Chapter 1

Groups of transformations

In this chapter we introduce the concepts of transformation groups and symmetry groups, and present as examples the symmetry groups of an equilateral triangle and of a circle, and the symmetric group S_n, the group of all permutations of n objects. A convenient way to present a permutation is as a product of commuting cycles.
The active and passive interpretations are different but equivalent ways to apply the mathematical formalism to physical problems.
The action of a transformation group on some set of points splits the set into disjoint subsets, called orbits. An orbit is called a homogeneous space, because any two points are transformed into each other. We may identify one arbitrary point of a homogeneous space with its fixed point group, and then all other points with left cosets of the fixed point group.
Different transformation groups may have the same multiplication table, they are then different realizations of the same abstract group. An abstract group has an associative group product, and is completely defined by its multiplication table. Abstract group theory will be the subject of the next chapter.

Group theory is the mathematical theory of symmetry, in physics, in mathematics, or anywhere. The basic insight is that symmetry can be described in terms of transformations. A *symmetry transformation*, or a *symmetry* for short, is a transformation of some object preserving certain properties of that object.

By definition, a transformation takes an initial state as input and produces a final state as output. When two transformations act one after the other, an initial state is transformed via an intermediate state into a final state. Forgetting the intermediate state, we see the process as one single transformation, the *functional composition* of the two transformations.

It follows immediately from these definitions that the composition of two symmetry transformations is again a symmetry transformation. Thus, the set of all symmetry transformations of a given object is a *symmetry group*, which is closed under the mathematical operation of functional composition. The composition operator is like a multiplication operator for transformations, and it provides the symmetry group with an interesting mathematical structure, a *group product*.

The symmetry group with its group product gives a precise meaning to the otherwise somewhat vague concept of the symmetry of an object. In most, if not in all, applications of group theory, the groups that turn up are symmetry groups of various kinds.

1.1 Symmetry of the equilateral triangle

An example to illustrate the concepts is the equilateral triangle shown in Figure 1.1. Its symmetry group consists of all possible transformations mapping it into itself with no distortions.

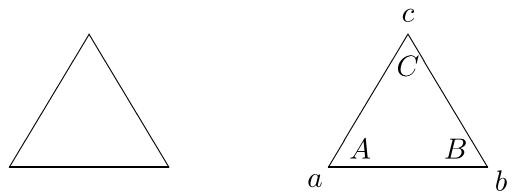

Figure 1.1: An equilateral triangle, unlabelled (left) and with its corners and corner positions labelled (right).

In order to describe the transformations of the triangle we label its corners, for example by A, B, C as shown in Figure 1.1. The corners are supposed to be identical, and we need to label them in order to keep track of how the triangle moves when we transform it. We also need to label the corner positions, for example by a, b, c as shown. The labels a, b, c stay in place when the triangle moves, they are our coordinate system.

There are altogether six different ways to transform the equilateral triangle into itself without distorting it. We may start by moving the corner A to any one of the three positions a, b, c. After that there are two free positions where we may put the corner B, and finally there is one position left for the corner C.

The six transformations may be classified as three rotations and three reflections. As an example, Figure 1.2 shows the horizontal reflection, that is, the reflection about the vertical line through the upper corner, at c. We denote this transformation by m_c, writing "m" for "mirror symmetry" and subscript "c" to indicate that the corner at c is fixed. Clearly there are two more reflection symmetries, to be denoted m_a and m_b.

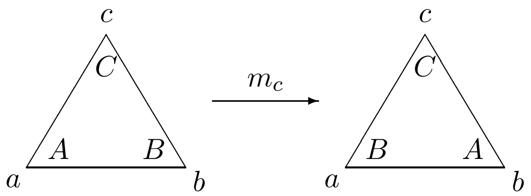

Figure 1.2: The horizontal reflection m_c.

Two more symmetry transformations of the equilateral triangle are the rotations by $2\pi/3$ in the positive (counterclockwise) and in the negative (clockwise) direction. Let us denote the positive rotation by r_+, and the negative rotation by r_-. A third, rather special kind of rotation is the one with zero rotation angle, this is the identity transformation, which we call I. To summarize, the symmetry group consists of the three rotations

$$\begin{array}{c}C\\A\;B\end{array}\xrightarrow{I}\begin{array}{c}C\\A\;B\end{array}\qquad\qquad\begin{array}{c}C\\A\;B\end{array}\xrightarrow{r_+}\begin{array}{c}B\\C\;A\end{array}\qquad\qquad\begin{array}{c}C\\A\;B\end{array}\xrightarrow{r_-}\begin{array}{c}A\\B\;C\end{array}\tag{1.1}$$

and the three reflections

$$\begin{array}{c} C \\ A\ B \end{array} \xrightarrow{m_a} \begin{array}{c} B \\ A\ C \end{array} \qquad \begin{array}{c} C \\ A\ B \end{array} \xrightarrow{m_b} \begin{array}{c} A \\ C\ B \end{array} \qquad \begin{array}{c} C \\ A\ B \end{array} \xrightarrow{m_c} \begin{array}{c} C \\ B\ A \end{array} \tag{1.2}$$

We have now introduced the characters of the play and are ready for the plot. Let two symmetry transformations act in succession, one after the other. For example, apply first the reflection m_a and afterwards the reflection m_b. The result is indeed a third symmetry transformation, the rotation r_-,

$$\begin{array}{c} C \\ A\ B \end{array} \xrightarrow{m_a} \begin{array}{c} B \\ A\ C \end{array} \xrightarrow{m_b} \begin{array}{c} A \\ B\ C \end{array} \quad \Longleftrightarrow \quad \begin{array}{c} C \\ A\ B \end{array} \xrightarrow{r_-} \begin{array}{c} A \\ B\ C \end{array} \tag{1.3}$$

If we apply the two reflections in the opposite order, first m_b and then m_a, the result is the rotation r_+ instead of r_-,

$$\begin{array}{c} C \\ A\ B \end{array} \xrightarrow{m_b} \begin{array}{c} A \\ C\ B \end{array} \xrightarrow{m_a} \begin{array}{c} B \\ C\ A \end{array} \quad \Longleftrightarrow \quad \begin{array}{c} C \\ A\ B \end{array} \xrightarrow{r_+} \begin{array}{c} B \\ C\ A \end{array} \tag{1.4}$$

We may use the symbol "$\circ$" to denote in an explicit way composition of transformations, thus we may write

$$m_b \circ m_a = r_- \ , \qquad m_a \circ m_b = r_+ \ . \tag{1.5}$$

But we will most often write the same relations simply as

$$m_b m_a = r_- \ , \qquad m_a m_b = r_+ \ . \tag{1.6}$$

1.2 Active and passive transformations

We think of these transformations here as *active* transformations, actually moving the triangle. Alternatively, we could have taken the *passive* point of view, letting the triangle stay in place and changing the coordinate system instead.

We will not worry too much here about the distinction between active and passive transformations. The two interpretations are equivalent according to the principle of relativity, which denies the existence of absolute space, so that it has no meaning to speak of the position of our triangle unless the position is defined relative to a coordinate system. In Figure 1.1, corner A is at position a, B is at position b and C at position c. It is the relation between the coordinate system and the triangle, $(aA)(bB)(cC)$, that defines the position. We may transform that relation *actively*, by moving the triangle, or *passively*, by changing the coordinate system, but the mathematics is the same in the end.

For example, if we rotate the triangle by $2\pi/3$ in the anticlockwise direction, we transform the position $(aA)(bB)(cC)$ into $(aC)(bA)(cB)$. The effect is the same if we rotate the rest of the universe by $2\pi/3$ in the clockwise direction.

It may seem strange to call it a passive transformation when we change the coordinate system by rotating the whole universe minus one small triangle. However, there is a lazy way to do the transformation. We need not rotate the universe, it is enough to just permute the labels a, b, c of the coordinate system, and then it is certainly justified to speak of a passive transformation.

In the real world, some kinds of active transformations are easy to perform, while others are more or less impossible. To actively rotate a physical object is usually easy, but there is no way to actively produce a mirror image of an object without disassembling and rebuilding it. A passive transformation, defined as an innocent relabelling of the coordinate system, is more easily performed.

To conclude, the active and passive points of view are different physical interpretations of the same mathematical formalism. Here we will concentrate on the mathematics, and will not commit ourselves to one or the other physical interpretation.

The relation between active and passive transformations is the subject of Problem 1.3 at the end of this chapter. See also Problem 1.4.

A different set of symmetry transformations

When we adopt the definition of a passive transformation as a permutation of the labels a, b, c of the coordinate system, we face an obvious alternative. Why not define a symmetry transformation as a permutation of the labels A, B, C of the triangle corners?

This would be a very natural approach. After all, the triangle is symmetric precisely because there is no visible difference between its three corners, there is no way to look at one corner and tell whether it is corner A, B, or C, unless we have already labelled it. The corner labels A, B, C are arbitrary, and we may interchange them as we like. The surprising fact is that this alternative definition leads to a different set of symmetry transformations.

We will return to this point in Section 1.11 below, but let us consider the reflection m_a as an example. In eq. (1.3), m_a acts on the triangle in the position $(aA)(bB)(cC)$, shown in Figure 1.1, and has the effect of interchanging the corners B and C. In eq. (1.4), the same reflection m_a acts on the triangle in the position $(aC)(bB)(cA)$, and has then the effect of interchanging the corners A and B.

Thus, our definition of the active transformation m_a is that it interchanges the corners at b and at c, leaving the corner at a unmoved. Our definition of the passive transformation m_a is that it interchanges the coordinate labels b and c, leaving the label a in place. Either way, m_a is defined with reference to the three coordinate labels a, b, c, and acts as a permutation of them. It does not act as a permutation of the three corners, or corner labels, A, B, C, since it permutes the corners differently depending on which position the triangle is in before it is transformed.

Our basic philosophy is that the position labels a, b, c play a more fundamental role than the corner labels A, B, C. It simply makes no sense to speak of transformations of the triangle, either active or passive, unless we refer somehow to a coordinate system.

1.3 The group multiplication table

Table 1.1 is the complete group multiplication table, in the format

	g
f	fg

The fact that the product of two symmetry transformations sometimes does not commute, so that $gf \neq fg$, means that the multiplication table is not completely symmetric about the main diagonal. The non-commutativity is the most dramatic difference between group products in general and the ordinary multiplication of numbers.

	I	r_+	r_-	m_a	m_b	m_c
I	I	r_+	r_-	m_a	m_b	m_c
r_+	r_+	r_-	I	m_c	m_a	m_b
r_-	r_-	I	r_+	m_b	m_c	m_a
m_a	m_a	m_b	m_c	I	r_+	r_-
m_b	m_b	m_c	m_a	r_-	I	r_+
m_c	m_c	m_a	m_b	r_+	r_-	I

Table 1.1: Multiplication table of the symmetry group of the equilateral triangle.

The identity ("do nothing") transformation I is a *unit* for functional composition, that is,

$$If = fI = f \tag{1.7}$$

for any transformation f. Furthermore, every symmetry transformation f has a unique *inverse* f^{-1} such that

$$f^{-1}f = ff^{-1} = I \; . \tag{1.8}$$

For example, we see from Table 1.1 that

$$r_+^{-1} = r_- \; , \qquad m_a^{-1} = m_a \; . \tag{1.9}$$

A reflection of the triangle is its own inverse.

The crucial property of the group product is that it is *associative*, that

$$f(gh) = (fg)h \quad \text{for every} \quad f, g, h \; . \tag{1.10}$$

We may verify this property directly from Table 1.1, for example we have that

$$r_+(m_a m_b) = r_+ r_+ = r_- \; , \qquad (r_+ m_a) m_b = m_c m_b = r_- \; . \tag{1.11}$$

The direct verification from the multiplication table takes a lot of work, since there are $6^3 = 216$ cases to consider. Fortunately it is unnecessary, because associativity holds *by definition* for composition of transformations. The composite transformations $f(gh)$ and $(fg)h$ are necessarily the same, because according to the definition of composition they are computed in exactly the same way: act first with h, then with g and finally with f.

Names of the triangle group

A set of group elements and an associative group product, with the properties that there exists a unit element in the group and an inverse to every group element, is all that it takes to make a group. The symmetry group of the equilateral triangle is the simplest example of a group with a non-commutative group product. It goes under different names in different contexts. As a crystal symmetry group it is called C_{3v}, where C_3 is the *subgroup* consisting of the three rotations, and the subscript "v" means that reflections are included. It is also identical to the group of all possible permutations (reorderings) of three objects, which is called S_3, the *symmetric group* of *degree* three.

Obvious generalizations are C_{nv} and S_n for $n = 4, 5, 6, \ldots$. Thus, C_{nv} is the symmetry group of a regular polygon with n corners, and it has $2n$ elements: n rotations, including the identity transformation I, and n reflections. The rotations form the subgroup C_n, which is *cyclic*, meaning that all the rotations it contains can be generated by repeated applications of one single rotation, for example the rotation by an angle $2\pi/n$.

