{R}^{n\times n}$.

$\diamond$ **Exercise 69.** Show that the mapping

$$\rho_n : \mathbb{C}^{n\times n} \to \mathbb{R}^{2n\times 2n}, \qquad Z = X + iY \mapsto \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}$$

is an injective *ring homomorphism.*

Hence we can *identify* the complex matrix $Z = X + iY$ with the $2n \times 2n$ real matrix $\rho_n(Z)$. Let

$$\mathbb{J} = \mathbb{J}_{2n} := \begin{bmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{bmatrix} \in \mathsf{SL}\,(2n, \mathbb{R}).$$

Then we can write

$$\rho_n(Z) = \rho_n(X + iY) = \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix} - \begin{bmatrix} Y & 0 \\ 0 & Y \end{bmatrix} \mathbb{J}.$$

$\diamond$ **Exercise 70.** First verify that

$$\mathbb{J}^2 = -I_{2n} \quad \text{and} \quad \mathbb{J}^\top = -\mathbb{J}$$

and then show that, for $Z \in \mathbb{C}^{n \times n}$,

$$\rho_n(\bar{Z}) = \rho_n(Z)^\top \iff X = X^\top \quad \text{and} \quad Y = Y^\top.$$

We see that $\rho_n(\mathsf{GL}\,(n, \mathbb{C}))$ is a closed subgroup of $\mathsf{GL}\,(2n, \mathbb{R})$, so *any linear Lie subgroup* G *of* $\mathsf{GL}\,(n, \mathbb{C})$ *can be viewed as a linear Lie subgroup of* $\mathsf{GL}\,(2n, \mathbb{R})$ (by identifying it with its image $\rho_n(\mathsf{G})$ under ρ_n). The following characterizations are sometimes useful:

$$\begin{aligned} \rho_n(\mathbb{C}^{n \times n}) &= \left\{A \in \mathbb{R}^{n \times n} \,:\, A\mathbb{J} = \mathbb{J}A\right\} \\ \rho_n(\mathsf{GL}\,(n, \mathbb{C})) &= \left\{A \in \mathsf{GL}\,(2n, \mathbb{R}) \,:\, A\mathbb{J} = \mathbb{J}A\right\}. \end{aligned}$$

Note : In a slight abuse of notation, the real symplectic group $\mathsf{Sp}\,(2n, \mathbb{R})$ is related to the unitary group $\mathsf{U}\,(n)$ by

$$\mathsf{Sp}\,(2n, \mathbb{R}) \cap \mathsf{O}\,(2n) = \mathsf{U}\,(n).$$

Problems (16–25)

(16) Consider a matrix $A \in \Bbbk^{n \times n}$.

(a) Assume that $\operatorname{rank} A = k$; show that there exist matrices $P, Q \in \mathsf{GL}\,(n, \Bbbk)$ such that

$$A = P \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix} Q.$$

(b) Verify that the sequence $(A_r)_{r \in \mathbb{N}}$ in $\mathsf{GL}\,(n, \Bbbk)$ with

$$A_r = P \begin{bmatrix} I_k & 0 \\ 0 & \frac{1}{r} I_{n-k} \end{bmatrix} Q$$

converges to A. Hence deduce that the set $\mathsf{GL}\,(n, \Bbbk)$ is *dense* in $\Bbbk^{n \times n}$. (A set whose closure is the whole space is said to be **dense** in the space.)

(17) Let $A, B \in \Bbbk^{n\times n}$. By using the result of PROBLEM 16 or otherwise, prove that the matrices AB and BA have the same *characteristic polynomial* and hence the same eigenvalues. (The **characteristic polynomial** of A is defined by $\mathrm{char}_A(\lambda) := \det(\lambda\mathbf{1} - A) \in \Bbbk[\lambda]$.)
[HINT: For an alternative proof, compare the determinants of the two products of the block matrices $\begin{bmatrix} A & -\lambda I_n \\ I_n & 0 \end{bmatrix}$ and $\begin{bmatrix} B & -\lambda I_n \\ -I_n & A \end{bmatrix}$.]

(18) (a) Determine the *center* $\mathsf{Z}(\mathsf{GL}(n,\Bbbk))$ of the general linear group $\mathsf{GL}(n,\Bbbk)$.
(b) Show that
(i) $\mathsf{Z}(\mathsf{GL}(n,\Bbbk))$ and $\mathsf{SL}(\mathsf{n},\Bbbk)$ are *normal* subgroups of $\mathsf{GL}(n,\Bbbk)$.
(ii) $\mathsf{GL}^+(n,\mathbb{R})$ is a *normal* subgroup of $\mathsf{GL}(n,\mathbb{R})$ (see section 4.3.1).
(iii) for each subset $\mathsf{M} \subseteq \Bbbk^{n\times n}$, the **centralizer**

$$\mathsf{Z}_{\mathsf{GL}(n,\Bbbk)}(\mathsf{M}) := \{A \in \mathsf{GL}(n,\Bbbk) \,:\, AX = XA \text{ for all } X \in \mathsf{M}\}$$

is a *closed* subgroup of $\mathsf{GL}(n,\Bbbk)$.

(19) Let $A \in \mathsf{GL}(n,\mathbb{R})$.
(a) Show that the symmetric matrix $S = A^\top A$ is *positive definite* (i.e., its eigenvalues are all positive real numbers). Deduce that S has a positive definite (real) symmetric square root, i.e., there is a positive definite symmetric matrix S_1 such that $S_1^2 = S$.
(b) Show that the matrix $S_1^{-1}A$ is orthogonal.
(c) If $PR = QS$, where P, Q are positive definite symmetric matrices and $R, S \in \mathsf{O}(n)$, show that $P^2 = Q^2$.
(d) Let S_2 be a positive definite symmetric matrix for which $S_2^2 = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$. Show that $S_2 = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$.
(e) Show that A can be *uniquely* expressed as $A = PR$, where P is a positive definite symmetric matrix and $R \in \mathsf{O}(n)$. If $\det A > 0$, show that $R \in \mathsf{SO}(n)$. (Such factorization is called **polar decomposition** of A.)

(20) Let $a \in \mathbb{R} \setminus \mathbb{Q}$. Show that

$$\mathsf{G} = \left\{ \begin{bmatrix} e^{it} & 0 \\ 0 & e^{iat} \end{bmatrix} \,:\, t \in \mathbb{R} \right\}$$

is a subgroup of $\mathsf{GL}(2,\mathbb{C})$, and then find a sequence of matrices in G which converges to $-I_2 \notin \mathsf{G}$. This means that G is not a linear Lie group.
[HINT : By taking $t = (2n+1)\pi$ for a suitably chosen $n \in \mathbb{Z}$, we can make ta *arbitrarily close* to an odd integer multiple of π, $(2m+1)\pi$ say. It is sufficient to show that for any positive integer N, there exist $n, m \in \mathbb{Z}$ such that $|(2n+1)a - (2m+1)| < \frac{1}{N}$·]

(21) Define the inner product $\langle \cdot , \cdot \rangle_{k,\ell}$ on $\mathbb{R}^{k+\ell}$ by the formula

$$\langle x, y\rangle_{k,\ell} := -x_1y_1 - \cdots - x_ky_k + x_{k+1}y_{k+1} + \cdots + x_{k+\ell}y_{k+\ell}.$$

The **pseudo-orthogonal group** $\mathsf{O}\,(k,\ell)$ consists of all matrices $A \in \mathsf{GL}\,(k{+}\ell, \mathbb{R})$ which preserve this inner product (i.e., such that $\langle Ax, Ay\rangle_{k,\ell} = \langle x, y\rangle_{k,\ell}$ for all $x, y \in \mathbb{R}^{k+\ell}$).

(a) Verify that $\mathsf{O}\,(k,\ell)$ is a *linear Lie subgroup* of $\mathsf{GL}\,(k+\ell, \mathbb{R})$.

(b) Let

$$Q = \operatorname{diag}\,(-1, \dots, -1, 1, \dots, 1) = \begin{bmatrix} -I_k & 0 \\ 0 & I_\ell \end{bmatrix}.$$

Prove that a matrix $A \in \mathsf{GL}\,(k{+}\ell, \mathbb{R})$ is in $\mathsf{O}\,(k,\ell)$ $\underline{\text{if and only if}}$ $A^\top QA = Q$. Hence deduce that $\det A = \pm 1$.

(c) Verify that $\mathsf{SO}\,(k,\ell) := \mathsf{O}\,(k,\ell) \cap \mathsf{SL}\,(k+\ell, \mathbb{R})$ is a *linear Lie subgroup* of $\mathsf{SL}\,(k+\ell, \mathbb{R})$.

(22) Show that

(a) The matrix $A = \begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}$ is in $\mathsf{SO}\,(1,1)$.

(b) For every $s, t \in \mathbb{R}$

$$\begin{bmatrix} \cosh s & \sinh s \\ \sinh s & \cosh s \end{bmatrix} \begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix} = \begin{bmatrix} \cosh(s+t) & \sinh(s+t) \\ \sinh(s+t) & \cosh(s+t) \end{bmatrix}.$$

(c) Every element (matrix) of $\mathsf{O}\,(1,1)$ can be written in one of the four forms

$$\begin{bmatrix} \cosh t & \sinh t \\ \sinh t & \cosh t \end{bmatrix}, \begin{bmatrix} -\cosh t & \sinh t \\ \sinh t & -\cosh t \end{bmatrix}, \begin{bmatrix} \cosh t & -\sinh t \\ \sinh t & -\cosh t \end{bmatrix}, \begin{bmatrix} -\cosh t & -\sinh t \\ \sinh t & \cosh t \end{bmatrix}.$$

(Since $\cosh t$ is always positive, there is no overlap among the four cases. Matrices of the first two forms have determinant one; matrices of the last two forms have determinant minus one.)

(23) Given $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathsf{GL}\,(2n, \mathbb{R})$, show that $A \in \mathsf{Sp}\,(2n, \mathbb{R})$ $\underline{\text{if and only if}}$ $a^\top c$ and $b^\top d$ are symmetric and $a^\top d - c^\top b = \mathbf{1}$.

(24) Let $\mathbb{Z}^n \leq \mathbb{R}^n$ be the *discrete* subgroup of vectors with integer entries and set

$$\mathsf{GL}\,(n, \mathbb{Z}) := \{A \in \mathsf{GL}\,(n, \mathbb{R}) \;:\; A\,(\mathbb{Z}^n) = \mathbb{Z}^n\}.$$

Show that $\mathsf{GL}\,(n, \mathbb{Z})$ is a linear Lie group. (This linear group consists of $n \times n$ matrices over (the ring) $\mathbb{Z}$ with determinant ± 1.)

