Advanced Mathematics

Lecture in WS 2023/24 at the KFU Graz

Axel Maas

Contents

Chapter 1

Introduction

This lecture covers a number of mathematical aspects, which are important in modern theoretical physics. The largest, and most central, part is group theory, or more precisely the theory of groups and algebras, as it plays a quite important and prominent role in modern physics. The reason is that nature exhibits symmetries, and these are consequences of the symmetry groups in their basic structure. These symmetries can appear in many disguises. The simplest one is that the symmetry can be identified by the naked eye, e. g. in snow flakes or in the form of crystals. Then they manifest themselves in conservation laws, which e. g. forces the orbit of a planet to be planar. And finally, they surface as degeneracies, like the fact that without magnetic field the magnetic levels of the hydrogen atom have all the same energies.

The second part deals with functions of complex variables, and their analytical structure. That is of central importance for many aspects of theoretical physics, but especially features of spectra and scattering processes. An example in quantum mechanics is the fact that poles of propagators determine the energy spectra.

Finally, some basic elements of manifolds will be needed, which is necessary for spaces with curvatures, but are also a common foundation for the other subjects. The most notable application in physics is, of course, general relativity, but also various internal spaces throughout theoretical physics are manifolds.

It will be seen that many aspects of group theory reemerge in both manifolds and function theory. That is not surprising as in both cases groups enter in the foundation of them. This again justifies the central part played by groups in the following.

The aim of this lecture is not to provide an explanation of any of the physical phenomena. The aim is to provide the mathematical tools to do so. E. g. one of the tools are group and algebra theory. Implementing these tools such as the observed symmetry, conservation, and degeneracy patterns observed in experiment are correctly reproduced by the theoretical description is the role of the corresponding physics lectures. Nonetheless, many examples will be drawn from physics. Also, the group of rotations and especially of spin will play very central roles, so that many mathematical concepts will have an immediate physical analog.

Chapter 2

Manifolds

In physics one usually encounters an underlying space(-time), like Euclidean space-time and Minkowski space-time. However, these are relatively rigid structures. Many phenomena, like general relativity or more involved internal spaces associated with gauge symmetries, require a more general concept. This concept is provided by manifolds.

2.1 Manifolds as maps

Manifolds as a structure are defined foremost operationally by a map: A manifold is a set $\mathcal{M}$ of elements $m \in \mathcal{M}$, of which overlapping subsets can be mapped to overlapping subsets of $\mathbb{R}^n$, $f(m) \to x$. This implies that manifolds are locally $\mathbb{R}^n$. Hence, every element of the set can be uniquely labeled by a vector in $\mathbb{R}^n$, and there is a minimum n for which this uniqueness can be guaranteed. This is called the dimension of the manifold.

The probably best known example are the $n-1$-spheres S^{n-1}. They are commonly defined by a map from $\mathbb{R}^n$ such that[1] $x_i^2 = 1$. This shows especially how the manifold is only defined via a subset of the $\mathbb{R}^n$. The dimension of the manifold is given by the minimal number of dimensions of the underlying $\mathbb{R}^n$ subset. In the case of the $n-1$-sphere, the dimensionality is $n-1$, as the condition on the sum renders one parameter dependent. However, attempting to use this to reduce the underlying $\mathbb{R}^n$ to $\mathbb{R}^{n-1}$ yields a non-trivial, but solvable, problem.

It is possible to label elements of the set by introducing coordinates by an injective map $\phi(x)$ from the $\mathbb{R}^n$ to the manifold, $\phi(x) \to X(m)$ with $m \in \mathcal{M}$ and $\phi^{-1}(X(m)) = x$. If every element of the manifold can be given coordinates using in this way by a single map $\phi(x)$ it is called a simple manifold. An example of a simple manifold is the one-dimensional unit circle S^1, which is obtained from a finite interval in $\mathbb{R}$ by, e. g., the map $\phi(x) = \exp(ix)$. This already shows that the map is necessarily not unique. E. g. $\phi(x) = 1 + \exp(ix)$ would work as well. Also, while in this case the set S^1 is represented by points in the complex plane, this is not the only possibility, and in general the map can be as abstract as a table.

[1] Throughout this lecture the Einstein convention is used. Thus exactly doubly appearing indices are always summed over their index range, as long as not stated explicitly otherwise.

If the manifold is not simple, then the overlapping feature allows for the construction of an atlas, which is the set of all coordinate systems needed to describe the complete manifold. E. g., given two subsets with two coordinate maps ϕ_1 and ϕ_2, the injectivity allows to construct uniquely coordinates $X(m)$ for any element m in the manifold in the overlap region by transition functions $\phi_2(\phi_1^{-1}(X^1(m)))$ and $\phi_1(\phi_2^{-1}(X^2(m)))$, as they need to satisfy $\phi_1^{-1}(X^1(m)) = x = \phi_1^{-2}(X^2(m))$. Using such transition functions, it is possible to move from any element in the manifold to another. The atlas is then the collection $\mathcal{A} = \{\phi_i\}$. Thus, a manifold with coordinates is defined by the set $\{\mathcal{M}, \mathcal{A}\}$.