S_n is the group of all possible permutations of n objects. It has $n!$ elements. In general C_{nv} is a subgroup of S_n, because every symmetry transformation of a regular polygon can be looked upon as a permutation of the corner positions, when we take the passive point of view. The case $n = 3$ is special in the sense that C_{nv} is the whole of S_n. This is not true when $n > 3$, as is obvious just by a counting of elements: $2n = n! = 6$ for $n = 3$, but $2n < n!$ for $n > 3$.

1.4 Symmetry of the circle

The limit of C_{nv} when $n \to \infty$ is a group which we might call $C_{\infty v}$. It is the symmetry group of the circle, more often known as the *orthogonal group* in two dimensions, O(2) for short. Like every group C_{nv} it consists of rotations and reflections, and the rotations form a subgroup of O(2) which is called the *special* orthogonal group, SO(2). A rotation symmetry of the circle is given by a continuously variable rotation angle, and also a general reflection depends on a continuously variable angle. Thus SO(2) and O(2) are both *continuous* groups, or what amounts to the same, they are *Lie groups*.

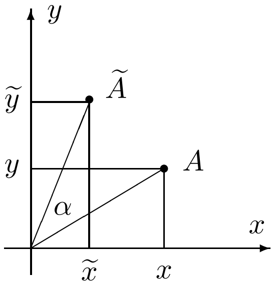

Figure 1.3: Rotation by an angle α.

A rotation by an angle α transforms a point A in the plane with coordinates (x, y) into a point $\widetilde{A}$ with coordinates $(\widetilde{x}, \widetilde{y})$, as shown in Figure 1.3. We find that

$$\begin{aligned} \widetilde{x} &= x\cos\alpha - y\sin\alpha \ , \\ \widetilde{y} &= x\sin\alpha + y\cos\alpha \ . \end{aligned} \tag{1.12}$$

Another way to write the transformtion is the matrix equation

$$\begin{pmatrix} \widetilde{x} \\ \widetilde{y} \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} . \tag{1.13}$$

The property of the transformation which allows us to write it in matrix form is that $\widetilde{x}$ and $\widetilde{y}$ are linear functions of x and y. Thus, a simple way to represent the rotation by α is by means of the 2×2 matrix

$$\mathbf{R}(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} . \tag{1.14}$$

Note that $\mathbf{R}(\alpha+\pi) = -\mathbf{R}(\alpha)$ and $\mathbf{R}(\alpha+2\pi) = -\mathbf{R}(\alpha+\pi) = \mathbf{R}(\alpha)$. The rotation by a zero angle is the identity *transformation* I, which is represented by the identity *matrix* $\mathbf{I}$,

$$\mathbf{R}(0) = \mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} . \tag{1.15}$$

Reflections, like rotations, are linear transformations. The general reflection in the plane, which leaves some direction $(\cos\alpha, \sin\alpha)$ invariant and inverts the orthogonal direction $(-\sin\alpha, \cos\alpha)$, is given by the matrix $\mathbf{M}(2\alpha)$, where

$$\mathbf{M}(\alpha) = \begin{pmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{pmatrix} . \tag{1.16}$$

Note that

$$\mathbf{M}(\alpha) = \mathbf{R}(\alpha)\,\mathbf{M}(0) = \mathbf{M}(0)\,\mathbf{R}(-\alpha) , \tag{1.17}$$

where

$$\mathbf{M}(0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \qquad (\mathbf{M}(0))^2 = \mathbf{I} . \tag{1.18}$$

With the matrix representation of linear transformations, the functional composition of linear transformations corresponds to matrix multiplication. From the above relations, and from the relation

$$\mathbf{R}(\alpha)\,\mathbf{R}(\beta) = \mathbf{R}(\alpha+\beta) , \tag{1.19}$$

we derive the multiplication table of the orthogonal group O(2), Table 1.2.

	$\mathbf{R}(\beta)$	$\mathbf{M}(\beta)$
$\mathbf{R}(\alpha)$	$\mathbf{R}(\alpha+\beta)$	$\mathbf{M}(\alpha+\beta)$
$\mathbf{M}(\alpha)$	$\mathbf{M}(\alpha-\beta)$	$\mathbf{R}(\alpha-\beta)$

Table 1.2: Multiplication table of O(2), the symmetry group of the circle.

We see that two rotations in the plane always commute,

$$\mathbf{R}(\alpha)\,\mathbf{R}(\beta) = \mathbf{R}(\alpha+\beta) = \mathbf{R}(\beta)\,\mathbf{R}(\alpha) . \tag{1.20}$$

A rotation and a reflection commute if the rotation is either $\mathbf{R}(0) = \mathbf{I}$ or $\mathbf{R}(\pi) = -\mathbf{I}$, but in no other case. A reflection $\mathbf{M}(\alpha)$ commutes with itself and with the orthogonal reflection $\mathbf{M}(\alpha+\pi) = -\mathbf{M}(\alpha)$, but with no other reflection. Thus, the rotation group SO(2) is commutative, whereas the complete orthogonal group O(2) is non-commutative.

The main property distinguishing rotations and reflections in the plane is the sign of the determinant,

$$\det \mathbf{R}(\alpha) = 1 , \qquad \det \mathbf{M}(\alpha) = -1 . \tag{1.21}$$

Note that $\mathrm{SO}(2) = \{\mathbf{R}(\alpha)\}$ is a *connected* subset of $\mathrm{O}(2)$, we may move continuously within $\mathrm{SO}(2)$ from any rotation matrix $\mathbf{R}(\alpha)$ to any other $\mathbf{R}(\beta)$: we simply change the rotation angle continuously from α to β.

The set of reflections, $\{\mathbf{M}(\alpha)\}$, is another connected subset of $\mathrm{O}(2)$. But there can be no continuous path within $\mathrm{O}(2)$ between a rotation and a reflection. To see this, note that the determinant $\det \mathbf{A}$ is a continuous function of the matrix $\mathbf{A}$. Since $\det \mathbf{A} = \pm 1$ for every $\mathbf{A} \in \mathrm{O}(2)$, it is impossible to change $\mathbf{A}$ continuously, within $\mathrm{O}(2)$, from a rotation with $\det \mathbf{A} = +1$ into a reflection with $\det \mathbf{A} = -1$.

This proves that $\mathrm{O}(2)$ consists of two separate *connection components*, the rotation group $\mathrm{SO}(2) = \{\mathbf{R}(\alpha)\}$ and the set of reflections $\{\mathbf{M}(\alpha)\}$.

1.5 Transformation groups

So far we have seen examples of transformation groups. We will now proceed to give a general definition, postponing until the next chapter the more abstract definition of a group by the group axioms.

It is possible to regard group theory as the theory of transformation groups, and no more than that. Most applications of group theory exemplify this view. It is even valid in a more general sense, because of the result known as Cayley's theorem, that every group may be regarded as a transformation group, acting by the group multiplication as a group of permutations of its own group elements. Cayley's theorem is a rather simple observation, but is considered important enough to be called a theorem.

From such a point of view the meaning of the group axioms is just that they summarize the basic properties of transformation groups. We will take a less extreme attitude here.

To prepare for the formal definition of a transformation group, we review some basic concepts.

Functions

Given two sets X and Y, either $X = Y$ or $X \neq Y$. We denote the number of elements in X and Y, respectively, by $|X|$ and $|Y|$, these may be finite or infinite numbers.

A *function* (or transformation, mapping) $f : X \to Y$ assigns to every element $x \in X$ a unique element $f(x) \in Y$. The set of all functions $f : X \to Y$ is denoted by Y^X, or else by $Y^{|X|}$. For example, if $X = \{1, 2, 3\}$, then $Y^X = Y^3$ is the set of all triplets (y_1, y_2, y_3) with each $y_i \in Y$.

Group transformations are not completely arbitrary functions, an important restriction is that they must be invertible. For example, a symmetry transformation of the equilateral triangle may be described as a function $f : X \to X$, where $X = \{a, b, c\}$ is the set of the three possible positions of the corners of the triangle. To be specific, the counterclockwise rotation r_+, interpreted as an active transformation, moves the corner at a to position b, it moves the corner at b to c and the corner at c to a, thus we may summarize its action as follows,

$$r_+(a) = b\ , \qquad r_+(b) = c\ , \qquad r_+(c) = a\ . \tag{1.22}$$

If we interpret it as a passive transformation, it relabels the corner position a as b, the position b as c, and the position c as a.

Table 1.3 shows how to represent all six symmetries of the equilateral triangle in this way, as functions. For each of the six functions in the table there is an inverse function in the same table, such that the action of the inverse function cancels the action of the function, and vice versa.

	a	b	c
I	a	b	c
r_+	b	c	a
r_-	c	a	b
m_a	a	c	b
m_b	c	b	a
m_c	b	a	c

Table 1.3: Symmetry transformations of the equilateral triangle as permutations of corner positions.

The *image* of a subset $A \subset X$ under the mapping $f : X \to Y$ is the subset

$$f(A) = \{ f(x) \mid x \in A \} \subset Y \; , \tag{1.23}$$

while the *inverse image* of a subset $B \subset Y$ is the subset

$$f^{-1}(B) = \{ x \in X \mid f(x) \in B \} \subset X \; . \tag{1.24}$$

Consider now the equation $f(x) = y$. We say that f is *onto* (or surjective) if for every $y \in Y$ there exists *at least* one solution $x \in X$, that is, if $f(X) = Y$.

f is *one to one* (or injective) if for every $y \in Y$ there exists *at most* one solution $x \in X$.

If $f : X \to Y$ is both onto and one to one, then by definition it is *invertible* (or bijective), and the *inverse* of f is the function $f^{-1} : Y \to X$ defined such that $x = f^{-1}(y)$ is the unique solution of the equation $f(x) = y$. In particular, an invertible function $f : X \to X$ is called a *permutation* of the elements of X.

Note that the notation $f^{-1}(B)$ for the inverse image of a set B does not presuppose that the inverse function f^{-1} exists. The inverse image of a set always exists as a set, which could be empty or contain one point or many points. The inverse function exists as a point mapping if and only if the inverse image of one point is always one point.

An important special case is when X and Y are finite sets with the same number of elements, $|X| = |Y|$. Then the two conditions that the function $f : X \to Y$ is onto or is one to one, are equivalent. In particular, when X is finite, then any function $f : X \to X$ is either onto and one to one, or it is neither of the two.

If the sets X and Y are finite, but not necessarily $|X| = |Y|$, then a necessary and sufficient condition for the existence of at least one function $f : X \to Y$ which is onto, is that $|X| \geq |Y|$. Similarly, there exists at least one function $f : X \to Y$ which is one to one, if and only if $|X| \leq |Y|$. Hence, there exists at least one invertible function $f : X \to Y$ if and only if the two sets X and Y have the same number of elements, $|X| = |Y|$.

If the two sets X and Y are both infinite, then the equation $|X| = |Y|$ has no meaning, unless we *define* it to mean that there exists at least one invertible function $f : X \to Y$.

Composition of functions

Given two functions $f : X \to Y$ and $g : Y \to Z$, then the *composite function* $g \circ f : X \to Z$, or more simply $gf : X \to Z$, is defined such that

$$gf(x) = g(f(x)) \quad \forall x \in X. \tag{1.25}$$

The central group axiom, the associative law, states that the two functions $h(gf)$ and $(hg)f$ are the same whenever $f : X \to Y$, $g : Y \to Z$ and $h : Z \to W$. It holds automatically for function composition, and in order to prove it, we simply write out the definitions,

$$\begin{aligned} [h(gf)](x) &= h([gf](x)) = h(g(f(x))) \ , \\ [(hg)f](x) &= [hg](f(x)) = h(g(f(x))) \ . \end{aligned} \tag{1.26}$$

Here we have been very explicit in our notation. Thus, for example, gf is a function mapping X into Z, $h(gf)$ is a function mapping X into W, and the values of these functions at the arbitrary point $x \in X$ are written as $[gf](x)$ and $[h(gf)](x)$, respectively.

The *identity transformation* on the set X is the function $I : X \to X$, or more explicitly $I_X : X \to X$, defined such that

$$I(x) = x \quad \forall x \in X \ . \tag{1.27}$$

I_Y is a *left unit* and I_X a *right unit* for every function $f : X \to Y$. That is, $I_Y f = f I_X = f$.

An interesting special case is when $f : X \to Y$ and $g : Y \to X$. In this case not only the composite function $gf : X \to X$ is well-defined, but also the function $fg : Y \to Y$. We see that if $X \neq Y$, then gf and fg are necessarily different functions, since gf maps X into X, while fg maps Y into Y.

If $f : X \to Y$ and $g : Y \to X$, as above, and if $gf = I_X$, then by definition g is a *left inverse* of f, whereas f is a *right inverse* of g. If $fg = I_Y$, then g is a right inverse of f and f a left inverse of g. We leave it as an exercise to verify the following statements.

– If f is invertible, then f^{-1} is both a left inverse and a right inverse of f.
– f has at least one left inverse if and only if it is one to one.
– f has at least one right inverse if and only if it is onto.
– f has both a left and a right inverse if and only if it is invertible.
– If g is a left inverse and h a right inverse of f, then $g = h = f^{-1}$.