(25) Verify the folowing set of equalities :

$$\begin{aligned}\rho_n(\mathsf{U}\,(n)) &= \mathsf{O}\,(n) \cap \rho_n(\mathsf{GL}\,(n, \mathbb{C}))\\ &= \mathsf{O}\,(n) \cap \mathsf{Sp}\,(2n, \mathbb{R})\\ &= \rho_n(\mathsf{GL}\,(n, \mathbb{C})) \cap \mathsf{Sp}\,(2n, \mathbb{R}).\end{aligned}$$

5. The Matrix Exponential

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

It follows that

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

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

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

Definition 61. The mapping

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

is called the **matrix exponential map**.

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

◇ **Exercise 74.** Show that

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

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

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

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

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

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

□

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

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

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

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

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

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

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

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

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

◇ **Exercise 76.**

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

◇ **Exercise 79.** Compute

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

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

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

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

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

Hence we have the following result

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

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

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

So the mapping

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

is defined and *differentiable* with

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

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

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

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

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

5.2. **Some useful formulas.**

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

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

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

Proof. Consider the difference

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

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

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

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

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

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

5.2.2. *Second formula.*

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

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

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

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

□

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

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

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

5.2.3. *Third formula.*

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

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

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

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

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

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

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

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

Then

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

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

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

□

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

THEOREM 38 (LIOUVILLE'S FORMULA). *For $A \in \Bbbk^{n\times n}$ we have*

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

FIRST PROOF (using the second definition of the exponential) : We have

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

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

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

It only remains to note that (for any $a \in \Bbbk$)

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

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

SECOND PROOF (using differential equations) : Consider the curve

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

Then (by PROPOSITION 37 applied to the curve γ)

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

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

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

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

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

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

and

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

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

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

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

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

The matrix

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

is strictly upper triangular, and so

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

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

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

□

The exponential mapping

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

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

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

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

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

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

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

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

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

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

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

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

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

(the Lie algebra of the Heisenberg group H_3)

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

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

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

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

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

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

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

Hence

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

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

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

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

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

with

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

Similar considerations lead to

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

where

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

This gives

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

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

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

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

and so

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

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

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

We also have

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

We have proved the following result.

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

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

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

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

and write

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

with

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

Then

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

where

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

Also

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

with

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

We see that

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

and by taking norms we get

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

The following estimation holds.

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

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

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

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

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

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

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

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

with

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

As $r \to \infty$,

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

and hence

$$rC_r \to U + V.$$

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

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

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

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

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

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

with (as $r \to \infty$)

$$rC_r \to U + V.$$

We also have

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

where

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

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

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

where

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

and

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

Combining these we get

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

where

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

One can verify that

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

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

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

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

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

and hence

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

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

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

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

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

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

(a) $\mathrm{Ad}_g(\lambda A + \mu B) = \lambda\mathrm{Ad}_g(A) + \mu\mathrm{Ad}_g(B)$.

(b) $\mathrm{Ad}_g([A,B]) = [\mathrm{Ad}_g(A), \mathrm{Ad}_g(B)]$.

(c) $\mathrm{Ad}_{gh}(A) = \mathrm{Ad}_g(\mathrm{Ad}_h(A))$.

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

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

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

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

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

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

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

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

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

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

we have the following formula

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

Explicitly, the formula says that

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

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

Note : The mapping

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

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

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

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

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

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

Problems (26–32)

(26) A matrix $A \in \Bbbk^{n\times n}$ is **nilpotent** if $A^k = 0$ for some $k \geq 1$.

(a) Prove that a nilpotent matrix is singular.

(b) Prove that a *strictly upper triangular* matrix $A = [a_{ij}]$ (i.e. with $a_{ij} = 0$ whenever $i \geq j$) is nilpotent.

(c) Find two nilpotent matrices whose product is *not* nilpotent.

(27) Suppose that $A \in \Bbbk^{n\times n}$ and $\|A\| < 1$.

(a) Show that the matrix series

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

converges (in $\Bbbk^{n\times n}$).

(b) Show that the matrix $\mathbf{1} - A$ is *invertible* and find a formula for $(\mathbf{1} - A)^{-1}$.

(c) If A is *nilpotent*, determine $(\mathbf{1} - A)^{-1}$ and $\exp(A)$.

(28) Let $A \in \Bbbk^{n\times n}$.

(a) Prove that A is *nilpotent* if and only if all its eigenvalues are equal to zero.

(b) The matrix A is called **unipotent** if $\mathbf{1} - A$ is nilpotent (i.e., $(\mathbf{1} - A)^k = 0$ for some $k \geq 1$). Prove that A is *unipotent* if and only if all its eigenvalues are equal to 1.

(c) If A is a strictly upper triangular matrix, show that $\exp(A)$ is unipotent.

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

(30) If $A, B \in \Bbbk^{n\times n}$ commute, show that

$$\frac{d}{dt}\exp(A + tB)\bigg|_{t=0} = \exp(A)B = B\exp(A).$$

(This is a formula for the *derivative* of the exponential mapping exp at an arbitrary A, evaluated only at those B such that $AB = BA$. The general situation is more complicated.)

(31) Let $A, B \in \Bbbk^{n\times n}$.

(a) Verify that

$$\operatorname{ad}[A, B] = \operatorname{ad} A \operatorname{ad} B - \operatorname{ad} B \operatorname{ad} A = [\operatorname{ad} A,\, \operatorname{ad} B]\,.$$

(This means that $\operatorname{ad} : \Bbbk^{n\times n} \to \mathsf{End}(\Bbbk^{n\times n})$ is a *Lie algebra homomorphism.*)

(b) Show by induction that

$$(\operatorname{ad} A)^n(B) = \sum_{k=0}^{n} \binom{n}{k} A^k B(-A)^{n-k}.$$

(c) Show by direct computation that

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

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

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

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

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

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

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

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

6. Lie Algebras

Tangent space to a linear Lie group • Lie algebras • Homomorhisms of Lie algebras • Lie algebras of linear Lie groups: examples.

6.1. **Tangent space to a linear Lie group.**

Let $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ be a linear Lie group.

Definition 62. A **one-parameter subgroup** of G is a continuous mapping $\gamma : \mathbb{R} \to \mathsf{G}$ which satisfies (the homomorphism property)

$$\gamma(s+t) = \gamma(s)\gamma(t) \qquad (t, s \in \mathbb{R}).$$

Note : Recall that $\mathbb{R}$ can be viewed as a linear Lie group. Hence (the one-parameter subgroup) γ is a *continuous homomorphism* of linear Lie groups. It can be shown that *every one-parameter subgroup of* G *is differentiable at* 0 (in fact, differentiable at every $t \in \mathbb{R}$).

A one-parameter subgroup $\gamma : \mathbb{R} \to \mathsf{G}$ can be viewed as a *collection* $(\gamma(t))_{t\in\mathbb{R}}$ of linear transformations on $\Bbbk^n$ such that (for $t, s \in \mathbb{R}$)

- $\gamma(0) = \mathrm{id}_{\Bbbk^n}$.
- $\gamma(s+t) = \gamma(s)\gamma(t)$.
- $\gamma(t) \in \mathsf{G}$ depends continuously on t.

In other words, γ is a linear representation of the (Abelian) group $\mathbb{R}$ on (the vector space) $\Bbbk^n$. (So γ defines a continuous action of $\mathbb{R}$ on $\Bbbk^n$.) On the other hand, the (parametrized) curve $\gamma : \mathbb{R} \to \mathsf{G}$ has a *tangent vector* $\dot\gamma(0) \in \Bbbk^{n\times n}$ at $\gamma(0) = \mathbf{1}$.

Proposition 43. *Let* $\gamma : \mathbb{R} \to \mathsf{G}$ *be a one-parameter subgroup of* G. *Then* γ *is differentiable at every* $t \in \mathbb{R}$ *and*

$$\dot\gamma(t) = \dot\gamma(0)\,\gamma(t) = \gamma(t)\,\dot\gamma(0).$$

Proof. We have (for $t, h \in \mathbb{R}$)

$$\begin{aligned} \dot\gamma(t) &= \lim_{h\to 0} \frac{1}{h}\left(\gamma(t+h) - \gamma(t)\right) \\ &= \lim_{h\to 0} \frac{1}{h}\left(\gamma(h)\gamma(t) - \gamma(t)\right) \\ &= \left(\lim_{h\to 0} \frac{1}{h}\left(\gamma(h) - \mathbf{1}\right)\right)\gamma(t) \\ &= \dot\gamma(0)\,\gamma(t) \end{aligned}$$

and similarly

$$\dot\gamma(t) = \gamma(t)\,\dot\gamma(0).$$

□

We can now determine the form of all one-parameter subgroups of G.

THEOREM 44. *Let $\gamma : \mathbb{R} \to \mathsf{G}$ be a one-parameter subgroup of G. Then it has the form*

$$\gamma(t) = \exp(tA)$$

for some $A \in \Bbbk^{n\times n}$.

Proof. Let $A = \dot\gamma(0)$. This means that γ satisfies (the differential equation)

$$\dot\gamma(t) = A\,\gamma(t)$$

and is subject to (the initial condition)

$$\gamma(0) = \mathbf{1}.$$

This initial value problem (IVP) has the unique solution $\gamma(t) = \exp(tA)$. □

We cannot yet reverse this process and decide for which $A \in \Bbbk^{n\times n}$ the one-parameter subgroup

$$\gamma : \mathbb{R} \to \mathsf{GL}(n, \Bbbk), \quad t \mapsto \exp(tA)$$

actually takes values in G. (The answer involves the *Lie algebra* of G.)

NOTE : We have a curious phenomenon in the fact that although the definition of a one-parameter group only involves first order differentiability, the general form $\exp(tA)$ is always infinitely differentiable (and indeed *analytic*) as a function of t. This is an important characteristic of much of the *Lie theory*, namely that conditions of first order differentiability (and even continuity) often lead to much stronger conditions.

Let $\mathsf{G} \leq \mathsf{GL}(n, \Bbbk)$ be a linear Lie group. Recall that $\Bbbk^{n\times n}$ may be considered to be some Euclidean space $\mathbb{R}^m$.