Where this subtlety comes into play can be seen by the two-sphere S^2, which is as a two-dimensional manifold. It is not simple, as the plane $\mathbb{R}^2$ is insufficient to describe the whole sphere, because either the north pole or the south pole is ill-defined. Introducing as coordinate maps spherical coordinates, but once counting the azimuth angle from 0 and once from π gives two coordinate maps, which overlap everywhere except at the poles, and the transition function is $\theta - \pi$. The atlas contains then two coordinate systems. Of course, there could be additional coordinate systems. Thus, the atlas is usually defined to be the minimal set of necessary coordinate systems[2]. On the other hand, by having a higher-dimensional underlying $\mathbb{R}^n$, the atlas has again only one coordinate system, at the price of a redundancy. This is the trade-off often encountered. This is called an embedding. It is worthwhile to note that it is not guaranteed that such an embedding exists.

If there exists as set of maps and transition functions which cover the whole manifold and $\Phi_{ij}(x) = \phi_i^{-1}(\phi_j(x)) \in \mathbb{R}^n$ for overlapping patches i and j are differentiable in the usual sense as functions of $x \in \mathbb{R}^n$, the manifold is called differentiable. If the underlying real space is even-dimensional, and can be mapped to a complex vector space such that all necessary maps and transition functions are so-called holomorphic, the manifold is holomorphic. The definition of holomorphic will be given in section 3.3.2.

2.2 Topology

Just like vector spaces, which are not normed, manifolds have so far no concept of distance. By introducing a topology, i. e. a map $d(m, n)$ from two elements m and n of the manifold $\mathcal{M}$ into $\mathbb{R}$, such a distance measure is introduced. Usually, and essentially in the whole lecture, it will be required that $d(m, n) \geq 0$ with equality if and only if $m = n$, i. e. $d(m, m) = 0$. This condition can be relaxed, but then various statements no longer hold. A manifold equipped with a topology, $\{\mathcal{M}, \mathcal{A}, d\}$, is called a topological manifold.

With this, it is possible to introduce the usual concept of closeness. Especially, this allows to define the neighborhood of an element m by including all points n which satisfy $d(m, n) < \delta$, with δ some number. Usually, this is called a ball. Isolated points are elements for which a value of δ exist, for which no elements other than m itself remains inside the ball. Conversely, an accumulation (or limit) point is whenever for any δ there are always at least one other element within the ball. It should be noted that the closeness of m and n in the manifold do not imply (necessarily) the closeness of the corresponding

[2]Sometimes the atlas is considered to be any set of complete coordinate systems, and one with the minimal number of needed ones is called minimal.

point x and y in the Euclidean sense of the underlying $\mathbb{R}^n$. However, in practice this happens often at least within a common coordinate patch.

A (sub)set of a topological manifold is bounded, if the distance between any two elements is finite. Closed and open (sub)sets are defined using Cauchy sequences[3] of elements. A closed set then requires that the limit of every Cauchy sequence converging within the set is also within the set. Otherwise the set is called open. If it is both bounded and closed, it is called compact. For any finite-dimensional manifold it can be shown that all closed subsets are also bounded. Compactness can also be established by the Heine-Borel overlapping criterion that a set is compact if any union of open sets containing the set contains a finite number of sets which also contains the original set. Note that in a topological manifold this implies that therefore an atlas exists such that points close in the manifold are also close in the Euclidean norm in the underlying $\mathbb{R}^n$. This gives a specific case of the situation mentioned above.

Conversely, the Heinel-Borel overlapping criterion allows to consider cases in which the topology can also take values zero for distinct points, or may even become negative. In that case, points can be considered close, if the are within the same coordinate patch, and the Euclidean distance in the underlying $\mathbb{R}^n$ is smaller than some number.

Given a topology and an atlas, it may be possible to introduce a metric. If there exists coordinates $X(x)$ and $y(Y)$ for points in the manifold, and a (semi-)scalar product $(X-Y)^\dagger g(X-Y)$ can be defined for $X(m)$ and $Y(n)$ such that it equals $d(X,Y) = d(X(m),Y(n))$, then this defines a metric g on the manifold. Of course, the metric will depend on the choice of coordinate system, as usual. It will also in general depend on X and Y, $g(X,Y)$. A metric is said to be compatible, if $g(X,Y) = g(X) = g(Y)$, if $d(X,Y)$ is infinitesimal. Note that while both the coordinates and the metric will depend on the choice of coordinate system, the definition requires that the distance $d(X,Y)$ does not. Thus, changes of the metric need to compensate changes in the coordinates. This provides rules for the how the metric changes under a coordinate transformation, and allows to define covariant (transforms like a coordinate) and contravariant (transforms opposite to a coordinate) tensors in the tangential space, at least inside a coordinate patch. The metric will then necessarily be able to transforms covariant object into a contravariant one. Likewise, an inverse metric, in the sense of being the inverse of the metric but transforming covariantly, will do the opposite.