Here is part of the proof of the last statement: if $gf = I_X$ and $fh = I_Y$, then

$$g = gI_Y = g(fh) = (gf)h = I_X h = h \ . \tag{1.28}$$

Note that the equation $gf = I_X$ does not in general imply that $fg = I_Y$, a left inverse is not always a right inverse. Or vice versa, a right inverse is not always a left inverse.

As an example, take $X = \{0\}$, $Y = \{1, 2\}$, $f(0) = 1$ and $g(1) = g(2) = 0$. Then $gf = I_X$, but $fg \neq I_Y$ because $fg(2) = 1$. Indeed f has no right inverse and g no left inverse. In this example $|X| = 1 < |Y| = 2$.

As a second example of a similar kind, take $X = Y = \{0, 1, 2, \ldots\}$, let $f(x) = x + 1$ for all $x \in X$, and let $g(0) = 0$, $g(x) = x - 1$ for $x > 0$. The function $f : X \to X$ is one to one but not onto, while $g : X \to X$ is onto but not one to one. This is possible because X is an infinite set.

In the following definition we require all functions to be invertible, so that we need not worry about the distinction between left and right inverses.

Definition 1.1 *A* "transformation group" *on some set X is a set F of invertible functions $f : X \to X$ which is closed under inversion and function composition.*

That is, $f^{-1} \in F$ and $fg \in F$ for all $f, g \in F$.

The "order" *of F is the number of elements, $|F|$.*

It follows immediately from the definition that every transformation group on X contains the identity transformation $I = I_X$, since $ff^{-1} = f^{-1}f = I$ for any arbitrary $f \in F$.

I is a (left and right) unit element of F, that is, $If = fI = f$ for every $f \in F$.

The largest possible transformation group on some set X is the set S_X (or $S_{|X|}$) of all permutations (invertible transformations) over X. Every transformation group on X is contained in S_X as a subgroup.

1.6 Symmetry and symmetry groups

Consider some set X, containing any kind of objects, and assume that we want to study some kind of structure on X. It may be any structure whatsoever, for example

- an algebraic structure, such as a group product, or the linear structure of a vector space;
- a metric (a table of distances between points in X);
- a topology (a collection of subsets of X called open sets);
- a classification of the members of X into two or more classes, for example as solutions or non-solutions of a set of equations;
- the equations might be, for example, polynomial equations, or the equations of motion of a physical system;
- any combination of structures.

Whatever the set X and the structure we consider, it is natural to call f a *symmetry transformation*, or simply a *symmetry*, if it is an invertible transformation of X onto itself preserving that structure. The set F of all such symmetry transformations is closed under composition and inversion, so that by Definition 1.1 it is a transformation group, which we call the *symmetry group* of the structure.

Symmetry transformations and symmetry groups are often given more specific names, depending on what kind of structure they preserve. For example,

- a *fixed point group* preserves one given element of X;
- a *space group*, or *crystal group*, preserves a crystal lattice;
- a *homeomorphism* preserves a topology;
- an *isomorphism* preserves an algebraic structure;
- a *linear group* preserves linear structure;
- an *isometry* preserves a metric;
- an *orthogonal* or *unitary group* preserves linear structure and metric.

It is rather obvious that symmetry groups are important, but it is sometimes surprising just how much information one can get from the symmetry group alone. A famous example from quantum mechanics is the understanding of the periodic system of the chemical elements, based on rotational symmetry together with the Pauli exclusion principle. In general, large symmetry groups, and perhaps less obviously, non-commutative ones, are the most useful.

1.7 The symmetric group S_n

Given two sets X and Y and an invertible function $f : X \to Y$. In this case the permutation groups S_X and S_Y are isomorphic, and the mapping $g \mapsto fgf^{-1}$ from S_X to S_Y is an isomorphism.

Definition 1.2 *The* "symmetric group" S_n *of* "degree" $n = 1, 2, \ldots$ *is the group of all permutations on the set* $\{1, 2, \ldots, n\}$*, or on any other set with* n *elements.*

The order of S_n is $|S_n| = n!$. To count the group elements, simply note that if we want to define an arbitrary permutation $f \in S_n$, we have at first n possibilities for defining $f(1)$, next $n-1$ possibilities for $f(2)$, and so on, until finally there is only one possibility left for defining $f(n)$. That makes altogether $n!$ possibilities.

Even and odd permutations

Definition 1.3 *The* "sign" *of a permutation* p *is*

$$\mathrm{sgn}(p) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} \frac{p(j) - p(i)}{j - i} = \prod_{i<j} \frac{p(j) - p(i)}{j - i} = \pm 1 \; . \tag{1.29}$$

p *is* "even" *if* $\mathrm{sgn}(p) = +1$ *and* "odd" *if* $\mathrm{sgn}(p) = -1$.

Whether the product runs over $i < j$ or $i > j$ is immaterial, the important point is that every pair $i \neq j$ of integers in the range $1, \ldots, n$ must be included once but not twice.

If $p, q \in S_n$, then

$$\begin{aligned}
\mathrm{sgn}(pq) &= \prod_{i<j} \frac{pq(j) - pq(i)}{j - i} = \prod_{i<j} \frac{p(q(j)) - p(q(i))}{q(j) - q(i)} \, \frac{q(j) - q(i)}{j - i} \\
&= \left(\prod_{k<l} \frac{p(l) - p(k)}{l - k} \right) \left(\prod_{i<j} \frac{q(j) - q(i)}{j - i} \right) = \mathrm{sgn}(p) \; \mathrm{sgn}(q) \; .
\end{aligned} \tag{1.30}$$

We define either $k = q(i)$, $l = q(j)$, if $q(i) < q(j)$, or $k = q(j)$, $l = q(i)$, if $q(j) < q(i)$. To summarize in words, the sign of the product is the product of signs,

$$\begin{array}{lclcl}
\text{even} \times \text{even} &=& \text{odd} \times \text{odd} &=& \text{even} \; , \\
\text{even} \times \text{odd} &=& \text{odd} \times \text{even} &=& \text{odd} \; .
\end{array} \tag{1.31}$$

Thus the mapping $\mathrm{sgn} : S_n \to \{1, -1\}$ is a group *homomorphism*: it preserves multiplication. It is not an *isomorphism*, since it is not one to one, except when $n = 1$ or $n = 2$.

Definition 1.4 *The* "alternating group" *of degree* n *is the subgroup* A_n *of even permutations in* S_n.

The alternating group A_n is exactly half of S_n. To see that there are the same number of odd and even permutations, just pick one arbitrary odd permutation p. Whenever q is an even permutation, pq is odd, and the correspondence $q \leftrightarrow pq$ is one to one between the even and odd permutations.

Cycles

Given a permutation $p \in S_n$ we can start with an arbitrary i_1 and define $i_2 = p(i_1)$, $i_3 = p(i_2)$, and so on. Since p acts on a finite set, sooner or later the sequence i_k must start repeating. Let i_L be the last number in the sequence which is not a repetition of an earlier number, then we must have $p(i_L) = i_1$, because otherwise p would not be invertible. If for example $p(i_L) = i_2$ there would be two different solutions $x = i_1$ and $x = i_L$ to the equation $p(x) = i_2$. This motivates the following notation.

Definition 1.5 *We write $s = (i_1, i_2, \ldots, i_L) = (i_1\, i_2\, \ldots\, i_L)$ and say that the permutation s is a* "cycle" *of length L (an* "L-cycle"*) if*

$$s(i_1) = i_2\ , \quad \ldots, \quad s(i_{L-1}) = i_L\ , \quad s(i_L) = i_1\ , \qquad s(j) = j \quad \forall j \neq i_1, \ldots, i_L\ . \tag{1.32}$$

A "transposition" *is a cycle of length 2.*

Two cycles $(i_1\, i_2\, \ldots\, i_L)$ and $(j_1\, j_2\, \ldots\, j_K)$ are "disjoint" *if*

$$\{i_1, i_2, \ldots, i_L\} \cap \{j_1, j_2 \ldots, j_K\} = \emptyset\ . \tag{1.33}$$

Inverting a cycle is easy, in fact,

$$(i_1\, i_2\, \ldots\, i_L)^{-1} = (i_L\, \ldots\, i_2\, i_1)\ . \tag{1.34}$$

The *order* of a cycle s is its length L. This means that s repeated L times is the identity transformation, $s^L = I$, whereas $s^n \neq I$ for $0 < n < L$.

Obviously, disjoint cycles commute. Every permutation can be factorized into a product of disjoint cycles, and the factors are unique although their ordering is not. Furthermore, every cycle and therefore every permutation can be written as a product of transpositions, in fact,

$$(i_1\, i_2\, \ldots\, i_L) = (i_1\, i_2)(i_2\, i_3)\ldots(i_{L-1}\, i_L)\ . \tag{1.35}$$

All transpositions are odd, hence a permutation is even or odd depending on whether it is the product of an even or odd number of transpositions. In particular, a cycle of length L is odd if L is even and even if L is odd.

In order to get used to the cycle notation, take as an example $p = (1\,2)$ and $q = (2\,3)$. That is, $p(1) = 2$, $p(2) = 1$, $p(i) = i$ otherwise, and similarly for q. Then $pq = (1\,2\,3)$, because

$$\begin{aligned} pq(1) &= p(q(1)) = p(1) = 2\ , \quad pq(2) = p(q(2)) = p(3) = 3\ , \\ pq(3) &= p(q(3)) = p(2) = 1\ , \quad pq(i) = p(q(i)) = p(i) = i \quad \text{otherwise}\ . \end{aligned} \tag{1.36}$$

The three examples

$$(1\,2)(1\,2) = I\ , \quad (1\,2)(2\,3) = (1\,2\,3)\ , \quad (1\,2)(3\,4) = (1\,2\,3)(2\,3\,4)\ , \tag{1.37}$$

are enough to show that every product of two transpositions and therefore every even permutation is either the identity, or it can be written as a product of 3-cycles. Obviously, since 3-cycles are even, an odd permutation can not be a product of 3-cycles only.

Theorem 1.6 *The transpositions (2-cycles) generate all of S_n (if $n \geq 2$), the 3-cycles generate the subgroup A_n of even permutations (if $n \geq 3$).*

1.8 Other ways to factorize permutations

The factorization of an arbitrary permutation $p \in S_n$ into disjoint cycles is unique. But there are other ways to factorize p into cycles. The following scheme uses a set of standard, but non-commuting, cycles, and also leads to a unique factorization.

Since S_{n-1} consists of all permutations of $1, 2, \ldots, n-1$, it may be regarded as a subgroup of S_n. Consider, as an example, S_7 as a subgroup of S_8, and consider the cyclic permutation

$$p = (2\,4\,8\,5) \in S_8 \;. \tag{1.38}$$

The main point here is that $p \notin S_7$, in fact, $p(8) = 5 \neq 8$. We observe that $p(4) = 8$, and we introduce a standard cycle

$$c = (8\,7\,6\,5\,4) \;, \tag{1.39}$$

with a decreasing sequence of consecutive numbers, and with $c(4) = 8$. If we now define

$$q = pc^{-1} = (2\,4\,8\,5)(4\,5\,6\,7\,8) = (2\,4)(5\,6\,7) \;, \tag{1.40}$$

then $q \in S_7$, because $q(8) = p(c^{-1}(8)) = p(4) = 8$. In this way we factorize $p \in S_8$ as $p = qc$, where $q \in S_7$, and where $c \in S_8$ is a standard cycle including the number 8.

Next, we observe that $q(6) = 7$, hence we introduce a standard cycle

$$d = (7\,6) \;, \tag{1.41}$$

with $d(6) = 7$, and we define

$$r = qd^{-1} = (2\,4)(5\,6\,7)(6\,7) = (2\,4)(5\,6) \;. \tag{1.42}$$

This gives the factorization $q = rd$, with $r \in S_6$. And so on, until we arrive at the factorization

$$p = c_2 c_3 c_4 c_5 c_6 c_7 c_8 \tag{1.43}$$

into cycles of a standard type, $c_8 = (8\,7\,6\,5\,4)$, $c_7 = (7\,6)$ $c_6 = (6\,5)$, $c_5 = I$, $c_4 = (4\,3\,2)$, $c_3 = (3\,2)$, $c_2 = I$. In general, $c_k \in S_k$. Our construction here shows that the factorization is unique.

In general, when we factorize an arbitrary permutation $p \in S_8$ in this way, c_8 may be one of the 8 different cycles I, $(8\,7)$, $(8\,7\,6)$, down to $(8\,7\,6\,5\,4\,3\,2\,1)$. There are 7 possibilities for c_7, and so on. Which proves again that the total number of different permutations in S_8 is $8! = 40\,320$.

It should be clear how to generalize this procedure from S_8 to S_n.