DEFINITION 63. A (parametrized) **curve** in G is a *differentiable* mapping $\gamma : (a, b) \subseteq \mathbb{R} \to \Bbbk^{n\times n}$ such that

$$\gamma(t) \in \mathsf{G} \quad \text{for all } \ t \in (a, b).$$

The derivative

$$\dot\gamma(t) := \lim_{h\to 0} \frac{1}{h}\left(\gamma(t+h) - \gamma(t)\right) \in \Bbbk^{n\times n}$$

is called the **tangent vector** to γ at $\gamma(t)$. We will usually assume that $a < 0 < b$.

◇ **Exercise 83.** Given two curves $\gamma, \sigma : (a, b) \to \mathsf{G}$, we define a new curve, the **product curve**, by

$$(\gamma\,\sigma)(t) := \gamma(t)\,\sigma(t).$$

Show that (for $t \in (a, b)$)

$$(\gamma\,\sigma)^{\cdot}(t) = \gamma(t)\,\dot\sigma(t) + \dot\gamma(t)\,\sigma(t).$$

◇ **Exercise 84.**

(a) Let $\gamma : (-1,1) \to \mathbb{R}^{3\times 3}$ be given by

$$\gamma(t) := \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{bmatrix}.$$

Show that γ is a curve in $\mathsf{SO}\,(3)$ and find $\dot\gamma(0)$. Show that

$$(\gamma^2)^{\cdot}(0) = 2\dot\gamma(0).$$

(b) Let $\sigma : (-1,1) \to \mathbb{R}^{3\times 3}$ be given by

$$\sigma(t) := \begin{bmatrix} 0 & 0 & 0 \\ 0 & \cos t & \sin t \\ 0 & -\sin t & \cos t \end{bmatrix}.$$

Calculate $\dot\sigma(0)$. Write the matrix $\gamma(t)\,\sigma(t)$ and verify that

$$(\gamma\,\sigma)^{\cdot}(0) = \dot\gamma(0) + \dot\sigma(0).$$

◇ **Exercise 85.** Let $\alpha : (-1,1) \to \mathbb{C}^{n\times n}$ be given by

$$\alpha(t) := \begin{bmatrix} e^{i\pi t} & 0 & 0 \\ 0 & e^{i\frac{\pi t}{2}} & 0 \\ 0 & 0 & e^{i\frac{\pi t}{2}} \end{bmatrix}.$$

Show that α is a curve in $\mathsf{U}\,(3)$. Calculate $\dot\alpha(0)$.

Definition 64. The **tangent space** to G at $A \in \mathsf{G}$ is the set

$$T_A\,\mathsf{G} := \{\dot\gamma(0) \in \mathbb{k}^{n\times n} \,:\, \gamma \text{ is a curve in } \mathsf{G} \text{ with } \gamma(0) = A\}.$$

Proposition 45. *The set $T_A\,\mathsf{G}$ is a real vector subspace of $\mathbb{k}^{n\times n}$.*

Proof. Let $\alpha, \beta : (a,b) \to \mathbb{k}^{n\times n}$ be two curves in G through A (i.e., $\alpha(0) = \beta(0) = A$). Then

$$\gamma : (a,b) \to \mathbb{k}^{n\times n}, \quad t \mapsto \alpha(t)\,A^{-1}\,\beta(t)$$

is also a curve in G with $\gamma(0) = A$. We have

$$\dot\gamma(t) = \dot\alpha(t)\,A^{-1}\,\beta(t) + \alpha(t)\,A^{-1}\,\dot\beta(t)$$

and hence

$$\dot\gamma(0) = \dot\alpha(0)\,A^{-1}\,\beta(0) + \alpha(0)\,A^{-1}\,\dot\beta(0) = \dot\alpha(0) + \dot\beta(0)$$

which shows that $T_A\,\mathsf{G}$ is closed under (vector) addition.

Similarly, if $\lambda \in \mathbb{R}$ and $\alpha : (a,b) \to \mathbb{k}^{n\times n}$ is a curve in G with $\alpha(0) = A$, then

$$\eta : (a,b) \to \mathbb{k}^{n\times n}, \quad t \mapsto \alpha(\lambda t)$$

is another such curve. Since $\dot\eta(0) = \lambda\,\dot\alpha(0)$, we see that $T_A\,\mathsf{G}$ is closed under (real) scalar multiplication. So $T_A\,\mathsf{G}$ is a (real) vector subspace of $\mathbb{k}^{n\times n}$. □

Note : Since the vector space $\Bbbk^{n\times n}$ is *finite dimensional*, so is (the tangent space) $T_A\,\mathsf{G}$.

Definition 65. If $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ is a linear Lie group, its **dimension** is the dimension of the (real) vector space $T_{\mathbf{1}}\,\mathsf{G}$ ($\mathbf{1}$ is the identity matrix). So

$$\dim\,\mathsf{G} := \dim_{\mathbb{R}}\,T_{\mathbf{1}}\,\mathsf{G}.$$

Note : If the linear Lie group G is complex, then its *complex dimension* is

$$\dim_{\mathbb{C}}\,\mathsf{G} := \dim_{\mathbb{C}}\,T_{\mathbf{1}}\,\mathsf{G}.$$

◇ **Exercise 86.** Show that the matrix group $\mathsf{U}\,(1)$ has dimension 1.

Note : The *only* connected linear Lie groups (up to isomorphism) of dimension 1 are $\mathbb{T}^1 = \mathsf{U}\,(1)$ and $\mathbb{R}$, and of dimension 2 are $\mathbb{R}^2, \mathbb{T}^1 \times \mathbb{R}, \mathbb{T}^2$, and $\mathsf{AGL}^+\,(1,\mathbb{R})$.

Example 66. *The real general linear group* $\mathsf{GL}\,(n,\mathbb{R})$ *has dimension* n^2. The determinant function $\det : \mathbb{R}^{n\times n} \to \mathbb{R}$ is *continuous* and $\det\,(\mathbf{1}) = 1$. So there is some ϵ-ball about $\mathbf{1}$ in $\mathbb{R}^{n\times n}$ such that, for each A in this ball, $\det A \neq 0$ (i.e., $A \in \mathsf{GL}\,(n,\mathbb{R})$). If $B \in \mathbb{R}^{n\times n}$, then define a curve σ in $\mathbb{R}^{n\times n}$ by

$$\sigma(t) := \mathbf{1} + tB.$$

Then $\sigma(0) = \mathbf{1}$ and $\dot\sigma(0) = B$, and (for small t) $\sigma(t) \in \mathsf{GL}\,(n,\mathbb{R})$. Hence the tangent space $T_{\mathbf{1}}\,\mathsf{GL}\,(n,\mathbb{R})$ is all of $\mathbb{R}^{n\times n}$ which has dimension n^2. So $\dim\,\mathsf{GL}\,(n,\mathbb{R}) = n^2$.

◇ **Exercise 87.** Show that the dimension of the *complex* general linear group $\mathsf{GL}\,(n,\mathbb{C})$ is $2n^2$.

Proposition 46. *Let* $\mathsf{Sk\text{-}sym}\,(n)$ *denote the set of all skew-symmetric matrices in* $\mathbb{R}^{n\times n}$. *Then* $\mathsf{Sk\text{-}sym}\,(n)$ *is a linear subspace of* $\mathbb{R}^{n\times n}$ *and its dimension is* $\frac{n(n-1)}{2}\cdot$

Proof. If $A, B \in \mathsf{Sk\text{-}sym}\,(n)$, then

$$(A+B)^\top + (A+B) = A^\top + A + B^\top + B = 0$$

so that $\mathsf{Sk\text{-}sym}\,(n)$ is closed under (vector) addition.

It is also closed under scalar multiplication, for if $A \in \mathsf{Sk\text{-}sym}\,(n)$ and $\lambda \in \mathbb{R}$, then $(\lambda\,A)^\top = \lambda\,A^\top$ so that

$$(\lambda\,A)^\top + \lambda\,A = \lambda\,(A^\top + A) = 0.$$

To check the dimension of $\mathsf{Sk\text{-}sym}\,(n)$ we construct a basis. Let E_{ij} denote the matrix whose entries are all zero except the ij-entry, which is 1, and the ji-entry, which is -1. If we define these E_{ij} *only* for $i < j$, we can see that they form a *basis* for $\mathsf{Sk\text{-}sym}\,(n)$. It is easy to compute that there are

$$(n-1) + (n-2) + \cdots + 2 + 1 = \frac{n(n-1)}{2}$$

of them. □

◇ **Exercise 88.** Show that if σ is a curve through the identity (i.e., $\sigma(0) = \mathbf{1}$) in the orthogonal group $\mathsf{O}(n)$, then $\dot{\sigma}(0)$ is skew-symmetric.

NOTE : It follows that $\dim \mathsf{O}(n) \leq \frac{n(n-1)}{2} \cdot$ (Later we will show that this estimation is an equality.)

6.2. **Lie algebras.** We will adopt the notation $\mathfrak{g} := T_{\mathbf{1}}\,\mathsf{G}$ for this real vector subspace of $\Bbbk^{n\times n}$. In fact, $\mathfrak{g}$ has a more interesting algebraic structure, namely that of a *Lie algebra.*

NOTE : It is customary to use *lower case Gothic* (Fraktur) characters (such as $\mathfrak{a}, \mathfrak{g}$ and $\mathfrak{h}$) to refer to Lie algebras.

DEFINITION 67. A (real) **Lie algebra** $\mathfrak{a}$ is a real vector space equipped with a product

$$[\cdot,\cdot] : \mathfrak{a}\times\mathfrak{a}\to\mathfrak{a}, \qquad (x,y)\mapsto [x,y]$$

such that (for $\lambda,\mu\in\mathbb{R}$ and $x,y,z\in\mathfrak{a}$)

(LA1) $[x,y] = -[y,x]$.
(LA2) $[\lambda x + \mu y, z] = \lambda\,[x,z] + \mu\,[y,z]$.
(LA3) $[x,[y,z]] + [y,[z,x]] + [z,[x,y]] = 0$.

The product $[\cdot,\cdot]$ is called the *Lie bracket* of the Lie algebra $\mathfrak{a}$.

NOTE : (1) Condition (LA3) is called the *Jacobi identity.* So the Lie bracket $[\cdot,\cdot]$ of (the Lie algebra) $\mathfrak{a}$ is a *skew-symmetric bilinear* mapping (on $\mathfrak{a}$) which satisfies the Jacobi identity. Hence Lie algebras are *non-associative* algebras. The Lie bracket plays for Lie algebras the same role that the associative law plays for associative algebras.

(2) While we can define *complex* Lie algebras (or, more generally, Lie algebras over any field), we shall only consider Lie algebras over $\mathbb{R}$.