If the metric has everywhere only positive eigenvalues, the manifold is called a Riemannian manifold, if not a pseudo-Riemannian manifold.

2.3 Paths and tangent spaces

Once a manifold with a topology is given, it is possible to define curves $X(t)$, where t is a real parameter, perhaps restricted to an interval, as a set of elements $m(t)$ labeled by the values of t and with assigned coordinates $X(t)$. Note that the curve may move through different patches, requiring to use transfer functions.

[3] A Cauchy sequence of elements $m_n \in \mathcal{M}$ is a sequence for which for any $\epsilon \in \mathbb{R}$ there exists an index n_ϵ such that for any $k > n_\epsilon$ and $l > n_\epsilon$ $d(m_k, m_l) < \epsilon$ holds.

A continous curve requires that for any ϵ there exists for any point $X(t)$ a point $X(t+\epsilon)$ such that $d(X(t), X(t+\epsilon)) < \delta$ for any arbitrarily small δ. As t is a real parameter, this likewise allows to introduce the concepts of differentiable curves as the rate of change of the coordinates at a point t. A curve is closed, if there exists a T such that $X(t) = X(t+T)$ for any t. If no such T exists, it is open. If the condition is fulfilled only for some, but not all t, the curve crosses itself.

The existence of differentiable curves allows to attach at every element of the manifold a vector space. To this end, there will be as many independent derivatives as there are dimensions. These derivatives can then be used to define base vectors. By defining their sum to be in the corresponding direction, this provides vector addition, and supplementing with a (usually real) body, this creates a vector space, the tangent space of the manifold at this point.

Curve are invariant under redefinitions of t, as the set of points does not change. A linear change is called an affine transformation. Thus, it is especially possible to map the parameter to the unit circle[4]. Note that the curve does not need to be closed for this.

2.4 Homotopy

Curves on manifolds can be classified. This classification has important consequences, as will be seen, e. g., in chapter 3. It plays an important roles in many cases and fields.

To construct the classification, consider two curves, $\mathcal{C}_i$, with their parameters t_i taking values on the unit circle. It is now possible to have both circles of parameters being located on the surface of a cylinder. Then the height h on the cylinder can be used to smoothly deform the parameter of both curves into each other. If the curves can in this process be themselves smoothly, i. e. remaining continous at every step of the deformation, deformed into each other, the two curves are said to be (freely) homotopic to each other. This can be stated by using the deformation parameter h, which is taken to be between zero and one. Then $X(h,t)$ is a continous function of both h and t, with $X(0,t)$ being $\mathcal{C}_1(t)$ and $X(1,t)$ being $\mathcal{C}_2(t)$. Note that the curves do not need to have the same length for this procedure, no matter how length is defined in this context.

All curves, which can be deformed into each other in this way are collected in a homotopy class. If the curves can be deformed to a point, i. e. the point as well as non-zero length curves are included in the class, the class is called the trivial class. If in a given space all curves belong to the trivial class, the space is called simply connected. This definition applies especially to manifolds. E. g., in a two-dimensional space with a point excluded, given a closed curve around such a puncture of the space, the curve cannot be deformed to a point. Its class is therefore non-trivial.

It may now happen that curves are such that it is not directly possible to just map the parameters in a one-to-one way on each other. Consider e. g. such non-contractible curves on the punctured space. Curves winding once and twice around the puncture cannot be deformed into each other. That can be visualized by having the parameter

[4]If t is not from a finite interval, this is still possible, as discussed in section 4.12.

wind around the cylinder n times, if the puncture is encircled n times. Likewise, curves which circle in opposite directions can also not be mapped into each other. Thus, for a such singly punctured space, the possible homotopy classes are given by the integers. Since combining multiple curves by following them one into another changes the integer, this can be considered the group of integers, $\mathbb{Z}$. Thus, $\mathbb{Z}$ is called the homotopy group of this space. The genus is then defined to be the number of such holes, i. e. non-identical possibilities to have non-contractible paths. E. g., the sphere has genus zero and the donut or the punctured plane genus one, .

In general, it is possible to define hypersurfaces in manifolds with topology. The original parameters can then be mapped to the n-dimensional sphere, rather than the 1-dimensional circle. If the space in which the curve resides is the space M, then the homotopy group $\pi_n(M)$ is defined to be the group of the homotopy class.