Later on, we will study linear representations of S_n and will then encounter the symmetrization operator

$$P = \sum_{p \in S_n} p \tag{1.44}$$

and the antisymmetrization operator

$$Q = \sum_{p \in S_n} \operatorname{sgn}(p)\, p \; . \tag{1.45}$$

These may be factorized as

$$P = P_2 P_3 \cdots P_n \;, \qquad Q = Q_2 Q_3 \cdots Q_n \;, \tag{1.46}$$

where, for example,

$$\begin{aligned} P_5 &= I + (5\,4) + (5\,4\,3) + (5\,4\,3\,2) + (5\,4\,3\,2\,1) = I + (5\,4)\, P_4 \;, \\ Q_5 &= I - (5\,4) + (5\,4\,3) - (5\,4\,3\,2) + (5\,4\,3\,2\,1) = I - (5\,4)\, Q_4 \;. \end{aligned} \tag{1.47}$$

A minimal set of transpositions generating S_n

The set of all transpositions is in general a redundant set of generators for the symmetric group S_n. To have a minimal set of generators, take for example the $n-1$ transpositions

$$T_i = (i, i+1) \;, \quad i = 1, 2, \ldots, n-1 \; . \tag{1.48}$$

They satisfy the following basic relations,

$$\begin{aligned} (T_i)^2 &= I \;, \\ T_i\, T_j &= T_j\, T_i \qquad \text{if} \qquad |i-j| \geq 2 \;, \\ T_{i+1}\, T_i\, T_{i+1} &= T_i\, T_{i+1}\, T_i \; . \end{aligned} \tag{1.49}$$

The last relation is proved simply by one example,

$$\begin{aligned} T_1\, T_2 &= (1\,2)\;(2\,3) = (1\,2\,3) \;, \\ (T_1\, T_2)\, T_1 &= (1\,2\,3)\;(1\,2) = (1\,3) \;, \\ T_2\, (T_1\, T_2) &= (2\,3)\;(1\,2\,3) = (1\,3) \; . \end{aligned} \tag{1.50}$$

Every product of a finite number of the generators T_i can be written in a standard form by a mechanical procedure using the following substitution rules,

$$\begin{aligned} (T_i)^2 &\to I \;, \\ T_j\, T_i &\to T_i\, T_j \qquad \text{if} \qquad j \geq i+2 \;, \\ (T_j\, T_{j-1}\, \ldots\, T_{i+1}\, T_i)\, T_j &\to T_{j-1}\, (T_j\, T_{j-1}\, \ldots\, T_{i+1}\, T_i) \qquad \text{if} \qquad j \geq i+1 \; . \end{aligned} \tag{1.51}$$

The longest string which is not changed by these rules, is

$$p = T_1\, (T_2\, T_1)\, (T_3\, T_2\, T_1)\, \ldots\, (T_{n-1}\, T_{n-2}\, \ldots\, T_2\, T_1) = c_2\, c_3\, \ldots\, c_n \;, \tag{1.52}$$

where each c_j is a cycle of length j,

$$c_j = T_{j-1}\, T_{j-2}\, \ldots\, T_2\, T_1 = (j, j-1, \ldots, 2, 1) \; . \tag{1.53}$$

The most general standard string, not changed by the rules, has the form

$$q = c_{2,i_2}\, c_{3,i_3}\, \ldots\, c_{n,i_n} \; . \tag{1.54}$$

We define $c_{j,i}$ with $i = 1, 2, \ldots, j$ as a cycle of length $j+1-i$. In particular, we define $c_{j,j} = I$, $c_{j,j-1} = T_{j-1} = (j, j-1)$, and

$$c_{j,i} = T_{j-1}\, T_{j-2}\, \ldots\, T_{i+1}\, T_i = (j, j-1, \ldots, i+1, i) \qquad \text{for} \qquad i < j \; . \tag{1.55}$$

We see that this factorization of an arbitrary $p \in S_n$ in terms of the transpositions $T_1, T_2, \ldots, T_{n-1}$ is the same as the cycle factorization introduced above.

1.9 Homogeneous spaces

We will now see that we may in principle learn all there is to know about the action of a transformation group by studying a special case called homogeneous spaces. Let F be a transformation group acting on a set X.

Definition 1.7 *The image under F of one point $x \in X$,*

$$F(x) = \{ f(x) \mid f \in F \} , \tag{1.56}$$

is called an "orbit"*.*

If for any two elements (or points) $x, y \in X$ there exists at least one $f \in F$ such that $y = f(x)$, then F is said to act "transitively" *on X, and X is called a* "homogeneous space"*.*

By definition, a homogeneous space X is an orbit, $X = F(x)$ for any $x \in X$. The converse is also true, that every orbit is a homogeneous space. For if y and z belong to the same orbit $F(x)$, say $y = f(x)$ and $z = g(x)$, then $z = h(y)$ when we choose $h = gf^{-1}$.

Note that we always have $x \in F(x)$, because the identity transformation I belongs to F, and $x = I(x)$.

It is also easy to see that for every $y \in F(x)$ we have $F(y) = F(x)$. In fact, let $y = f(x)$. We have $F(y) \subset F(x)$, because $z = h(y)$ implies that $z = g(x)$ with $g = hf$. And we have $F(x) \subset F(y)$, because $z = g(x)$ implies that $z = h(y)$ with $h = gf^{-1}$.

From these two facts, that $x \in F(x)$, and that $x \in F(y)$ implies $F(x) = F(y)$, it follows that every $x \in X$ belongs to exactly one orbit. In other words, X splits into disjoint orbits under the action of F. Since the orbits are homogeneous spaces, we have reduced the study of transformation groups and how they act to the study of homogeneous spaces.

In the standard mathematical terminology, the relation $y \in F(x)$ is an *equivalence relation* between x and y. It is

- *reflexive*: $x \in F(x)$ (because $x = I(x)$ and $I \in F$);
- *symmetric*: if $y \in F(x)$ then $x \in F(y)$ (because if $y = f(x)$ then $x = f^{-1}(y)$);
- and *transitive*: if $y \in F(x)$ and $z \in F(y)$, then $z \in F(x)$
  (because if $y = f(x)$ and $z = g(y)$, then $z = h(x)$ with $h = gf$).

An orbit $F(x) \subset X$ is an *equivalence class* with respect to the equivalence relation $y \in F(x)$ on X. The properties of an equivalence relation imply that every element $x \in X$ belongs to exactly one equivalence class.

We may summarize by a theorem. The part which we have not already proved will be proved below.

Theorem 1.8 *Let F be a transformation group on X. Then X is the union of (one or more) disjoint orbits, and every orbit is a homogeneous space.*

If X is a homogeneous space (one single orbit), then the number of elements in X is a divisor in the order of F. More precisely, $|F| = |X|\,|H|$, where H is the fixed point group of one arbitrary point in X.

Example

Take as an example the orthogonal group O(2), or the rotation group SO(2) $\subset$ O(2). An orbit either of O(2) or of SO(2) in the plane $\mathbf{R}^2$ is a circle of radius r,

$$C_r = \{ (x,y) \mid x^2 + y^2 = r^2 \} \ . \tag{1.57}$$

The whole plane is split into circles of different radii. A special case is the origin $(0,0)$, which is the circle C_0 of radius $r = 0$. Every circle is a homogeneous space of SO(2) and therefore of O(2): any point on the circle is rotated into any other point by some rotation.

1.10 Fixed point group and cosets

Assume in what follows that X is a homogeneous space of the transformation group F. Pick an arbitrary point $x \in X$. For every $y \in X$ introduce the (non-empty) set H_y of all transformations in F transforming x into y,

$$H_y = \{ f \in F \mid f(x) = y \} \ . \tag{1.58}$$

In particular, H_x is the *fixed point group* of x, we define

$$H = H_x = \{ h \in F \mid h(x) = x \} \ . \tag{1.59}$$

It is easy to verify that H is closed under composition and inversion, so that it is a transformation group, actually a subgroup of F.

Obviously, every $f \in F$ belongs to one and only one subset H_y. In other words, F is split in this way into disjoint subsets.

Every subset H_y is what we call a *left coset* of the fixed point group H, that is,

$$H_y = fH = \{ fh \mid h \in H\} \ , \tag{1.60}$$

where f is an arbitrary element in H_y. In order to prove this, assume that $f(x) = y$. We want to prove that $g(x) = y$ if and only if $g = fh$ with $h \in H$.

The proof of the "if" statement is very easy: if $g = fh$ with $h(x) = x$, then $g(x) = fh(x) = f(h(x)) = f(x) = y$. The proof of the "only if" statement is almost as easy: we have to solve the equation $g = fh$ for h, and the unique solution is $h = f^{-1}g$. Then $h \in H$, since $h(x) = f^{-1}g(x) = f^{-1}(g(x)) = f^{-1}(y) = x$.

When $H_y = fH$, the one to one correspondence $fh \leftrightarrow h$ between elements of H_y and of H proves that these two sets have the same number of elements. Thus, if we pick arbitrarily one transformation f_y from every coset H_y, every transformation $f \in F$ factorizes *in a unique way* as $f = f_y h$, where $y = f(x)$ and $h \in H$. This factorization counts the elements in F as $|F| = |X|\,|H|$, a result which is of course especially interesting when the group F is finite.

In summary, the correspondence between the point $y \in X$ and the left coset H_y is one to one and onto: to every point y corresponds a coset H_y, and to every coset fH corresponds a point $y = f(x)$. Thus we have actually proved an important theorem.

Definition 1.9 *The* "quotient space" *F/H is the set of all left cosets of H in F.*

Theorem 1.10 *A homogeneous space X of a transformation group F can be identified with the quotient space F/H, where H is the fixed point group of an arbitrary point $x \in X$.*

The order of F is $|F| = |X|\,|H|$.

Since the point $x \in X$ was chosen arbitrarily, it is natural to ask what happens if we choose a different point x'. Any two points x and x' in the homogeneous space X are connected by a transformation $g \in F$ such that $x' = g(x)$. It is easy to see that the two fixed point groups, H of x and H' of x', are then related in the following way,

$$H' = \{\, ghg^{-1} \mid h \in H \} \; . \tag{1.61}$$

In a compact notation we write $H' = gHg^{-1}$.

The mapping $h \mapsto h' = ghg^{-1}$ from H to H' is invertible, since there exists an inverse mapping $h' \mapsto h = g^{-1}hg$. This shows in particular that $|H| = |H'|$.

A mapping of this form is called a *conjugation.* It preserves group multiplication, that is, if $h = h_1h_2$ and $h' = ghg^{-1}$, $h'_1 = gh_1g^{-1}$, $h'_2 = gh_2g^{-1}$, then

$$h' = gh_1h_2g^{-1} = (gh_1g^{-1})(gh_2g^{-1}) = h'_1h'_2 \; . \tag{1.62}$$

An invertible mapping that preserves group multiplication is called an *isomorphism* of two groups. Thus the fixed point groups H and H' are isomorphic, they are the same group in the sense that they have the same multiplication table.

Example

A simple example is the symmetry group C_{3v} of the equilateral triangle, represented in Table 1.3 as a group of permutations of the corner positions a, b, c. The set $X = \{a, b, c\}$ is then a homogeneous space. The fixed point group of one corner position, say a, is $H = \{I, m_a\}$. Thus $|X| = 3$, $|H| = 2$, and $|C_{3v}| = 6 = |X|\,|H|$.

The fixed point group of the corner position b is $H' = \{I, m_b\}$. A transformation g with $g(a) = b$ is for example $g = m_c$, and it defines the following isomorphism from H to H',

$$I \mapsto m_c\, I\, {m_c}^{-1} = I \; , \qquad m_a \mapsto m_c\, m_a\, {m_c}^{-1} = m_b \; . \tag{1.63}$$

1.11 Permutation of boxes and permutation of balls

In the Sections 1.1 and 1.2 we introduced six symmetry transformations of the equilateral triangle, and observed that they may be understood as permutations of the coordinate labels a, b, c, but not as permutations of the corner labels A, B, C. It may be worthwhile to investigate the relation between these two kinds of permutations, partly as a matter of principle, and partly because the distinction may turn out to be useful.

Instead of the triangle as an example we consider three identical balls in three identical boxes, one ball in each box, as illustrated in Figure 1.4. It is possible to generalize to n identical balls in n identical boxes, but we will be specific and stay with the case $n = 3$. The system of boxes and balls is maximally symmetric, since any permutation of boxes and any permutation of balls is a symmetry transformation.

We label the boxes as a, b, c and the balls as A, B, C. Then we enumerate the six possible positions, or configurations, as follows,

$$\begin{aligned} &p_1 = (aA)(bB)(cC) \; , \qquad p_2 = (aB)(bC)(cA) \; , \qquad p_3 = (aC)(bA)(cB) \; , \\ &p_4 = (aA)(bC)(cB) \; , \qquad p_5 = (aC)(bB)(cA) \; , \qquad p_6 = (aB)(bA)(cC) \; . \end{aligned} \tag{1.64}$$

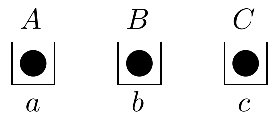

Figure 1.4: Three balls A, B, C in three boxes a, b, c. This is position p_1.