EXAMPLE 68. Let $\mathfrak{a} = \mathbb{R}^n$ and set (for all $x,y\in\mathbb{R}^n$)

$$[x,y] := 0.$$

The *trivial product* is a skew-symmetric bilinear multiplication (on $\mathbb{R}^n$) which satisfies the Jacobi identity and hence is a Lie bracket. $\mathbb{R}^n$ equipped with this product (Lie bracket) is a Lie algebra. Such a Lie algebra is called an *Abelian Lie algebra.*

◇ **Exercise 89.** Show that the *only* Lie algebra structure on (the vector space) $\mathbb{R}$ is the trivial one.

EXAMPLE 69. Let $\mathfrak{a} = \mathbb{R}^3$ and set (for $x,y\in\mathbb{R}^3$)

$$[x,y] := x\times y \qquad \text{(the cross product).}$$

For the *standard unit vectors* e_1, e_2, e_3 we have

$$[e_1, e_2] = -[e_2, e_1] = e_3, \quad [e_2, e_3] = -[e_3, e_2] = e_1, \quad [e_3, e_1] = -[e_1, e_3] = e_2.$$

Then $\mathbb{R}^3$ equipped with this bracket operation is a Lie algebra. In fact, as we will see later, this is the Lie algebra of (the matrix group) $\mathsf{SO}\,(3)$ and also of $\mathsf{SU}\,(2)$ in disguise.

Given two matrices $A, B \in \Bbbk^{n\times n}$, their **commutator** is

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

A and B *commute* (i.e., $AB = BA$) if and only if $[A, B] = 0$. The commutator $[\cdot, \cdot]$ is a product on $\Bbbk^{n\times n}$ satisfying conditions (LA1) – (LA3).

◇ **Exercise 90.** Verify the *Jacobi identity* for the commutator $[\cdot, \cdot]$.

The *real* vector space $\Bbbk^{n\times n}$ equipped with the commutator $[\cdot, \cdot]$ is a Lie algebra.

Note : The procedure to give $\Bbbk^{n\times n}$ a Lie algebra structure can be extended to any *associative* algebra. A Lie bracket can be defined in any associative algebra by the commutator $[x, y] = xy - yx$, making it a Lie algebra. Here the skew-symmetry condition (axiom) is clearly satisfied, and one can check easily that in this case the Jacobi identity for the commutator follows from the associativity law for the ordinary product.

There is another way in which Lie algebras arise in the study of algebras. A **derivation** d of a *non-associative* algebra $\mathcal{A}$ (i.e., a vector space endowed with a bilinear mapping $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$) is a linear mapping $\mathcal{A} \to \mathcal{A}$ satisfying the formal analogue of the Leibniz rule for differentiating a product (for all $x, y \in \mathcal{A}$)

$$d(xy) = (dx)y + x(dy).$$

(The concept of a derivation is an abstraction of the idea of a *first-order differential operator.*) The set of all derivations on $\mathcal{A}$ is clearly a vector subspace of the algebra $\mathsf{End}\,(\mathcal{A})$ of all linear mappings $\mathcal{A} \to \mathcal{A}$. Although the product of derivations is in general *not* a derivation, the commutator $d_1 \circ d_2 - d_2 \circ d_1$ of two derivations is again a derivation. Thus the set of all derivations of a non-associative algebra is a Lie algebra, called the *derivation algebra* of the given non-associative algebra.

Suppose that $\mathfrak{a}$ is a vector subspace of the Lie algebra $\Bbbk^{n\times n}$. Then $\mathfrak{a}$ is a **Lie subalgebra** of $\Bbbk^{n\times n}$ if it is closed under taking commutators of pairs of alements in $\mathfrak{a}$; that is,

$$A, B \in \mathfrak{a} \;\Rightarrow\; [A, B] \in \mathfrak{a}.$$

Of course, $\Bbbk^{n\times n}$ is a Lie subalgebra of itself.

Theorem 47. *If* $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ *is a linear Lie group, then the tangent space* $\mathfrak{g} = T_1\,\mathsf{G}$ *is a Lie subalgebra of* $\Bbbk^{n\times n}$.

Proof. We will show that two curves α, β in G with $\alpha(0) = \beta(0) = \mathbf{1}$, there is such a curve γ with $\dot\gamma(0) = [\dot\alpha(0),\, \dot\beta(0)]$, where $[\cdot, \cdot]$ is the matrix commutator.

Consider the mapping

$$F : (s,t) \mapsto F(s,t) := \alpha(s)\,\beta(t)\,\alpha(s)^{-1}.$$

This is clearly differentiable with respect to each of the variables s, t. For each s (in the domain of α), $F(s, \cdot)$ is a curve in G with $F(s,0) = \mathbf{1}$. Differentiating gives

$$\frac{d}{dt}F(s,t)\bigg|_{t=0} = \alpha(s)\,\dot\beta(0)\,\alpha(s)^{-1}$$

and so

$$\alpha(s)\,\dot\beta(0)\,\alpha(s)^{-1} \in \mathfrak{g}.$$

Since $\mathfrak{g}$ is a *closed subspace* of $\Bbbk^{n\times n}$ (any vector subspace is an intersection of hyperplanes), whenever this limit exists we also have

$$\lim_{s\to 0} \frac{1}{s}\Big(\alpha(s)\,\dot\beta(0)\,\alpha(s)^{-1} - \dot\beta(0)\Big) \in \mathfrak{g}.$$

◇ **Exercise 91.** Verify the following matrix version of the usual rule for differentiating an inverse :

$$\frac{d}{dt}\left(\alpha(t)^{-1}\right) = -\alpha(t)^{-1}\,\dot\alpha(t)\,\alpha(t)^{-1}.$$

We have

$$\begin{aligned}
\lim_{s\to 0} \frac{1}{s}\Big(\alpha(s)\,\dot\beta(0)\,\alpha(s)^{-1} - \dot\beta(0)\Big) &= \frac{d}{ds}\alpha(s)\,\dot\beta(0)\,\alpha(s)^{-1}\bigg|_{s=0}\\
&= \dot\alpha(0)\,\dot\beta(0)\,\alpha(0) - \alpha(0)\,\dot\beta(0)\,\alpha(0)^{-1}\dot\alpha(0)\,\alpha(0)^{-1}\\
&= \dot\alpha(0)\,\dot\beta(0)\,\alpha(0) - \alpha(0)\,\dot\beta(0)\,\dot\alpha(0)\\
&= \dot\alpha(0)\,\dot\beta(0) - \dot\beta(0)\,\dot\alpha(0)\\
&= [\dot\alpha(0),\, \dot\beta(0)].
\end{aligned}$$

This shows that $[\dot\alpha(0), \dot\beta(0)] \in \mathfrak{g}$, hence it must be of the form $\dot\gamma(0)$ for some curve γ. □

So, for each linear Lie group G, there is a Lie algebra $\mathfrak{g} = T_{\mathbf{1}}\,\mathsf{G}$. We call $\mathfrak{g}$ the **Lie algebra** of G.

NOTE : The essential phenomenon behind Lie theory is that *one may associate, in a natural way, to a linear Lie group* G *its Lie algebra* $\mathfrak{g}$. The Lie algebra is first of all a (real) vector space and secondly is endowed with a skew-symmetric bilinear product (the Lie bracket). Amazingly, *the group* G *is almost completely determined by* $\mathfrak{g}$ *and its Lie bracket.* Thus, for many purposes, one can replace G with $\mathfrak{g}$. Since G is a complicated nonlinear object and $\mathfrak{g}$ is just a vector space, it is usually vastly simpler to work with $\mathfrak{g}$. Otherwise intractable computations may become straightforward linear algebra; this is one source of the power of Lie theory.

6.3. **Homomorphisms of Lie algebras.** A suitable type of homomorphism $\mathsf{G} \to \mathsf{H}$ between linear Lie groups gives rise to a linear mapping $\mathfrak{g} \to \mathfrak{h}$ respecting the Lie algebra structures.

Definition 70. Let $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$, $\mathsf{H} \leq \mathsf{GL}\,(m, \Bbbk)$ be linear Lie groups and let $\Phi : \mathsf{G} \to \mathsf{H}$ be a continuous mapping. Then Φ is said to be **differentiable** if for every (differentiable) curve $\gamma : (a, b) \to \mathsf{G}$, the composite mapping $\Phi \circ \gamma : (a, b) \to \mathsf{H}$ is a (differentiable) curve with derivative

$$(\Phi \circ \gamma)^{\cdot}(t) = \frac{d}{dt}\Phi(\gamma(t))$$

and, if two (differentiable) curves $\alpha, \beta : (a, b) \to \mathsf{G}$ both satisfy the conditions

$$\alpha(0) = \beta(0) \quad \text{and} \quad \dot{\alpha}(0) = \dot{\beta}(0),$$

then $(\Phi \circ \alpha)^{\cdot}(0) = (\Phi \circ \beta)^{\cdot}(0)$.

Such a mapping Φ is a *differentiable homomorphism* if it is also a group homomorphism. A continuous homomorphism of matrix groups that is also *differentiable* is called a **Lie homomorphism**.

Note : The "technical restriction" in the definition of a Lie homomorphism is, in fact, *unnecessary.* (It turns out, but by no means easy to prove, that *every continuous homomorphism between Lie groups is differentiable* – in fact, analytic.)

If $\Phi : \mathsf{G} \to \mathsf{H}$ is the restriction of a differentiable mapping $\mathsf{GL}\,(n, \Bbbk) \to \mathsf{GL}\,(m, \Bbbk)$, then Φ is also a differentiable mapping.