A position p is an invertible function $p : \{a, b, c\} \mapsto \{A, B, C\}$. The six different positions, or functions, are listed in Table 1.4.

For the six permutations of boxes we use the names $I, r_+, r_-, m_a, m_b, m_c$, as before. They are listed in Table 1.3, and again in Table 1.4. Since the boxes are identical, it makes little difference whether we actually move the boxes, or whether we just interchange their labels.

Similarly, there are six permutations of balls, which we call $I, R_+, R_-, M_A, M_B, M_C$. They are also listed in Table 1.4. Since the balls are identical, we may move them around, or we may choose to just interchange their labels.

The group of permutations of boxes is the symmetric group S_3, and the group of permutations of balls is also S_3.

	a	b	c
p_1	A	B	C
p_2	B	C	A
p_3	C	A	B
p_4	A	C	B
p_5	C	B	A
p_6	B	A	C

	a	b	c
I	a	b	c
r_+	b	c	a
r_-	c	a	b
m_a	a	c	b
m_b	c	b	a
m_c	b	a	c

	A	B	C
I	A	B	C
R_+	B	C	A
R_-	C	A	B
M_A	A	C	B
M_B	C	B	A
M_C	B	A	C

Table 1.4: The six configurations, permutations of boxes, and permutations of balls.

Now start with a position

$$p = (uU)(vV)(wW) , \tag{1.65}$$

where $\{u, v, w\} = \{a, b, c\}$, in any order, and $\{U, V, W\} = \{A, B, C\}$, in any order. A permutation f of boxes acts on p to give

$$\widetilde{p} = f(p) = (f(u)U)(f(v)V)(f(w)W) . \tag{1.66}$$

For example, with $p = p_1 = (aA)(bB)(cC)$ we have $r_+(p) = (bA)(cB)(aC)$, as shown in Figure 1.5. By eq. (1.66) we have that

$$\widetilde{p}(f(u)) = U = p(u) , \qquad \widetilde{p}(f(v)) = V = p(v) , \qquad \widetilde{p}(f(w)) = W = p(w) . \tag{1.67}$$

The relation $\widetilde{p}(f(x)) = p(x)$ for $x = a, b, c$ means that p is the composite function $\widetilde{p} \circ f$,

$$\widetilde{p} \circ f = p , \qquad \widetilde{p} = p \circ f^{-1} . \tag{1.68}$$

Alternatively, a permutation f of the boxes may be looked upon as a permutation of the contents of the boxes. In effect, it moves the ball lying in box x into box $f(x)$. The important point is that the action of f depends on how the *boxes* are labelled but not at all on how the *balls* are labelled.

A permutation F of the balls, on the contrary, depends on how the balls are labelled but not on how the boxes are labelled. It acts on $p = (uU)(vV)(wW)$ to give

$$\widetilde{p} = F(p) = (uF(U))(vF(V))(wF(W)) \; . \tag{1.69}$$

For example, $R_-(p_1) = (aC)(bA)(cB) = r_+(p_1)$, as is also shown in Figure 1.5. In general, for a permutation of the balls,

$$\widetilde{p}(x) = F(p(x)) \; , \tag{1.70}$$

which means that $\widetilde{p}$ is the composite function $F \circ p$,

$$\widetilde{p} = F \circ p \; . \tag{1.71}$$

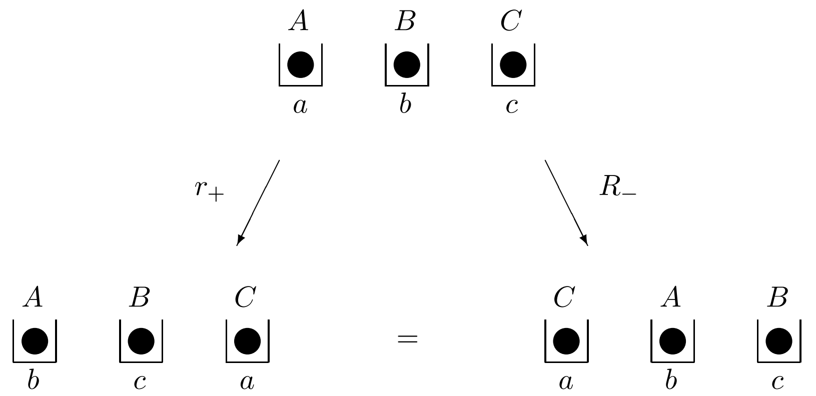

Figure 1.5: r_+ permutes the boxes, R_- permutes the balls.

Figure 1.5 shows that $r_+(p_1) = R_-(p_1)$. However, it is not true that r_+ and R_- are identical as transformations of the system. For example, with $p = p_6 = (aB)(bA)(cC)$ we have that $r_+(p) \neq R_-(p)$, as shown in Figure 1.6.

In conclusion, the symmetric group S_3 may permute either the boxes or the balls, and these are *two different actions of the same group on the same system.* The two actions commute, because composition of functions is associative. In fact,

$$F(f(p)) = F(p \circ f^{-1}) = F \circ (p \circ f^{-1}) = (F \circ p) \circ f^{-1} = f(F \circ p) = f(F(p)) \; . \tag{1.72}$$

Commutativity may also be proved directly, as follows. Let $p = (uU)(vV)(wW)$, then

$$\begin{aligned} F(f(p)) &= F(\,(f(u)U)\,(f(v)V)\,(f(w)W)\,) \\ &= (f(u)F(U))\,(f(v)F(V))\,(f(w)F(W)) \\ &= f(\,(uF(U))\,(vF(V))\,(wF(W))\,) \\ &= f(F(p)) \; . \end{aligned} \tag{1.73}$$

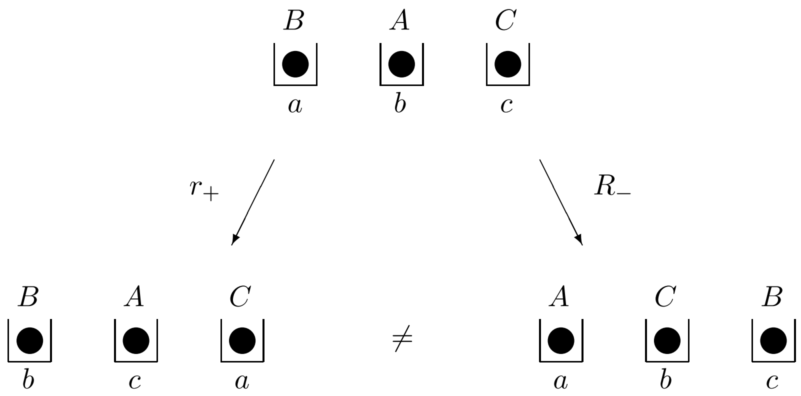

Figure 1.6: r_+ permutes the boxes, R_- permutes the balls.

We see that the full symmetry group of the system of boxes and balls is not just S_3, with $3! = 6$ group elements, it is actually the *direct product group* $S_3 \otimes S_3$, with $6^2 = 36$ elements. Two elements $F \in S_3$ and $f \in S_3$ define one element $\mathcal{F} = (F, f) \in S_3 \otimes S_3$, transforming a position p into

$$\mathcal{F}(p) = F \circ p \circ f^{-1} \, . \tag{1.74}$$

The group product $\mathcal{H} = \mathcal{F}\mathcal{G}$ of two elements $\mathcal{F} = (F, f)$ and $\mathcal{G} = (G, g)$ is $\mathcal{H} = (FG, fg)$, since

$$\begin{aligned} \mathcal{H}(p) &= \mathcal{F}(\mathcal{G}(p)) = F \circ (G \circ p \circ g^{-1}) \circ f^{-1} = (F \circ G) \circ p \circ (g^{-1} \circ f^{-1}) \\ &= (F \circ G) \circ p \circ (f \circ g)^{-1} \, . \end{aligned} \tag{1.75}$$

In the same way, we might have introduced 36 different symmetry transformations of the equilateral triangle, instead of just 6. However, that would certainly have introduced extra confusion, and it is not at all clear that it would have helped us to understand the symmetry of the triangle any better.

Problems

1. Given a composite function $h = gf$, where $f : X \to Y$, $g : Y \to Z$, $h : X \to Z$.

   Show that if f and g are both invertible, then h is invertible, and $h^{-1} = f^{-1}g^{-1}$.

   Show (by an example) that h may be invertible even if neither f nor g is invertible.

2. Let f and g be invertible functions, $f : X \to X$ and $g : Y \to Y$, and let p be a function, not necessarily invertible, $p : X \to Y$. Define a transformation h of the function p by the relation

   $$p \mapsto h(p) = g \circ p \circ f^{-1} = gpf^{-1} \; . \tag{1.76}$$

   Two transformations of this kind, $h_1(p) = g_1 p {f_1}^{-1}$ and $h_2(p) = g_2 p {f_2}^{-1}$, define a composite transformation $h = h_2 h_1$ such that $h(p) = h_2(h_1(p))$.

   Show that h is again of the form $h(p) = gpf^{-1}$, with $f = f_2 f_1$ and $g = g_2 g_1$.

3. In this exercise we look at the relation between the active and passive interpretations of the symmetry group of the equilateral triangle in Figure 1.1.

   The possible positions of the triangle are

   $$\begin{aligned} &1: \; (aA)(bB)(cC) \; , \quad 2: \; (aB)(bC)(cA) \; , \quad 3: \; (aC)(bA)(cB) \; , \\ &4: \; (aA)(bC)(cB) \; , \quad 5: \; (aC)(bB)(cA) \; , \quad 6: \; (aB)(bA)(cC) \; . \end{aligned} \tag{1.77}$$

   The active transformations introduced in eq. (1.1) and eq. (1.2) move the triangle corners A, B, C, and transform the triangle positions in the following way,

	1	2	3	4	5	6
I	1	2	3	4	5	6
r_+	3	1	2	6	4	5
r_-	2	3	1	5	6	4
m_a	4	6	5	1	3	2
m_b	5	4	6	2	1	3
m_c	6	5	4	3	2	1

   (1.78)

   A passive transformation does not move the triangle, but permutes the labels a, b, c of the coordinate system.

   Verify that the six different permutations of the labels a, b, c listed in Table 1.3 produce the six different transformations of triangle positions that are listed in the above table.

4. We may try to reproduce the transformation table in eq. (1.78) by permuting the triangle corners, labelled A, B, C, instead of permuting the corner positions a, b, c. If we define functions $I, R_+, R_-, M_A, M_B, M_C$ by the following table, then we may reproduce the first column of the transformation table,

$$
\begin{array}{c|ccc}
 & A & B & C \\
\hline
I & A & B & C \\
R_+ & B & C & A \\
R_- & C & A & B \\
M_A & A & C & B \\
M_B & C & B & A \\
M_C & B & A & C
\end{array}
\quad \Longrightarrow \quad
\begin{array}{c|cccccc}
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\hline
I & 1 & 2 & 3 & 4 & 5 & 6 \\
R_- & 3 & & & & & \\
R_+ & 2 & & & & & \\
M_A & 4 & & & & & \\
M_B & 5 & & & & & \\
M_C & 6 & & & & &
\end{array}
\tag{1.79}
$$

Complete the last table, and compare it with eq. (1.78). Conclusion?

5. If we start with $n-1$ group elements $T_1, T_2, \ldots, T_{n-1}$ (and the identity I), satisfying the basic relations listed in eq. (1.49), and if we include all products of these generators, we obtain the whole symmetric group S_n.

   We defined these generators in eq. (1.48) as the transpositions $T_i = (i, i+1)$. However, we may ignore the definition in eq. (1.48) and use only the relations in eq. (1.49).

   Take the special case $n = 3$, and use only the relations in eq. (1.49) to show that the group S_3 contains exactly six different group elements that are products of T_1 and T_2. Compute the multiplication table of of these six group elements.

Chapter 2

Group theory

This chapter is a summary of basic definitions and some central concepts and theorems of group theory. Every group may be regarded as a transformation group. For example, it acts on itself in two ways, by left and right translation.
Some of the central concepts are: Conjugation classes; subgroups; left and right cosets; normal (or invariant) subgroups; factor groups; homomorphism and isomorphism; direct and semidirect products of groups.

2.1 The group axioms

The essence of the concept of symmetry transformations is distilled into the following set of axioms.

Definition 2.1 *A* "group" *G is a set G with a* "group product" *such that*

(i) any two elements $a, b \in G$ have a unique product $ab \in G$;
(ii) $a(bc) = (ab)c \quad \forall a, b, c \in G$ (the associative law);
(iii) there exists at least one "left unit" *$e \in G$ such that $ea = a \quad \forall a \in G$;*
(iv) every $a \in G$ has at least one "left inverse" *$a^{-1} \in G$ such that $a^{-1}a = e$.*

In a "commutative" *or* "Abelian" *group we have in addition*

(v) $ab = ba \quad \forall a, b \in G$ (the commutative law).