Proposition 48. *Let* G, H, K *be linear Lie groups and let* $\Phi : \mathsf{G} \to \mathsf{H}, \Psi : \mathsf{H} \to \mathsf{K}$ *be differentiable homomorphisms.*

(a) *For each* $A \in \mathsf{G}$ *there is a linear mapping* $d\Phi_A : T_A\,\mathsf{G} \to T_{\Phi(A)}\,\mathsf{H}$ *given by*

$$d\Phi_A(\dot{\gamma}(0)) = (\Phi \circ \gamma)^{\cdot}(0).$$

(b) *We have*

$$d\Psi_{\Phi(A)} \circ d\Phi_A = d(\Psi \circ \Phi)_A.$$

(c) *For the identity mapping* $\mathbf{1}_{\mathsf{G}} : \mathsf{G} \to \mathsf{G}$ *and* $A \in \mathsf{G}$,

$$d\,(\mathbf{1}_{\mathsf{G}})_A = \mathbf{1}_{T_A\,\mathsf{G}}.$$

Proof. (a) The definition of $d\Phi_A$ makes sense since (by the definition of differentiability), given $X \in T_A\,\mathsf{G}$, for any curve γ with $\gamma(0) = A$ and $\dot{\gamma}(0) = X$, $(\Phi \circ \gamma)^{\cdot}(0)$ depends *only* on X and *not* on γ. The identities (b) and (c) are straightforward to verify. □

◇ **Exercise 92.** Verify that the map $d\Phi_A : T_A\,\mathsf{G} \to T_{\Phi(A)}\,\mathsf{H}$ is *linear.*

If $\Phi : \mathsf{G} \to \mathsf{H}$ is a differentiable homomorphism, then (since $\Phi(\mathbf{1}) = \mathbf{1}$) $d\Phi_{\mathbf{1}} : T_{\mathbf{1}}\,\mathsf{G} \to T_{\mathbf{1}}\,\mathsf{H}$ is a linear mapping, called the **derivative** of Φ and usually denoted by

$$d\Phi : \mathfrak{g} \to \mathfrak{h}.$$

Definition 71. Let $\mathfrak{g}, \mathfrak{h}$ be Lie algebras. A linear mapping $\phi : \mathfrak{g} \to \mathfrak{h}$ is a **homomorphism of Lie algebras** if (for $x, y \in \mathfrak{g}$)

$$\phi\,([x, y]) = [\phi\,(x), \phi\,(y)].$$

Theorem 49. *Let G, H be linear Lie groups and $\Phi : \mathsf{G} \to \mathsf{H}$ be a Lie homomorphism. Then the derivative $d\Phi : \mathfrak{g} \to \mathfrak{h}$ is a homomorphism of Lie algebras.*

Following ideas and notation in the proof of Theorem 47, for curves α, β in G with $\alpha(0) = \beta(0) = \mathbf{1}$, we can use the composite mapping

$$\Phi \circ F : (s, t) \mapsto \Phi(F(s, t)) = \Phi(\alpha(s))\Phi(\beta(t))\Phi(\alpha(s))^{-1}$$

to deduce $d\Phi([\dot\alpha(0), \dot\beta(0)]) = [d\Phi(\dot\alpha(0)), d\Phi(\dot\beta(0))]$.

◇ **Exercise 93.** Write down a full proof of Theorem 49.

Corollary 50. *Let G, H be linear Lie groups and $\Phi : \mathsf{G} \to \mathsf{H}$ be a Lie isomorphism of linear Lie groups. Then the derivative $d\Phi : \mathfrak{g} \to \mathfrak{h}$ is an isomorphism of Lie algebras.*

Proof. $\Phi^{-1} \circ \Phi$ is the identity, so

$$d\Phi^{-1} \circ d\Phi : T_{\mathbf{1}}\,\mathsf{G} \to T_{\mathbf{1}}\,\mathsf{G}$$

is the identity. Thus $d\Phi^{-1}$ is surjective and $d\Phi$ is injective.

Likewise, $\Phi \circ \Phi^{-1}$ is the identity, so $d\Phi \circ d\Phi^{-1}$ is the identity. Thus $d\Phi^{-1}$ is injective, and $d\Phi$ is surjective. The result now follows. □

Note : Isomorphic linear Lie groups have isomorphic Lie algebras. The converse (i.e., linear Lie groups with isomorphic Lie algebras are isomorphic) is *false.* For example, the rotation group $\mathsf{SO}\,(2)$ and the diagonal group

$$\mathsf{D}_1 = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & e^a \end{bmatrix} \,:\, a \in \mathbb{R} \right\} \leq \mathsf{AGL}^+\,(1, \mathbb{R})$$

both have Lie algebras isomorphic to $\mathbb{R}$ (the only Lie algebra structure on $\mathbb{R}$), but $\mathsf{SO}\,(2)$ is homeomorphic to a circle, while D_1 is homeomorphic to $\mathbb{R}$, so they are certainly *not* isomorphic.

However, the converse is – in a sense – *almost* true, so that the bracket operation on $\mathfrak{g}$ *almost* determines G as a group. After the existence of the Lie algebra, this fact is the most remarkable in Lie theory. Its precise formulation is known as *Lie's Third Theorem.*

6.4. **Lie algebras of linear Lie groups: examples.**

6.4.1. *The Lie algebras of* $\mathsf{GL}\,(n,\mathbb{R})$ *and* $\mathsf{GL}\,(n,\mathbb{C})$. Let us start with the *real* general linear group $\mathsf{GL}\,(n,\mathbb{R}) \subset \mathbb{R}^{n\times n}$. We have shown (see Example 90) that $T_{\mathbf{1}}\,\mathsf{GL}\,(n,\mathbb{R}) = \mathbb{R}^{n\times n}$. Hence *the Lie algebra* $\mathfrak{gl}(n,\mathbb{R})$ *of* $\mathsf{GL}\,(n,\mathbb{R})$ *consists of all* $n \times n$ *matrices (with real entries)*, with the commutator as the Lie bracket. Thus $\mathfrak{gl}\,(n,\mathbb{R}) = \mathbb{R}^{n\times n}$. It follows that

$$\dim\,\mathsf{GL}\,(n,\mathbb{R}) = \dim\,\mathfrak{gl}\,(n,\mathbb{R}) = n^2.$$

Similarly, the Lie algebra of the *complex* general linear group $\mathsf{GL}\,(n,\mathbb{C})$ is $\mathfrak{gl}\,(n,\mathbb{C}) = \mathbb{C}^{n\times n}$ and

$$\dim\,\mathsf{GL}\,(n,\mathbb{C}) = \dim_{\mathbb{R}}\,\mathfrak{gl}\,(n,\mathbb{C}) = 2n^2.$$

6.4.2. *The Lie algebras of* $\mathsf{SL}\,(n,\mathbb{R})$ *and* $\mathsf{SL}\,(n,\mathbb{C})$. For $\mathsf{SL}\,(n,\mathbb{R}) \leq \mathsf{GL}\,(n,\mathbb{R})$, suppose that $\alpha : (a,b) \to \mathsf{SL}\,(n,\mathbb{R})$ is a curve in $\mathsf{SL}\,(n,\mathbb{R})$ with $\alpha(0) = \mathbf{1}$. For $t \in (a,b)$ we have $\det\,\alpha(t) = 1$ and so

$$\frac{d}{dt}\det\,\alpha(t) = 0.$$

Using Proposition 37, it follows that $\operatorname{tr}\dot\alpha(0) = 0$ and thus

$$T_{\mathbf{1}}\,\mathsf{SL}\,(n,\mathbb{R}) \subseteq \operatorname{Ker}\operatorname{tr} := \left\{A \in \mathbb{R}^{n\times n} \,:\, \operatorname{tr}A = 0\right\}.$$

If $A \in \operatorname{Ker}\operatorname{tr} \subseteq \mathbb{R}^{n\times n}$, the curve $\alpha : (a,b) \to \mathbb{R}^{n\times n}, \quad t \mapsto \exp\,(tA)$ satisfies (the boundary conditions)

$$\alpha(0) = \mathbf{1} \quad \text{and} \quad \dot\alpha(0) = A.$$

Moreover, using Liouville's Formula, we get

$$\det\,\alpha(t) = \det\,\exp\,(tA) = e^{t\operatorname{tr}A} = 1.$$

Hence *the Lie algebra* $\mathfrak{sl}\,(n,\mathbb{R})$ *of* $\mathsf{SL}\,(n,\mathbb{R})$ *consists of all* $n \times n$ *matrices (with real entries) having trace zero*, with the commutator as the Lie bracket. Thus

$$\mathfrak{sl}\,(n,\mathbb{R}) = T_{\mathbf{1}}\,\mathsf{SL}\,(n,\mathbb{R}) = \{A \in \mathfrak{gl}\,(n,\mathbb{R}) \,:\, \operatorname{tr}A = 0\}.$$

Since $\operatorname{tr}A = 0$ imposes one condition on A, it follows that

$$\dim\mathsf{SL}\,(n,\mathbb{R}) = \dim_{\mathbb{R}}\,\mathfrak{sl}\,(n,\mathbb{R}) = n^2 - 1.$$

Similarly, *the Lie algebra of the complex special linear group* $\mathsf{SL}\,(n,\mathbb{C})$ *is*

$$\mathfrak{sl}\,(n,\mathbb{C}) = T_{\mathbf{1}}\,\mathsf{SL}\,(n,\mathbb{C}) = \{A \in \mathfrak{gl}\,(n,\mathbb{C}) \,:\, \operatorname{tr}A = 0\}$$

and

$$\dim\,\mathsf{SL}\,(n,\mathbb{C}) = \dim_{\mathbb{R}}\,\mathfrak{sl}\,(n,\mathbb{C}) = 2(n^2 - 1).$$

6.4.3. *The Lie algebras of* $\mathsf{O}\,(n)$ *and* $\mathsf{SO}\,(n)$. First, consider the orthogonal group $\mathsf{O}\,(n)$; that is,

$$\mathsf{O}\,(n) = \left\{A \in \mathsf{GL}\,(n,\mathbb{R}) \,:\, A^\top A = \mathbf{1}\right\} \leq \mathsf{GL}\,(n,\mathbb{R}).$$

Given a curve $\alpha : (a,b) \to \mathsf{O}\,(n)$ with $\alpha(0) = \mathbf{1}$, we have

$$\frac{d}{dt}\alpha(t)^T\alpha(t) = 0$$

and so

$$\dot{\alpha}(t)^\top\alpha(t) + \alpha(t)^\top\dot{\alpha}(t) = 0$$

which implies

$$\dot{\alpha}(0)^\top + \dot{\alpha}(0) = 0.$$

Thus we must have $\dot{\alpha}(0) \in \mathbb{R}^{n\times n}$ is *skew-symmetric*. So

$$T_{\mathbf{1}}\,\mathsf{O}\,(n) \subseteq \mathsf{Sk\text{-}sym}\,(n) = \left\{A \in \mathbb{R}^{n\times n} \,:\, A^\top + A = 0\right\}$$

(the set of all $n \times n$ skew-symmetric matrices in $\mathbb{R}^{n\times n}$).