The "order" *of the group G, denoted by $|G|$, is the number of elements in the set G (either finite or infinite).*

We need not be too careful in distinguishing between the group G and the set G of group elements, since the group product is usually unambiguous. In an Abelian (commutative) group one often writes the group "product" as $a + b$ instead of ab, and then 0 is the unit element, $-a$ the inverse of a and $na = n \times a$ the n-th power of a.

Axiom (iii) guarantees that every group has at least one element, the unit element e. The minimal group is the one that consists of e alone, it is the only group of order 1.

A somewhat less trivial example of a group is given in Table 2.1. It is a recommended exercise to check the group axioms in this example, and also to try to construct other groups

	e	a	b	c
e	e	a	b	c
a	a	e	c	b
b	b	c	e	a
c	c	b	a	e

Table 2.1: Multiplication table of the fourth order Abelian group called the *four-group*. It is called D_2 when it appears as a crystal point group.

of order two, three and four. This particular group has, for example, a physical application as the crystal point group D_2.

Given a group element a we define

$$a^0 = e, \qquad a^{n+1} = a^n a \quad \text{for} \quad n = 0, 1, 2, \dots \; . \tag{2.1}$$

Associativity means that there is no need for parentheses when we write products of more than two group elements. Thus,

$$\begin{aligned} a^3 &= a^2 a = (aa)a = a(aa) = aa^2 = aaa \; , \\ a^{m+n} &= a^m a^n = a^n a^m \quad \forall m, n = 0, 1, 2, \dots \; . \end{aligned} \tag{2.2}$$

The following result is so useful that we state it as a theorem.

Theorem 2.2 *The unit e is idempotent, $e^2 = e$, and is uniquely defined by this property.*

In fact, $ee = e$ by definition, and if $f^2 = f$, then

$$f = ef = (f^{-1}f)f = f^{-1}(ff) = f^{-1}f = e \; . \tag{2.3}$$

It follows that the left inverse a^{-1} is also a right inverse, $aa^{-1} = e$, since

$$(aa^{-1})^2 = (aa^{-1})(aa^{-1}) = a(a^{-1}(aa^{-1})) = a((a^{-1}a)a^{-1}) = a(ea^{-1}) = aa^{-1} \; . \tag{2.4}$$

The left unit is also a right unit, since

$$ae = a(a^{-1}a) = (aa^{-1})a = ea = a \; . \tag{2.5}$$

And the inverse a^{-1} is uniquely defined, for if $ba = e$, then

$$b = be = b(aa^{-1}) = (ba)a^{-1} = ea^{-1} = a^{-1} \; . \tag{2.6}$$

Or if $ab = e$, then

$$b = eb = (a^{-1}a)b = a^{-1}(ab) = a^{-1}e = a^{-1} \; . \tag{2.7}$$

In our definition of a group, instead of postulating the existence of a left unit and left inverses, we could just as well have postulated the existence of a right unit and right inverses.

Would it also be sufficient to postulate left unit and right inverses?

Uniqueness of the inverse implies that $(a^{-1})^{-1} = a$ and $(ab)^{-1} = b^{-1}a^{-1}$, since $aa^{-1} = e$ and

$$(b^{-1}a^{-1})(ab) = b^{-1}(a^{-1}a)b = b^{-1}b = e \; . \tag{2.8}$$

Taking $b = a$, we get that $(a^2)^{-1} = (a^{-1})^2$. In general, $(a^n)^{-1} = (a^{-1})^n$ for $n = 0, 1, 2, \ldots$, which motivates the definition

$$a^{-n} = (a^n)^{-1} \quad \text{when} \quad n > 0 \; . \tag{2.9}$$

We leave it as an exercise to prove the following relations for any two integers m and n, positive, negative or zero,

$$a^{m+n} = a^m a^n \; , \qquad a^{mn} = (a^m)^n \; . \tag{2.10}$$

Cyclic groups

Different powers of a are not necessarily different elements of G, it might well happen that $a^m = a^n$ for some $m > n$. In that case,

$$a^{m-n} = a^m a^{-n} = a^n a^{-n} = e \; . \tag{2.11}$$

Definition 2.3 *The* "order" *of the element $a \in G$ is the smallest integer $N > 0$ such that $a^N = e$. If no such N exists, a is of infinite order.*

G is a "cyclic group" *if it has a* "cyclic element" *(or* "generator"*) g such that all group elements are powers of g, positive, negative or zero. That is, if*

$$G = \{ \, g^n \mid n = 0, \pm 1, \pm 2, \ldots \, \} \; . \tag{2.12}$$

Obviously, all elements of a finite group are of finite order, and the order of a cyclic group is the same as the order of any one of its generators. Every cyclic group is Abelian, since $g^m g^n = g^n g^m = g^{m+n}$.

We may use two different notations for cyclic groups. There is a unique cyclic group of infinite order,

$$C_\infty = \{ \, g^n \mid n = 0, \pm 1, \pm 2, \ldots \, \} \; , \tag{2.13}$$

with $g^m \neq g^n$ when $m \neq n$. It has two generators, g and g^{-1}. It can be identified with the addition group of integers,

$$\mathbb{Z} = \{ \, 0, \pm 1, \pm 2, \ldots \, \} \; . \tag{2.14}$$

There is also a unique cyclic group of finite order N,

$$C_N = \{ \, g \, , \, g^2 \, , \, \ldots \, , \, g^{N-1} = g^{-1} \, , \, g^N = g^0 = e \, \} \; . \tag{2.15}$$

Generators are g and g^{-1}, and in fact any g^k such that k and N are relative prime (have no prime factor in common). It can be identified with the addition group of integers *modulo* N,

$$\mathbb{Z}_N = \{ \, 0 \, , \, 1 \, , \, 2 \, , \, \ldots \, , \, N-1 \equiv -1 \, , \, N \equiv 0 \, \} \; . \tag{2.16}$$

2.2 Left and right translation, conjugation

Theorem 2.4 (Cayley) *Every group G is a transformation group.*

To prove the theorem, simply note that the group element g transforms (or *translates*) an element $x \in G$ into gx. That is, we might think of g as a function $g : G \to G$ such that $g(x) = gx$. We will introduce a slightly different notation.

Definition 2.5 *Every $g \in G$ defines a* "left translation" *$L_g : G \to G$ and a* "right translation" *$R_g : G \to G$, such that*

$$L_g(x) = gx \;, \qquad R_g(x) = xg^{-1} \qquad \forall x \in G \;. \tag{2.17}$$

The group product corresponds to composition of left translations, $L_{gh} = L_g L_h$, and of right translations, $R_{gh} = R_g R_h$, as is easily verified,

$$\begin{aligned} L_{gh}(x) &= ghx = L_g(hx) = L_g(L_h(x)) = L_g L_h(x) \;, \\ R_{gh}(x) &= x(gh)^{-1} = xh^{-1}g^{-1} = R_g(xh^{-1}) = R_g(R_h(x)) = R_g R_h(x) \;. \end{aligned} \tag{2.18}$$

In an Abelian group, a right translation is simply the inverse of a left translation,

$$R_g(x) = xg^{-1} = g^{-1}x = L_g^{-1}(x) \;. \tag{2.19}$$

Clearly, different group elements define different left translations, $L_g = L_h$ if and only if $g = h$. Hence the group G can be identified with the transformation group consisting of all left translations. Or equivalently, if one prefers, with the group of right translations.

Definition 2.6 *A simultaneous left and right translation by g, $C_g = L_g R_g = R_g L_g$, which gives $C_g(x) = gxg^{-1}$, is a* "conjugation"*, and gxg^{-1} is the* "conjugate" *of x by g.*

The "conjugation class" *(or simply* "class"*) of $x \in G$ is the set of all elements that are conjugate to x,*

$$\mathrm{Cl}(x) = \{\, gxg^{-1} \mid g \in G \,\} \;. \tag{2.20}$$

The "centralizer" *(or* "commutant"*) of x is the set of all elements commuting with x,*

$$Z(x) = \{\, g \in G \mid gx = xg \,\} \;. \tag{2.21}$$

In every group the set $\{e\}$ containing the unit element alone is a conjugation class. A group is Abelian if and only if every conjugation class contains one single element.

The centralizer $Z(x)$ is a subgroup of G, it is the fixed point group of x when G acts on itself by conjugation. It is often denoted by $C_G(x)$, otherwise $C_G(x)$ would be a natural notation for the conjugation class $\mathrm{Cl}(x)$.

By Definition 1.7, the conjugation classes are the orbits when G acts on itself by conjugation. Therefore we have the following theorem, as a special case of Theorem 1.8.

Theorem 2.7 *Every element of G is a member of exactly one conjugation class. In other words, G is a union of disjoint conjugation classes.*

The number of elements in a conjugation class $\mathrm{Cl}(x)$ is a divisor of the order of the group,

$$|G| = |\,\mathrm{Cl}(x)|\;|Z(x)| \;. \tag{2.22}$$

2.3 Example: conjugation classes in S_n

If $s = (i_1\, i_2\, \ldots\, i_L)$ is a cycle and q is any permutation, then the conjugate of s by q, qsq^{-1}, is a cycle of the same length as s,

$$qsq^{-1} = (q(i_1)\, q(i_2)\, \ldots\, q(i_L)) \; . \tag{2.23}$$

In fact, qsq^{-1} maps $q(i_1)$ into $q(s(q^{-1}(q(i_1)))) = q(s(i_1)) = q(i_2)$, and so on.

Conversely, this formula shows that any two cycles r and s of the same length are conjugates of each other, $r = qsq^{-1}$ for some q. For if $r = (j_1\, j_2\, \ldots\, j_L)$, we can always find a q such that $q(i_k) = j_k$ for $k = 1, 2, \ldots, L$. In particular, $s^{-1} = (i_L\, \ldots\, i_2\, i_1)$ is a conjugate of $s = (i_1\, i_2\, \ldots\, i_L)$.

Theorem 2.8 *Two permutations in S_n are conjugate if and only if they have the same cycle structure, in the sense that the lengths of cycles correspond when both are factored into disjoint cycles. In particular, every permutation is conjugate to its inverse.*

The standard factorization of a permutation into commuting cycles has the form $p = s_1 s_2 \ldots s_M$, where each s_i is a cycle of length L_i, and where $L_1 \geq L_2 \geq \ldots \geq L_M \geq 1$. We count here also cycles of length 1. Every number $1, 2, \ldots, n$ belongs to one and only one of the cycles s_i, and therefore the lengths L_i form a partition of n, $L_1 + L_2 + \ldots + L_M = n$. The sign of p is simply

$$\begin{aligned} \operatorname{sgn}(p) &= \operatorname{sgn}(s_1) \operatorname{sgn}(s_2) \ldots \operatorname{sgn}(s_M) \\ &= (-1)^{L_1 - 1} (-1)^{L_2 - 1} \ldots (-1)^{L_M - 1} = (-1)^{n-M} \; , \end{aligned} \tag{2.24}$$

where M is the number of cycles of p.

Definition 2.9 *A* "partition of n" *is a decomposition of n as a sum of positive integers.*

Since addition is commutative, the order of the numbers in a partition is irrelevant. A useful convention is to write them in a decreasing (or non-increasing) sequence.

Theorem 2.10 *There is a one to one correspondence between conjugation classes in S_n and partitions of n.*

By Theorem 1.8, the number of elements in the conjugation class containing a given permutation $p \in S_n$ is

$$N_p = \frac{|S_n|}{|Z(p)|} = \frac{n!}{|Z(p)|} \; , \tag{2.25}$$

where the centralizer $Z(p) \subset S_n$ is the fixed point group of p,

$$Z(p) = \{\, q \in S_n \mid qpq^{-1} = p \,\} \; . \tag{2.26}$$

Assume that p, when factored into disjoint cycles, contains $\nu_L = 0, 1, 2, \ldots$ cycles of length L for $L = 1, 2, 3, \ldots$. Thus,

$$\sum_{L=1}^{\infty} \nu_L\, L = n \; . \tag{2.27}$$

This is another way to specify a partition of n. Then the order of the fixed point group $Z(p)$ is

$$|Z(p)| = \prod_{L=1}^{\infty} \nu_L!\, L^{\nu_L} \;, \tag{2.28}$$

and the number of elements in the conjugation class of p is

$$N_p = N(\nu_1, \nu_2, \ldots) = \frac{n!}{\displaystyle\prod_{L=1}^{\infty} \nu_L!\, L^{\nu_L}} \;. \tag{2.29}$$

In eq. (2.28) the reason for the factor $\nu_L!$ is that every permutation $q \in Z(p)$ may permute freely the ν_L cycles of length L. For every cycle of length L there is a factor L, because q may permute *cyclically* the L numbers belonging to that cycle.