On the other hand, if $A \in \mathsf{Sk\text{-}sym}\,(n) \subseteq \mathbb{R}^{n\times n}$, we consider the curve

$$\alpha : (a,b) \to \mathsf{GL}\,(n,\mathbb{R}), \quad t \mapsto \exp\,(tA).$$

Then

$$\begin{aligned}\alpha(t)^\top\alpha(t) &= \exp\,(tA)^\top \exp\,(tA)\\ &= \exp\,(tA^\top)\exp\,(tA)\\ &= \exp\,(-tA)\exp\,(tA)\\ &= \mathbf{1}.\end{aligned}$$

Hence we can view α as a curve in $\mathsf{O}\,(n)$. Since $\dot{\alpha}(0) = A$, this shows that

$$\mathsf{Sk\text{-}sym}\,(n) \subseteq T_{\mathbf{1}}\,\mathsf{O}\,(n)$$

and hence *the Lie algebra* $\mathfrak{o}(n)$ *of the orthogonal group* $\mathsf{O}\,(n)$ *consists of all* $n \times n$ *skew-symmetric matrices*, with the usual commutator as the Lie bracket. Thus

$$\mathfrak{o}(n) = T_{\mathbf{1}}\,\mathsf{O}\,(n) = \mathsf{Sk\text{-}sym}\,(n) = \left\{A \in \mathbb{R}^{n\times n} \,:\, A^\top + A = 0\right\}.$$

It follows that (see Proposition 46)

$$\dim\,\mathsf{O}\,(n) = \dim\,\mathfrak{o}(n) = \frac{n(n-1)}{2}\cdot$$

◇ **Exercise 94.** Show that if $A \in \mathsf{Sk\text{-}sym}\,(n)$, then $\operatorname{tr} A = 0$.

By Liouville's Formula, we have

$$\det \alpha(t) = \det \exp(tA) = 1$$

and hence $\alpha : (a,b) \to \mathsf{SO}(n)$, where $\mathsf{SO}(n)$ is the *special orthogonal group.* We have actually shown that *the Lie algebra of the special orthogonal group* $\mathsf{SO}(n)$ *is*

$$\mathfrak{so}(n) = \mathfrak{o}(n) = \left\{ A \in \mathbb{R}^{n\times n} \, : \, A^\top + A = 0 \right\}.$$

6.4.4. *The Lie algebra of* $\mathsf{SO}(3)$. We will discuss the Lie algebra $\mathfrak{so}(3)$ of the *rotation group* $\mathsf{SO}(3)$ in some detail.

◇ **Exercise 95.** Show that

$$\mathfrak{so}(3) = \left\{ \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \in \mathbb{R}^{3\times 3} \, : \, a,b,c \in \mathbb{R} \right\}.$$

The Lie algebra $\mathfrak{so}(3)$ is a three-dimensional real vector space. Consider the *rotations*

$$R_1(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}, \; R_2(t) = \begin{bmatrix} \cos t & 0 & \sin t \\ 0 & 1 & 0 \\ -\sin t & 0 & \cos t \end{bmatrix}, \; R_3 = \begin{bmatrix} \cos t & -\sin t & 0 \\ \sin t & \cos t & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then the mappings

$$\rho_i : t \mapsto R_i(t), \quad i = 1,2,3$$

are curves in $\mathsf{SO}(3)$ and clearly $\rho_i(0) = \mathbf{1}$. It follows that

$$\dot{\rho}_i(0) := A_i \in \mathfrak{so}(3), \quad i = 1,2,3.$$

These elements (matrices) are

$$A_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

◇ **Exercise 96.** Verify that the matrices A_1, A_2, A_3 form a basis for $\mathfrak{so}(3)$. (We shall refer to this basis as the *standard basis.*)

◇ **Exercise 97.** Compute all the Lie brackets (commutators) $[A_i, A_j]$, $i,j = 1,2,3$ and then determine the coefficients c_{ij}^k defined by

$$[A_i, A_j] = c_{ij}^1 A_1 + c_{ij}^2 A_2 + c_{ij}^3 A_3, \quad i,j = 1,2,3.$$

(These coefficients are called the **structure constants** of the Lie algebra. They completely determine the Lie bracket $[\cdot,\cdot]$.)

The Lie algebra $\mathfrak{so}(3)$ may be *identified* with (the Lie algebra) $\mathbb{R}^3$ as follows. We define the mapping

$$\widehat{\ }: \mathbb{R}^3 \to \mathfrak{so}\,(3), \quad x = (x_1, x_2, x_3) \mapsto \widehat{x} := \begin{bmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{bmatrix}.$$

This mapping is called the **hat map**.

◇ **Exercise 98.** Show that the hat map $\widehat{\ }: \mathbb{R}^3 \to \mathfrak{so}\,(3)$ is an *isomorphism* of vector spaces.

◇ **Exercise 99.** Show that (for $x, y \in \mathbb{R}^3$)

(a) $x \times y = \widehat{x}\, y$.
(b) $\widehat{x \times y} = [\widehat{x}, \widehat{y}]$.
(c) $x \bullet y = -\frac{1}{2}\mathrm{tr}\,(\widehat{x}\,\widehat{y})$.

Formula (b) says that the hat map is in fact an isomorphism of Lie algebras and so *we can identify the Lie algebra* $\mathfrak{so}(3)$ *with (the Lie algebra)* $\mathbb{R}^3$.

NOTE : For $x \in \mathbb{R}^3$ and $t \in \mathbb{R}$, the matrix exponential $\exp\,(t\,\widehat{x})$ is a *rotation* about (the axis) x through the angle $t\|x\|$. The following explicit formula for $\exp\,(\widehat{x})$ is known as RODRIGUES' FORMULA:

$$\exp\,(\widehat{x}) = \mathbf{1} + \frac{\sin\|x\|}{\|x\|}\,\widehat{x} + \frac{1}{2}\left[\frac{\sin\left(\frac{\|x\|}{2}\right)}{\frac{\|x\|}{2}}\right]^2 \widehat{x}^2.$$

This result says that *the exponential map*

$$\exp : \mathfrak{so}\,(3) \to \mathsf{SO}\,(3)$$

is onto. RODRIGUES' FORMULA is useful in *computational solid mechanics*, along with its quaternionic counterpart.

6.4.5. *The Lie algebras of* $\mathsf{U}\,(n)$ *and* $\mathsf{SU}\,(n)$. Consider the unitary group $\mathsf{U}\,(n)$; that is,

$$\mathsf{U}\,(n) = \{A \in \mathsf{GL}\,(n, \mathbb{C}) \,:\, A^*A = \mathbf{1}\}.$$

For a curve α in $\mathsf{U}\,(n)$ with $\alpha(0) = \mathbf{1}$, we obtain

$$\dot{\alpha}(0)^* + \dot{\alpha}(0) = 0$$

and so $\dot{\alpha}(0) \in \mathbb{C}^{n\times n}$ is skew-Hermitian. So

$$T_{\mathbf{1}}\,\mathsf{U}\,(n) \subseteq \mathsf{Sk\text{-}Herm}\,(n) = \left\{A \in \mathbb{C}^{n\times n} \,:\, A^* + A = 0\right\}$$

(the set of all $n \times n$ skew-Hermitian matrices in $\mathbb{C}^{n\times n}$).

If $H \in \mathsf{Sk\text{-}Herm}\,(n)$, then the curve

$$\alpha : (a, b) \to \mathsf{GL}\,(n, \mathbb{C}), \quad t \mapsto \exp\,(tH)$$

satisfies

$$\begin{aligned}\alpha(t)^*\alpha(t) &= \exp(tH)^*\exp(tH)\\ &= \exp(tH^*)\exp(tH)\\ &= \exp(-tH)\exp(tH)\\ &= \mathbf{1}.\end{aligned}$$

Hence we can view α as a curve in $\mathsf{U}(n)$. Since $\dot{\alpha}(0) = H$, this shows that

$$\mathsf{Sk\text{-}Herm}\,(n) \subseteq T_{\mathbf{1}}\,\mathsf{U}\,(n)$$

and hence *the Lie algebra* $\mathfrak{u}\,(n)$ *of the unitary group* $\mathsf{U}\,(n)$ *consists of all* $n \times n$ *skew-Hermitian matrices, with the usual commutator as the Lie bracket.* Thus

$$\mathfrak{u}(n) = T_{\mathbf{1}}\,\mathsf{U}\,(n) = \mathsf{Sk\text{-}Herm}\,(n) = \left\{H \in \mathbb{C}^{n\times n} \;:\; H^* + H = 0\right\}.$$

It follows that (see **Problem 39**)

$$\dim \mathsf{U}\,(n) = \dim_{\mathbb{R}} \mathfrak{u}\,(n) = n^2.$$

The special unitary group $\mathsf{SU}\,(n)$ can be handled in a similar way. Again we have

$$\mathfrak{su}\,(n) = T_{\mathbf{1}}\,\mathsf{SU}\,(n) \subseteq \mathsf{Sk\text{-}Herm}\,(n).$$

If $\alpha : (a, b) \to \mathsf{SU}\,(n)$ is a curve with $\alpha(0) = \mathbf{1}$ then, as in the analysis for $\mathsf{SL}\,(n, \mathbb{R})$, we have $\operatorname{tr}\dot{\alpha}(0) = 0$. Writing

$$\mathsf{Sk\text{-}Herm}^0\,(n) := \{H \in \mathsf{Sk\text{-}Herm}\,(n) \;:\; \operatorname{tr} H = 0\}$$

this gives $\mathfrak{su}\,(n) \subseteq \mathsf{Sk\text{-}Herm}^0\,(n)$. On the other hand, if $H \in \mathsf{Sk\text{-}Herm}^0\,(n)$ then the curve

$$\alpha : (a, b) \to \mathsf{U}\,(n), \quad t \mapsto \exp(tH)$$

takes values in $\mathsf{SU}\,(n)$ and $\dot{\alpha}(0) = H$. Hence

$$\mathfrak{su}\,(n) = T_{\mathbf{1}}\,\mathsf{SU}\,(n) = \mathsf{Sk\text{-}Herm}^0\,(n) = \left\{H \in \mathbb{C}^{n\times n} \;:\; H^* + H = 0 \text{ and } \operatorname{tr} H = 0\right\}.$$

Note : For a linear Lie group $\mathsf{G} \leq \mathsf{GL}\,(n, \mathbb{R})$ (with Lie algebra $\mathfrak{g}$), the following are true (and can be used in determining Lie algebras of linear Lie groups).

- The mapping

  $$\exp_{\mathsf{G}} : \mathfrak{g} \to \mathsf{GL}\,(n, \mathbb{R}), \quad X \mapsto \exp(X)$$

  has image contained in G, $\exp_{\mathsf{G}}(\mathfrak{g}) \subseteq \mathsf{G}$. We will normally write $\exp_{\mathsf{G}} : \mathfrak{g} \to \mathsf{G}$ for the exponential mapping on G (and sometimes even just exp). In general, the exponential mapping $\exp_{\mathsf{G}}$ is neither one-to-one nor onto.
- If G is compact and connected, then $\exp_{\mathsf{G}}$ is onto.
- The mapping $\exp_{\mathsf{G}}$ maps a neighborhood of 0 in $\mathfrak{g}$ bijectively onto a neighborhood of $\mathbf{1}$ in G.

◇ **Exercise 100.** Verify that the exponential map

$$\exp_{\mathsf{U}\,(1)} : \mathbb{R} \to \mathsf{U}\,(1) = \mathbb{S}^1, \qquad t \mapsto e^{it}$$

is onto but *not* one-to-one.

Example 72. The exponential map

$$\exp_{\mathsf{SL}\,(2,\mathbb{R})} : \mathfrak{sl}\,(2,\mathbb{R}) \to \mathsf{SL}\,(2,\mathbb{R})$$

is *not* onto. Let

$$A = \begin{bmatrix} \lambda & 0 \\ 0 & \frac{1}{\lambda} \end{bmatrix} \quad \text{with } \lambda < -1.$$

We see that $A \in \mathsf{SL}\,(2,\mathbb{R})$ and we shall show that A is *not* of the form $\exp(X)$ with $X \in \mathfrak{sl}\,(2,\mathbb{R})$. If $A = \exp(X)$, then the eigenvalues of A are of the form e^a and e^b, where a and b are the eigenvalues of X. Suppose $\lambda = e^a$ and $\frac{1}{\lambda} = e^b$. Then

$$a = -b + 2k\pi\, i, \quad k \in \mathbb{Z}.$$

However, since λ is negative, a is actually complex and therefore its conjugate is also an eigenvalue; that is, $b = \bar{a}$. This gives a as pure imaginary. Then

$$1 = |e^a| = |\lambda| = -\lambda$$

which contradicts the assumption that $\lambda < -1$.

6.4.6. *The Lie algebra of* $\mathsf{SU}\,(2)$. We will discuss the Lie algebra $\mathfrak{su}\,(2)$ in some detail.

◇ **Exercise 101.** Show that

$$\mathfrak{su}\,(2) = \left\{ \begin{bmatrix} ci & -b+ai \\ b+ai & -ci \end{bmatrix} \in \mathbb{C}^{2\times 2} \ : \ a,b,c \in \mathbb{R} \right\}.$$

The Lie algebra $\mathfrak{su}\,(2)$ is a three-dimensional *real* vector space. Consider the matrices

$$H_1 = \frac{1}{2}\begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad H_2 = \frac{1}{2}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad H_3 = \frac{1}{2}\begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

Clearly,

$$H_i \in \mathfrak{su}\,(2), \quad i = 1,2,3.$$

◇ **Exercise 102.** Verify that the matrices H_1, H_2, H_3 form a basis for $\mathfrak{su}\,(2)$.

◇ **Exercise 103.** Compute all the Lie brackets (commutators) $[H_i, H_j]$, $i,j = 1,2,3$ and then determine the structure constants of the Lie algebra $\mathfrak{su}\,(2)$.

Consider the mapping

$$\phi : \mathbb{R}^3 \to \mathfrak{su}\,(2), \quad x = (x_1,x_2,x_3) \mapsto x_1H_1 + x_2H_2 + x_3H_3.$$

$\diamond$ **Exercise 104.** Show that the mapping $\phi : \mathbb{R}^3 \to \mathfrak{su}\,(2)$ is an *isomorphism* of Lie algebras ($\mathbb{R}^3$ with the cross product).

Thus *we can identify the Lie algebra* $\mathfrak{su}\,(2)$ *with (the Lie algebra)* $\mathbb{R}^3$.

NOTE : The Lie algebras $\mathfrak{su}\,(2)$ and $\mathfrak{so}\,(3)$ look the same algebraically (they are isomorphic). An explicit isomorphism (of Lie algebras) is given by

$$\psi : x_1H_1 + x_2H_2 + x_3H_3 \mapsto x_1A_1 + x_2A_2 + x_3A_3.$$

This suggests that there might be a close relationship between the matrix groups themselves. Indeed there is a (surjective) *Lie homomorphism* $\mathsf{SU}\,(2) \to \mathsf{SO}\,(3)$ whose derivative (at $\mathbf{1}$) is ψ. Recall the adjoint representation

$$\mathrm{Ad} : \mathsf{SU}\,(2) \to \mathsf{Aut}\,(\mathfrak{su}(2)), \quad A \mapsto \mathrm{Ad}_A\,(: U \mapsto AUA^*).$$

Each Ad_A is a *linear isomorphism* of $\mathfrak{su}\,(2)$. Ad_A is actually an *orthogonal transformation* on $\mathfrak{su}\,(2)$ (the mapping $(X, Y) \mapsto -\mathrm{tr}\,(XY)$ is an inner product on $\mathfrak{su}\,(2)$) and so Ad_A corresponds to an element of $\mathsf{O}\,(3)$ (in fact, $\mathsf{SO}\,(3)$). The mapping

$$\overline{\mathrm{Ad}} : \mathsf{SU}\,(2) \to \mathsf{SO}\,(3), \quad A \mapsto \mathrm{Ad}_A$$

turns out to be a continuous homomorphism of matrix groups that is differentiable (i.e., a Lie homomorphism) and such that its derivative $d\,\overline{\mathrm{Ad}} : \mathfrak{su}\,(2) \to \mathfrak{so}\,(3)$ is ψ.

6.4.7. *The Lie algebras of* $\mathsf{T}\,(n, \Bbbk)$ *and* $\mathsf{UT}\,(n, \Bbbk)$. Let $\alpha : (a, b) \to \mathsf{T}\,(n, \Bbbk)$ be a curve in $\mathsf{T}\,(n, \Bbbk)$ with $\alpha(0) = \mathbf{1}$. Then $\dot{\alpha}(0)$ is upper triangular. Moreover, using the argument for $\mathsf{GL}\,(n, \Bbbk)$ we see that given any upper triangular matrix $A \in \Bbbk^{n\times n}$, there is a curve

$$\sigma : (-\epsilon, \epsilon) \to \Bbbk^{n\times n}, \quad t \mapsto \mathbf{1} + tA$$

such that $\sigma(t) \in \mathsf{T}\,(n, \Bbbk)$ and $\dot{\sigma}(0) = A$. Hence *the Lie algebra* $\mathfrak{t}\,(n, \Bbbk)$ *of* $\mathsf{T}\,(n, \Bbbk)$ *consists of all* $n \times n$ *upper triangular matrices*, with the usual commutator as the Lie bracket. Thus

$$\mathfrak{t}\,(n, \Bbbk) = T_{\mathbf{1}}\,\mathsf{T}\,(n, \Bbbk) = \left\{A \in \Bbbk^{n\times n} \;:\; a_{ij} = 0 \;\text{ for } i > j\right\}.$$

It follows that

$$\dim\,\mathsf{T}\,(n, \Bbbk) = \dim_{\mathbb{R}}\,\mathfrak{t}\,(n, \Bbbk) = \frac{n(n+1)}{2}\,\dim_{\mathbb{R}}\,\Bbbk.$$

An upper triangular matrix $A \in \Bbbk^{n\times n}$ is *strictly upper triangular* if all its diagonal entries are 0. Then *the Lie algebra of the unipotent group* $\mathsf{UT}\,(n, \Bbbk)$ *consists of all* $n \times n$ *strictly upper triangular matrices*, with the usual commutator as the Lie bracket. So

$$\mathfrak{ut}\,(n, \Bbbk) = T_{\mathbf{1}}\,\mathsf{UT}\,(n, \Bbbk) = \left\{A \in \Bbbk^{n\times n} \;:\; a_{ij} = 0 \;\text{ for } i \geq j\right\}.$$

$\diamond$ **Exercise 105.** Find $\dim_{\mathbb{R}}\,\mathfrak{ut}\,(n, \Bbbk)$.

Problems (33–44)

(33) Let $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ be a linear Lie group.

(a) Prove that if $A = \dot{\alpha}(0) \in T_{\mathbf{1}}\,\mathsf{G}$, then $\exp(A) \in \mathsf{G}$. (This means that the matrix exponential map $\exp : \Bbbk^{n\times n} \to \mathsf{GL}\,(n, \Bbbk)$ maps the Lie algebra $\mathfrak{g} = T_{\mathbf{1}}\,\mathsf{G}$ into G.)

(b) Hence deduce that

$$T_{\mathbf{1}}\,\mathsf{G} = \{X \in \Bbbk^{n\times n} \,:\, \exp(tX) \in \mathsf{G} \text{ for all } t \in \mathbb{R}\}.$$

(34) Let G be a linear Lie group. Prove that the following statements are logically equivalent.

(a) Any two elements of G may be joined by a path in G.

(b) G is *not* the disjoint union of two non-empty open sets.

(c) G is generated by any neighborhood of $\mathbf{1}$.

(d) G is generated by $\exp(\mathfrak{g})$. (A subset of G *generates* G if every element of G is a finite product of elements of the subset and their inverses; in this case, it means that every element of G is of the form $\exp(X_1)\exp(X_2)\cdots\exp(X_k)$ for some $X_1, X_2, \ldots, X_k$ in the Lie algebra $\mathfrak{g}$ of G.)

(35) Let G be a linear Lie group with associated Lie algebra $\mathfrak{g}$. Prove that the group G is Abelian if and only if the Lie algebra $\mathfrak{g}$ is Abelian (i.e., $[x, y] = 0$ for all $x, y \in \mathfrak{g}$).

(36) Let $\mathfrak{g}$ be a (real) Lie algebra. A vector subspace $\mathfrak{k}$ of $\mathfrak{g}$ is called an **ideal** if

$$[x, y] \in \mathfrak{k} \quad \text{for all } x \in \mathfrak{g},\ y \in \mathfrak{k}.$$

(a) Verify that any ideal $\mathfrak{k}$ is a Lie subalgebra (of $\mathfrak{g}$).

(b) Show that the **center** of $\mathfrak{g}$

$$\mathfrak{z}\,(\mathfrak{g}) := \{x \in \mathfrak{g} \,:\, [x, y] = 0 \text{ for all } y \in \mathfrak{g}\}$$

is an ideal in $\mathfrak{g}$.