Take just two specific examples. The conjugation class of the identity transformation I has $\nu_1 = n$ and $\nu_2 = \nu_3 = \ldots = 0$. Therfore the number of elements is

$$N_I = N(n) = \frac{n!}{n!\, 1^n} = 1 \;. \tag{2.30}$$

The conjugation class of transpositions has $\nu_1 = n - 2$, $\nu_2 = 1$ and $\nu_3 = \nu_4 = \ldots = 0$. Therefore the number of elements is

$$N(n-2, 1) = \frac{n!}{((n-2)!\, 1^{n-2})(1!\, 2^1)} = \frac{n(n-1)}{2} = \binom{n}{2} \;. \tag{2.31}$$

This is the number of ways to pick two numbers out of n.

2.4 Subgroups

Definition 2.11 *A subset $H \subset G$ is a* "subgroup" *of the group G if it is a group under the group product of G.*

A subgroup H of G is "normal" *(or* "invariant"*) if it is invariant under conjugation, that is, if $ghg^{-1} \in H$ for all $h \in H$ and $g \in G$.*

The group G is "simple" *if it has no normal subgroup besides the two trivial cases $H = G$ and $H = \{e\}$.*

In an Abelian group $ghg^{-1} = h$, so that every subgroup is normal.

The unit element of the subgroup H must be idempotent, and hence identical to the unit e of G, by Theorem 2.2. Therefore the inverse a^{-1} of a must also be the same in the two groups G and H. It is easy to prove

Theorem 2.12 *A necessary and sufficient condition for $H \subset G$ to be a subgroup is that it is closed under multiplication and inversion, that is,*

$$ab \in H \quad \textit{and} \quad a^{-1} \in H \quad \forall a, b \in H \;. \tag{2.32}$$

If H is finite (even if G is infinite), then closure under multiplication is sufficient.

In fact, when H is finite and closed under multiplication, every element $a \in H$ must be of finite order $N \geq 1$, hence $e = a^N \in H$ and $a^{-1} = a^{N-1} = a^{2N-1} \in H$.

The trivial normal subgroups G and $\{e\}$ have already been mentioned. The next example of a subgroup is the cyclic group generated by one element $a \in G$,

$$\{ a^n \mid n = 0, \pm 1, \pm 2, \dots \} . \tag{2.33}$$

The order of the element a is just the order of the subgroup it generates. If a is of finite order N, the subgroup is

$$\{a, a^2, \dots, a^{N-1} = a^{-1}, a^N = e\} . \tag{2.34}$$

More generally, every subset $A \subset G$ generates a subgroup of G. This is the smallest subgroup containing A, and it consists of the elements of A and their inverses together with all (finite) products of these. A subset A invariant under conjugation, such that $gAg^{-1} = A$ for all $g \in G$, generates a subgroup which is normal, since

$$ga^{-1}g^{-1} = (gag^{-1})^{-1} \qquad \text{and} \qquad g(ab)g^{-1} = (gag^{-1})(gbg^{-1}) . \tag{2.35}$$

Definition 2.13 *The* "centre" *(or* "commutant"*) of a group G is the normal subgroup*

$$Z(G) = \{ g \in G \mid gh = hg \quad \forall h \in G \} . \tag{2.36}$$

More generally, if $S \subset G$, the "centralizer" *(or* "commutant"*) of S is the subgroup*

$$Z(S) = \{ g \in G \mid gs = sg \quad \forall s \in S \} . \tag{2.37}$$

The "normalizer" *of S is the subgroup*

$$N(S) = \{ g \in G \mid gsg^{-1} \in S \quad \forall s \in S \} . \tag{2.38}$$

Obviously, we always have $Z(G) \subset Z(S) \subset N(S)$.

Definition 2.14 *The* "commutator" *of two elements $a, b \in G$ is $aba^{-1}b^{-1}$. The set of all commutators in G generates the* "commutator subgroup" *$Q(G)$.*

The commutator subgroup $Q(G)$ is normal, because the set of commutators is invariant under conjugation. In fact, let $c = aba^{-1}b^{-1}$ be a commutator and let $g \in G$. Then gcg^{-1} is a commutator,

$$gcg^{-1} = (gag^{-1})(gbg^{-1})(gag^{-1})^{-1}(gbg^{-1})^{-1} . \tag{2.39}$$

2.5 Cosets

Definition 2.15 *Let H be a subgroup of the group G, and let $g \in G$. The set*

$$gH = \{ gh \mid h \in H \} , \tag{2.40}$$

which is H left-translated by g, is a "left coset" *of H. The right-translated copy of H,*

$$Hg = \{ hg \mid h \in H \} , \tag{2.41}$$

is a "right coset"*. The* "coset space" *(or* "quotient space"*) G/H is the set of all left cosets. Similarly, $H\backslash G$ is the set of all right cosets.*

In analogy to the definitions of gH and Hg we may define

$$gHg^{-1} = \{ ghg^{-1} \mid h \in H \} . \qquad (2.42)$$

When H is a subgroup, so is gHg^{-1}, and H is normal if and only if $gHg^{-1} = H$ for every $g \in G$. The equation $gHg^{-1} = H$ can be right multiplied by g to give $gH = Hg$. Thus, for a normal subgroup the distinction between left and right cosets is superfluous, and this is seen to be precisely the defining property of a normal subgroup.

Theorem 2.16 *Let $a, b \in G$. Then $a \in aH$. And $a \in bH$ implies that $aH = bH$. Therefore every element of G is a member of exactly one left coset of H.*

The proof is trivial. In fact, $a \in aH$ because $a = ae$ and $e \in H$. And if $a \in bH$, then $a = bh$ for some $h \in H$, hence $aH = bhH = bH$. There is of course a similar theorem for right cosets.

There is a one to one correspondence, $h = g^{-1}(gh) \leftrightarrow gh$, between H and gH, so that they contain the same number of elements, $|H| = |gH|$. Similarly, $|H| = |Hg|$. This means that the subgroup H of G partitions G into disjoint left (or equivalently right) cosets with $|H|$ elements in each. Which proves

Theorem 2.17 (Lagrange) *The order $|G|$ of a finite group G is a multiple of the order $|H|$ of any subgroup H. More precisely,*

$$|G| = |H|\,|G/H| = |H|\,|H\backslash G| . \qquad (2.43)$$

In particular, $|G|$ is a multiple of the order of any single element in G.

The last point follows because every group element generates a cyclic subgroup.

This theorem completes the theory of finite groups of order p where p is a prime number. It implies that in such a group every element must be of order either 1 or p. Since e is the only element of order 1, all other elements are cyclic, and the group is cyclic.

Both left and right translations act transitively on a group G, so every group is a homogeneous space. The left translations on G are also transformations of the quotient space G/H of left cosets, when H is any subgroup of G. A coset aH is translated by $g \in G$ into

$$g(aH) = (ga)H = gaH . \qquad (2.44)$$

The action of G on G/H is also transitive, so every coset space G/H is a homogeneous space. In fact, this is the most general kind of homogeneous space, by Theorem 1.10.

Obviously, since left and right cosets are in a one to one correspondence, a homogeneous space X identified with the left coset space G/H can be identified in a similar manner with the right coset space $H\backslash G$.

2.6 Example: normal subgroups of S_n

As an example, let us determine all the normal subgroups of the symmetric group S_n.

For any two permutations $p, q \in S_n$ we have that

$$\begin{aligned} \operatorname{sgn}(qpq^{-1}) &= \operatorname{sgn}(q)\,\operatorname{sgn}(p)\,\operatorname{sgn}(q^{-1}) = \operatorname{sgn}(p)\,\operatorname{sgn}(q)\,\operatorname{sgn}(q^{-1}) \\ &= \operatorname{sgn}(p)\,\operatorname{sgn}(qq^{-1}) = \operatorname{sgn}(p) . \end{aligned} \qquad (2.45)$$

Thus the conjugate of every even permutation is always even, so that the alternating group A_n, consisting of all even permutations, is a normal subgroup of S_n. For $n > 1$ it is exactly half of S_n. We want to prove the following remarkable result.

Theorem 2.18 *With one single exception in the case $n = 4$, A_n is the only non-trivial normal subgroup of S_n (non-trivial if $n > 2$).*

Any normal subgroup H must contain the entire conjugation class of every element it contains. For example, if it contains one transposition, it contains every transposition and hence all of S_n. Therefore the interesting case is when H does not contain any transposition.

Assume that H contains at least one element $p \neq I$. Factorize p into disjoint cycles, $p = s_1 s_2 \ldots s_M$, of lengths $L_1 \geq L_2 \geq \ldots \geq L_M$. Just to be specific we will assume that $s_1 = (1\,2\,\ldots\,L_1)$, this means that we may have to replace p by one of its conjugates. There are now two cases, either $L_1 > 2$ or $L_1 = 2$.

If $L_1 > 2$, take $q = r s_2^{-1} \ldots s_M^{-1}$ with $r = (1, 2, L_1, L_1 - 1, \ldots, 3)$. Then q is a conjugate of p, so that $q \in H$ and $pq = s_1 r = (1\,3\,2) \in H$. Since H contains one 3-cycle, by Theorem 1.6 it contains all of A_n, hence either $H = A_n$ or $H = S_n$. To prove that $H = A_n$ or $H = S_n$ when H is a subgroup of S_n containing A_n, just count elements: since $n!/2 \leq |H| \leq n!$, and $|H|$ is a divisor of $n!$, the only possibilities are $|H| = n!/2$ or $|H| = n!$.

If $L_1 = 2$, then every L_i is either 2 or 1, and each of the cycles s_i is its own inverse. The case when p is a transposition, $L_2 = 1$, has been dealt with already, so we assume that $L_2 = 2$. We have already assumed that $s_1 = (1\,2)$, and we can further assume that $s_2 = (3\,4)$. Take now $q = (1\,3)(2\,4)s_3 \ldots s_M$. Then $q \in H$ and $(1\,4)(2\,3) = pq \in H$. If $n \geq 5$, we can conclude further that

$$(1\,2\,3) = (1\,2)(4\,5)(2\,3)(4\,5) \in H \;, \tag{2.46}$$

and that $H = A_n$ or $H = S_n$, as before.

There is only one very special case left. This is the normal subgroup of S_4 consisting of the 4 permutations I, $(1\,2)(3\,4)$, $(1\,3)(2\,4)$ and $(1\,4)(2\,3)$.

2.7 Factor groups

There is a natural way to define the product of two subsets $A, B \subset G$,

$$AB = \{\, ab \mid a \in A,\ b \in B \,\} \;. \tag{2.47}$$

A subgroup H is also a coset, $H = hH$ for every $h \in H$, and is idempotent, $HH = H$. The weaker relation $HH \subset H$ means simply that H is closed under multiplication. If in addition H is normal, then left and right cosets are identical, and the product of two cosets aH and $bH = Hb$ is again a coset,

$$(aH)(bH) = a(Hb)H = a(bH)H = (ab)(HH) = abH \;. \tag{2.48}$$

This multiplication rule for cosets implies that $H = eH$ acts as a left unit,

$$H(aH) = (eH)(aH) = eaH = aH \;, \tag{2.49}$$

and that $a^{-1}H$ is an inverse for aH,

$$(a^{-1}H)(aH) = a^{-1}aH = eH = H \;. \tag{2.50}$$

Theorem 2.19 *The coset space G/H is a group if H is a normal subgroup of G.*

In this case, G/H is called a *factor group*, or a *quotient group*.

2.8 Homomorphisms and isomorphisms

Definition 2.20 *Let G and H be two groups. A function $\rho : G \to H$ is a* "homomorphism" *if it preserves group multiplication, that is if*

$$\rho(ab) = \rho(a)\rho(b) \quad \forall a, b \in G \; . \tag{2.51}$$

If in addition ρ is an invertible function, then it is an "isomorphism"*, and the two groups are* "isomorphic"*, $G \simeq H$.*

Isomorphic groups are identical as abstract groups, and if they are different it is because they appear in different contexts, for example as different subgroups of the same group.

A homomorphism $\rho : G \to H$ maps the unit element e_G of G to the unit element e_H of H. In fact, $\rho(e_G)$ equals e_H because it is idempotent,

$$\rho(e_G)\rho(e_G) = \rho(e_G e_G) = \rho(e_G) \; . \tag{2.52}$$

Furthermore, if $a \in G$ and $b = \rho(a)$, then $b^{-1} = \rho(a^{-1})$. For the inverse of b is unique, and

$$\rho(a^{-1})\, b = \rho(a^{-1})\, \rho(a) = \rho(a^{-1}a) = \rho(e_G) = e_H \; . \tag{2.53}$$

Definition 2.21 *The* "kernel" $\operatorname{Ker}\rho$ *of a homomorphism $\rho : G \to H$ consists of the elements of G which map to the unit element of H,*

$$\operatorname{Ker}\rho = \rho^{-1}(e_H) = \{\, a \in G \mid \rho(a) = e_H \,\} \; . \tag{2.54}$$

Similarly, we denote by $\operatorname{Img}\rho$ *the* "image" *of G under the function ρ,*

$$\operatorname{Img}\rho = \rho(G) = \{\, \rho(a) \mid a \in G \,\} \; . \tag{2.55}$$

Theorem 2.22 $\operatorname{Img}\rho$ *is a subgroup of H, and* $\operatorname{Ker}\rho$ *is a normal subgroup of G.*

Proof. $\operatorname{Img}\rho$ and $\operatorname{Ker}\rho$ are both subgroups because they are closed under multiplication and inversion. $\operatorname{Ker}\rho$ is normal because if $a \in \operatorname{Ker}\rho$ and $g \in G$, then $gag^{-1} \in \operatorname{Ker}\rho$ since

$$\rho(gag^{-1}) = \rho(g)\rho(a)\rho(g^{-1}) = \rho(g)\, e_H\, (\rho(g))^{-1} = e_H \; . \tag{2.56}$$

QED

Theorem 2.23 *Let K be a normal subgroup of the group G, and define $\rho : G \to G/K$ by $\rho(a) = aK$. Then ρ is a homomorphism and* $\operatorname{Ker}\rho = K$*.*

Proof. ρ is a homomorphism because of the multiplication rule for cosets of a normal subgroup, $(aK)(bK) = (ab)K$. To prove that $\operatorname{Ker}\rho = K$, observe that (i) K is the unit element of G/K; (ii) $a \in \operatorname{Ker}\rho$ means that $\rho(a) = K$, or what is the same, $aK = K$; (iii) when K is a subgroup of G, then $aK = K$ if and only if $a \in K$.