(c) Show that the vector subspace

$$[\mathfrak{g}, \mathfrak{g}] := \operatorname{span}\{[x, y] \,:\, x, y \in \mathfrak{g}\}$$

is an ideal in $\mathfrak{g}$. (It is called the **commutator subalgebra**.)

(d) Show that the set

$$\mathfrak{sl}\,(n, \Bbbk) := \{x \in \Bbbk^{n\times n} \,:\, \operatorname{tr} x = 0\}$$

is an ideal in $\Bbbk^{n\times n}$. (It is called the **special linear Lie algebra.)**

(37) Show that if $\phi : \mathfrak{g}_1 \to \mathfrak{g}_2$ is a Lie algebra homomorphism, then the kernel $\operatorname{Ker}\phi$ is an ideal of $\mathfrak{g}_1$, and the image $\operatorname{Im}\phi$ is a Lie subalgebra of $\mathfrak{g}_2$.

(38) Let $\mathsf{G} \leq \mathsf{GL}\,(n, \Bbbk)$ be a linear Lie group. Prove that if H is a *normal* subgroup of G, then $T_{\mathbf{1}}\,\mathsf{H}$ is an *ideal* of the Lie algebra $T_{\mathbf{1}}\,\mathsf{G}$.

(39) A matrix $A \in \mathbb{C}^{n\times n}$ is called **skew-Hermitian** if $A^* + A = 0$.

(a) Show that the diagonal terms of a skew-Hermitian matrix are purely imaginary and hence deduce that the set $\mathsf{Sk\text{-}Herm}\,(n)$ of all skew-Hermitian matrices in $\mathbb{C}^{n\times n}$ is *not* a vector space over $\mathbb{C}$.

(b) Prove that $\mathsf{Sk\text{-}Herm}\,(n)$ is a *real* vector space of dimension

$$n + 2\,\frac{n(n-1)}{2} = n^2.$$

(c) If σ is a curve through the identity in $\mathsf{U}\,(n)$, show that $\dot{\sigma}(0)$ is skew-Hermitian and hence

$$\dim \mathsf{U}\,(n) \leq n^2.$$

(40) Consider the set (of $n \times n$ **skew-symmetric** matrices)

$$\mathfrak{so}\,(n) = \{x \in \mathbb{R}^{n\times n} \, : \, x^\top + x = 0\}.$$

(It is called the special **orthogonal Lie algebra**.)

(a) Show that $\mathfrak{so}\,(n)$ is a Lie subalgebra of $\mathbb{R}^{n\times n}$.

(b) Show that the Lie algebra $\mathfrak{so}\,(3)$ contains no ideals other than itself and (the trivial ideal) $\{0\}$. (Such a Lie algebra is called **simple**.)
[Hint : Show that any non-trivial ideal must contain *all* the elements of the standard basis.]

(41) For each of the following linear Lie group G, determine its Lie algebra $\mathfrak{g}$ and hence its dimension.

(a) $\mathsf{G} = \left\{A \in \mathsf{GL}\,(2,\mathbb{R}) \, : \, A^\top Q\, A = Q\right\}$, where $Q = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

(b) $\mathsf{G} = \left\{A \in \mathsf{GL}\,(2,\mathbb{R}) \, : \, A^\top Q\, A = Q\right\}$, where $Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$.

(c) $\mathsf{G} = \mathsf{AGL}\,(3,\mathbb{R})$.

(d) $\mathsf{G} = \mathsf{H}_3$.

(e) $\mathsf{G} = \mathsf{G}_4 \leq \mathsf{UT}^u\,(4,\mathbb{R})$ from **Exercise 65**.

(f) $\mathsf{G} = \mathsf{E}\,(n)$.

(g) $\mathsf{G} = \mathsf{SE}\,(n)$.

(42) (a) Show that the Lie algebra of the symplectic group $\mathsf{Sp}\,(2n,\mathbb{R})$ is

$$\mathfrak{sp}\,(2n,\mathbb{R}) = \left\{A \in \mathbb{R}^{2n\times 2n} \, : \, A^\top \mathbb{J} + \mathbb{J}A = 0\right\}.$$

(b) If

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathfrak{sl}\,(2n,\mathbb{R})$$

show that $A \in \mathfrak{sp}\,(2n, \mathbb{R})$ <u>if and only if</u>

$$d = -a^\top, \quad c = c^\top, \quad \text{and} \quad b = b^\top.$$

(c) Calculate the dimension of $\mathfrak{sp}\,(2n, \mathbb{R})$.

(43) Show that the Lie algebra of the Lorentz group Lor is

$$\mathfrak{lor} = \left\{ A \in \mathbb{R}^{4\times 4} \ : \ SA + A^\top S = 0 \right\} = \left\{ \begin{bmatrix} 0 & a_1 & a_2 & a_3 \\ -a_1 & 0 & a_4 & a_5 \\ -a_2 & -a_4 & 0 & a_6 \\ a_3 & a_5 & a_6 & 0 \end{bmatrix} \ : \ a_1, a_2, a_3, a_4, a_5, a_6 \in \mathbb{R} \right\}.$$

(44) Consider the linear Lie group $\Bbbk^\times = \mathsf{GL}\,(1, \Bbbk)$. (Its Lie algebra is clearly $\Bbbk$.)

(a) Show that the determinant function

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

is a *Lie homomorphism* (i.e., a continuous homomorphism of linear Lie groups that is also *differentiable*; cf. DEFINITION 70).

(b) Show that the induced *homomorphism* of Lie algebras (i.e., the derivative of det) is the trace function

$$\mathrm{tr} : \Bbbk^{n\times n} \to \Bbbk.$$

(c) Derive from (b) that (for $A, B \in \Bbbk^{n\times n}$)

$$\mathrm{tr}\,(AB) = \mathrm{tr}\,(BA).$$

7. Groups and Geometry (**Optional**)

Geometries • The "Erlanger Programm" • Classical geometries • Other geometries (in the sense of Klein).

7.1. **Geometries.** ....

....

7.2. **The "Erlanger Programm".** ....

....

7.3. **Classical geometries.** ....

....

7.3.1. *Euclidean geometry.*

7.3.2. *Spherical geometry.*

7.3.3. *Elliptic geometry.*

7.3.4. *Hyperbolic geometry.*

7.3.5. *Affine geometry.*

7.3.6. *Projective geometry.*

7.4. **Other geometries (in the sense of Klein).** ....

....

7.4.1. *Galilean geometry.*

7.4.2. *Lorentzian geometry.* ....

....

References

[1] M.A. Armstrong, *Groups and Symmetry*, Springer, 1988.

[2] A. Arvanitoyeorgos, *An Introduction to Lie Groups and the Geometry of Homogeneous Spaces*, American Mathematical Society, 2003.

[3] M. Atiyah, What is geometry ?, *Math. Gazette* **66**(437)(1982), 179–184.

[4] A. Baker, *Matrix Groups. An Introduction to Lie Group Theory*, Springer, 2002.

[5] W. Barker, R. Howe, *Continuous Symmetry. From Euclid to Klein*, American Mathematical Society, 2007.

[6] A.F. Beardon, *Algebra and Geometry*, Cambridge University Press, 2005.

[7] M. Berger, *Geometry I*, Springer, 1987.

[8] M. Berger, *Geometry II*, Springer, 1987.

[9] G.S. Birman, K. Nomizu, Trigonometry in Lorentzian geometry, *Amer. Math. Monthly* **91**(9)(1984), 543–549.

[10] R.P. Burn, *Groups. A Path to Geometry*, Cambridge University Press, 1985.

[11] J.J. Callahan, *The Geometry of Spacetime. An Introduction to Special and General Relativity*, Springer, 2000.

[12] S.S. Chern, From triangles to manifolds, *Amer. Math. Monthly* **86**(5)(1979), 339–349.

[13] S.S. Chern, What is geometry ?, *Amer. Math. Monthly* **97**(8)(1990), 679–686.

[14] J.N. Clelland, *From Frenet to Cartan: The Method of Moving Frames*, American Mathematical Society, 2017.

[15] V. Climenhaga, A. Katok, *From Groups to Geometry and Back*, American Mathematical Society, 2017.

[16] M. Henle, *Modern Geometries. Non-Euclidean, Projective, and Discrete* (Second Edition), Prentice Hall, 2001.

[17] R. Howe, Very basic Lie theory, *Amer. Math. Monthly* **90**(9)(1983), 600–623.

[18] B. Iversen, *Hyperbolic Geometry*, Cambridge University Press, 1992.

[19] B. Komrakov, A. Churyumov, B. Doubrov, Two-Dimensional Homogeneous Spaces, International Sophus Lie Center, Oslo, 1993.

[20] R.S. Millman, Kleinian transformation geometry, *Amer. Math. Monthly* **84**(5)(1977), 338–349.

[21] G.l. Naber, *The Geometry of Minkowski Spacetime. An Introduction to the Mathematics of the Special Theory of Relativity*, Dover, 1992.

[22] P.M. Neumann, G.A. Stoy, E.C. Thompson, *Groups and Geometry*, Oxford University Press, 1994.

[23] J.G. Ratcliffe, *Foundations of Hyperbolic Manifolds* (Second Edition), Springer, 2006.

[24] E.G. Rees, *Notes on Geometry*, Springer, 1983.

[25] M. Reid, B. Szendröi, *Geometry and Topology*, Cambridge University Press, 2005.

[26] W. Rossmann, *Lie Groups. An Introduction Through Linear Groups*, Oxford University Press, 2002.

[27] R.W. Sharpe, *Differential Geometry. Cartan's Generalization of Klein's Erlangen Program*, Springer, 1997.

[28] T.Q. Sibley, *Thinking Geometrically: A Survey of Geometries*, American Mathematical Society, 2015.

[29] A.B. Sossinsky, *Geometries*, American Mathematical Society, 2012.

[30] J. Stillwell, *Geometry of Surfaces*, Springer, 1992.

[31] J. Stillwell, *The Four Pillars of Geometry*, Springer, 2005.

[32] J. Stillwell, *Naive Lie Theory*, Springer, 2008.

[33] K. Tapp, *Matrix Groups for Undergraduates* (Second Edition), American Mathematical Society, 2016.

[34] I.M. Yaglom, *A Simple Non-Euclidean Geometry and Its Physical Basis*, Springer, 1979.

[35] I.M. Yaglom, *Felix Klein and Sophus Lie. Evolution of the Idea of Symmetry in the Nineteenth Century*, Birkhäuser, 1988.

Department of Mathematics, Rhodes University, 6140 Grahamstown, South Africa
E-mail address: c.c.remsing@ru.ac.za