QED

This is actually the most general kind of homomorphism, by

Theorem 2.24 *Let $\rho : G \to H$ be a homomorphism, and let $K = \operatorname{Ker}\rho$. If $h \in \operatorname{Img}\rho$, the inverse image $\rho^{-1}(h)$ is both a left and a right coset of K.*

Hence, ρ is an isomorphism between $\operatorname{Img}\rho$ and the factor group G/K.

ρ is an isomorphism between G and H if and only if $\operatorname{Img}\rho = H$ and $\operatorname{Ker}\rho = \{e_G\}$.

Proof. We want to prove that $\rho^{-1}(h)$ is a left and right coset. Let $a \in \rho^{-1}(h)$, that is, $\rho(a) = h$. Then $aK \subset \rho^{-1}(h)$, because

$$\rho(aK) = \rho(a)\rho(K) = he_H = h \; . \tag{2.57}$$

The opposite inclusion $\rho^{-1}(h) \subset aK$ also holds, because $a^{-1}\rho^{-1}(h) \subset K$, in fact,

$$\rho(a^{-1}\rho^{-1}(h)) = (\rho(a))^{-1}\rho(\rho^{-1}(h)) = h^{-1}h = e_H \; . \tag{2.58}$$

It follows that $\rho^{-1}(h) = aK$. Since K is normal, $aK = Ka$.

QED

Definition 2.25 *An isomorphism of the group G with itself is an* "automorphism"*. The automorphisms of G form a group, which we call* $\operatorname{Aut}(G)$*.*

An "inner automorphism" *of G is a conjugation, $\rho(a) = gag^{-1}$ for a given $g \in G$, an* "outer automorphism" *is not of this form. The inner automorphisms form a group called $\mathcal{I}(G)$, which is a subgroup of* $\operatorname{Aut}(G)$*.*

2.9 Direct and semidirect product

Definition 2.26 *Given a group G with two subgroups K and H, where K is normal, and where every $g \in G$ can be decomposed in a unique way as $g = kh$ with $k \in K$ and $h \in H$. Then G is a* "semidirect product" *of K and H, and we write $G = K \times H$.*

A special case is the "direct product" *$G = K \otimes H$, where $kh = hk$ for all $k \in K$ and $h \in H$.*

An even more special case is the "direct sum" *$G = K \oplus H$, which is a direct product where K and H are both Abelian groups and the group operation is written as addition instead of multiplication.*

The notation $G = K \times H$ for the semidirect product is consistent with the fact that the set G is a Cartesian product between the two sets K and H.

Note that if G is Abelian, then both K and H are Abelian and G is the direct sum $K \oplus H$. On the other hand, a semidirect product group $G = K \times H$ which is not a direct product, is always non-Abelian, even when K and H are both Abelian groups. This is therefore a way of constructing non-Abelian groups out of Abelian groups.

Knowing the group products in the subgroups K and H is almost enough if one wants to reconstruct the group product in the semidirect product $G = K \times H$. The product of two elements $g_1 = k_1h_1$ and $g_2 = k_2h_2$ of G is

$$g_1g_2 = k_1h_1k_2h_2 = [k_1(h_1k_2h_1^{-1})][h_1h_2] \; . \tag{2.59}$$

Here $h_1k_2h_1^{-1} \in K$, since K is assumed to be a normal subgroup of G, therefore the two square brackets are group products in K and in H, respectively.

Let us introduce the notation

$$h(k) = C_h(k) = hkh^{-1} . \tag{2.60}$$

By conjugation in the group G, every $h \in H$ acts on K as an automorphism, called C_h or simply h. In other words, the mapping $C : h \mapsto C_h$ is a homomorphism $C : H \to \mathrm{Aut}(K)$. This homomorphism defines uniquely the group product in G, given the group products of K and H. The characteristic property of the direct product is that H acts trivially on K, $h(k) = k$ for all $h \in H$ and $k \in K$.

This discussion suggests the following theorem.

Theorem 2.27 *Given two groups K and H, and an action of H on K, $k \mapsto h(k)$ where $k \in K$ and $h \in H$, such that*

$$h(k_1k_2) = h(k_1)h(k_2) , \qquad [h_1h_2](k) = h_1(h_2(k)) . \tag{2.61}$$

Let G be the Cartesian product of the sets K and H, and define the group product of G as follows. Given three elements $g, g_1, g_2 \in G$, $g = (k, h)$, $g_1 = (k_1, h_1)$, and $g_2 = (k_2, h_2)$. We define $g = g_1g_2$ to mean that

$$k = k_1h_1(k_2) , \qquad h = h_1h_2 . \tag{2.62}$$

This definition turns G into a group, with K and H as subgroups, by the identifications

$$(k, e_H) = k \in K , \qquad (e_K, h) = h \in H . \tag{2.63}$$

The group G is a semidirect product of K and H. In particular, K is a normal subgroup of G.

Proof. We prove associativity of the defined product, and leave the rest of the proof as an exercise. For any three elements $g_1, g_2, g_3 \in G$ we have that

$$\begin{aligned} g_1(g_2g_3) &= g_1(k_2h_2(k_3), h_2h_3) = (k_1h_1(k_2h_2(k_3)), h_1(h_2h_3)) \\ &= (k_1h_1(k_2)h_1(h_2(k_3)), (h_1h_2)h_3) = (k_1h_1(k_2), h_1h_2)g_3 = (g_1g_2)g_3 . \end{aligned} \tag{2.64}$$

QED

Example

From the cyclic groups C_3 of order three and C_2 of order two we may construct the direct product $C_3 \otimes C_2 = C_6$ and the semidirect product $C_3 \times C_2 = S_3$.

In fact, let $a \in G$ and $b \in G$ be group elements of order three and two, $a^3 = b^2 = e$. Thus, a generates a subgroup $C_3 \subset G$, b generates a subgroup $C_2 \subset G$, and G contains (at least) the six different elements e, a, a^2, b, ab, a^2b.

Assume first that a and b commute, and define $c = ab = ba$. Since $c^n = a^nb^n$, we see that c generates a cyclic group of order six, containing the elements $c = ab$, $c^2 = a^2$, $c^3 = b$, $c^4 = a$, $c^5 = a^2b$, $c^6 = e$. This is the direct product group $C_3 \otimes C_2 = C_6$.

The second alternative is that $bab^{-1} = a^2$, or equivalently $ba = a^2 b$, since the mapping $a \mapsto a^2$ is an isomorphism of C_3. Again the six elements $e, a, a^2, b, ab, a^2 b$ define a group. For example, the product of ab and $a^2 b$ is

$$(ab)(a^2 b) = a(ba)ab = a(a^2 b)ab = a^3 (ba)b = (a^2 b)b = a^2 b^2 = a^2 \ . \tag{2.65}$$

This group is the semidirect product $C_3 \times C_2 = S_3$, which is non-Abelian.

2.10 Group extension

When $G = K \times H$, there is one single element of H in every coset of K, and therefore the correspondence $h \leftrightarrow hK$ between H and G/K is an isomorphism. Thus, the semidirect product $G = K \times H$ is one particular solution to the non-trivial problem of *group extension*: given two groups H and K, find a group G such that K is (isomorphic to) a normal subgroup of G and H is (isomorphic to) G/K.

Definition 2.28 *G is an* "extension" *of H by K if $H \simeq G/K$.*

The group extension problem has always one trivial solution, the direct product $K \otimes H$, which exists for any two groups. A non-trivial semidirect product $K \times H$ exists as soon as there exists a non-trivial homomorphism $H \to \mathrm{Aut}(K)$.

There exist however more general group extensions than semidirect products. In fact the relation $H = G/K$ does not imply that H is a subgroup of G, or more precisely it does not imply that there exists a subgroup of G with exactly one element in every left coset of K and therefore isomorphic with $H = G/K$.

A very simple counterexample is C_4, the cyclic group of order four. It has four elements a, a^2, a^3 and $a^4 = e$, and one non-trivial subgroup $K = \{e, a^2\} \simeq C_2$. Since it is Abelian, every subgroup is normal, and the quotient group C_4/C_2, consisting of the two cosets

$$K = \{e, a^2\} \ , \qquad aK = \{a, a^3\} \ , \tag{2.66}$$

is again isomorphic to C_2. Thus C_4 is an extension of C_2 by C_2. But it is not a semidirect product, because there is no C_2 subgroup of C_4 having one element in each of the cosets K and aK. The only semidirect product of two C_2 groups is the direct product $C_2 \otimes C_2 \simeq D_2$.

The simple groups, which by definition have no non-trivial normal subgroups, are not extensions of smaller groups. By the Jordan–Hölder theorem, stated here without proof, the finite simple groups are the basic building blocks from which all finite groups can be built. Thus they play a role in group theory somewhat like the role of prime numbers in number theory. One parallel is for example the fact that the factorization of finite groups into simple groups is essentially unique, like the prime factorization of integers is unique.

Theorem 2.29 (Jordan–Hölder) *All finite groups can be constructed from the finite simple groups by repeated group extensions. Such a construction of a finite group from finite simple groups is called a decomposition series.*

Any two decomposition series of the same finite group involve the same simple groups with the same multiplicities, but not necessarily in the same order.

Unlike the multiplication of two numbers, however, the construction of a group extension from two smaller groups is in general a problem with many solutions, and there is not even known a general theory for how it can be done.

The classification of all finite simple groups was completed just a few years ago. It is contained in hundreds of papers covering thousands of pages in mathematical journals, and is called “the enormous theorem”. However, this is still only a first small step towards a complete classification of all finite groups, since the problem of group extension has not been solved in general.

Problems

1. Consider the symmetry group of the equilateral triangle, Table 1.1.

   Find all its subgroups. Which of these are normal?

2. Prove that the only finite groups that are Abelian and simple, are the cyclic groups of order p where p is a prime number.

3. Recall that the alternating group A_n is defined as the subgroup of the symmetric group S_n consisting of all even permutations. Let $p \in S_n$, and define

$$\begin{aligned} C_1(p) &= \{qpq^{-1} \mid q \in S_n\} \;, \\ C_2(p) &= \{qpq^{-1} \mid q \in A_n\} \;. \end{aligned} \tag{2.67}$$

   Obviously, $C_2(p) \subset C_1(p)$. Show that $C_2(p) = C_1(p)$ if and only if there exists at least one odd permutation $q \in S_n$ commuting with p (so that $qpq^{-1} = p$).

   Show also that if $C_1(p) \neq C_2(p)$, that is, if there exists no odd permutation q commuting with p, then $C_2(p)$ contains exactly half the elements of $C_1(p)$.

   Hint: One possible way to prove both these results is by counting. In fact, $C_1(p)$ and $C_2(p)$ are homogeneous spaces of the two groups S_n and A_n, respectively. By Theorem 1.8 they contain a number of elements equal to the order of the group, either $|S_n| = n!$ or $|A_n| = n!/2$, divided by the order of the fixed point group, which consists of all $q \in S_n$, or $q \in A_n$, with $qpq^{-1} = p$.

4. Using the results from Problem 3 above, describe the conjugation classes of the alternating group A_n.

5. Prove that the alternating group A_n is simple if $n \geq 5$.

   One way to proceed is as follows. Assume that H is a normal subgroup of A_n. We want to prove that either $H = \{e\}$ or $H = A_n$. Since A_n is the only nontrivial normal subgroup of S_n, it is sufficient to prove that H must be normal as a subgroup of S_n.

6. The concepts of inner and outer automorphisms are introduced in Definition 2.25.

   Show that the group of inner automorphisms, $\mathcal{I}(G)$, is isomorphic to $G/Z(G)$, where $Z(G)$ is the centre of G. In mathematical notation, $\mathcal{I}(G) \simeq G/Z(G)$.

   Show that $\mathcal{I}(G)$ is a normal subgroup of the group of all automorphisms, $\mathrm{Aut}(G)$.