E. g. in a simply-connected space, say $\mathbb{C}$, all curves belong to the same class, and thus there is only the trivial homotopy group, denoted $\pi_1(\mathbb{C}) = 0$. However, consider the two-dimensional torus. Curves which wrap n-times around the ring and m-times around the inner hole cannot be deformed continuously into each other, when they have $n' \neq n$ and/or $m' = m$. Thus, there are two integers labelling the homotopy classes and thus the group is $\pi_1(T) = \mathbb{Z}^2$.

The group structure is imposed by defining group composition to be composition of curves from different classes. Spaces with different homotopy group(s) are globally different, and thus there cannot exists homorphisms between them. The homotopy groups therefore capture the global structure of manifolds. It can be shown that all homotopy groups $\pi_{n>1}(M)$ are Abelian, but it does not hold true that if M is 'bigger' than S^n it is always trivial. E. g. $\pi_3(S^2) \neq 0$, which is stated here without proof.

Determining homotopy groups is generally relatively difficult, but they are known for many cases, and for certain classes of spaces there exist constructive approaches to obtain them.

Chapter 3

Complex functions

Complex-valued functions of complex variables are a common sight in physics, especially in quantum physics. At the same time, singularity structures play an important role in quantum physics as well, e. g. in the resolvent. These two issues seem to be unrelated at first. There is, however, a deep connection between them, and its formalization is the basis of function theory. Interpreting $\mathbb{C}^n$ as a manifold with topology allows to sees these as implementations of homotopy classes.

3.1 Patches and curves

Unsurprisingly, the arena of function theory are complex vector spaces $\mathbb{C}^n$. These are isomorphic to real vector spaces $\mathbb{R}^{2n}$. Of course, the multiplication of two complex numbers is not a simple structure on the real space. Most of the following can be straightforwardly continued to $n > 1$, and thus the primary arena will be the complex plane $\mathbb{C}$. These are just the ordinary complex numbers, endowed with the known structures and simple operations.

Most of the usual statements and definitions of analysis can be carried over. Only that complex numbers cannot be ordered requires occasionally to resort to the absolute value of a complex number. Especially, convergence of a series z_n is given by a convergence of $|z_n|$. In turn, by virtue of the triangle inequality, this is implied by convergence of both $|\Re z_n|$ and $|\Im z_n|$. Boundedness is also imposed in terms of the absolute value, and is in turn implied by the boundedness of both real part and imaginary part. Using the absolute value, it is possible to translate criteria of convergence to complex, infinite series from real numbers.

For infinite sums convergence can be obtained as well from the usual concept of real numbers. Especially important in the following is the Cauchy criterion, i. e. there exists some indices n and m such that

$$\left| \sum_{k=m+1}^{n} z_k \right| \leq \epsilon$$

for any arbitrary small ϵ and all indices larger than n and m. An infinite sum is absolutely convergent if the sum of $|z_n|$ converges. The usual criteria for absolute convergence can

be carried over from real numbers when using again the absolute value.

Definitions of subsets, open sets, closed sets, etc. are carried over from manifolds by using the absolute value of differences of complex numbers as a topology. Especially useful will be the definition that a set is bounded if the limit of any convergent series inside the set also belongs to the set. Also useful is the concept of a limit or (ac)cumulation point z of a set, which here takes the form that for any arbitrarily small ϵ there exist a second point z_ϵ in the same set such that

$$|z - z_\epsilon| < \epsilon.$$

Including all limit points into a set generates the bounded hull of the set. Moreover, if all subsequences converging to the same limit as the full sequence are contained in a set, this set is compact.

Together with the condition that a set is bounded if all elements of the set are bounded the important statement is valid that any closed and bounded set is compact. These concepts play an important role for the viability of statements on complex functions, as it will be important on which kind of sets they have which properties.

Also necessary will be the implementation of (continous, oriented) curves or paths $\mathcal{C} = z(t)$ in $\mathbb{C}^n$, which map some (continous, oriented) real interval $[t_1, t_2]$ into the complex numbers. It will be important to differentiate between curves which do not cross themselves, and those which do. The length of a curve is defined as

$$l_\mathcal{C} = \int_{t_1}^{t_2} dt \left| \frac{dz(t)}{dt} \right|,$$

and thus by the usual integration. If this integral exists in the Riemannian sense and is finite, the curve is called rectifiable. If $z(t_1) = z(t_2)$ the curve is called closed. In the following any oriented, closed path will be taken to be traversed in the mathematically positive sense by convention. The interior of a closed curve are all points left to the closed curve when traversing it in this sense. It should be noted that this is an implementation of the paths in 2.4. At the moment, all curves are belonging to the trivial homotopy class, and the complex plane is simple.

This concept is used to define two important features. A (sub)set G of $\mathbb{C}$ is also called a patch. It is called connected if any two points in it can be connected by a curve, which is entirely within G. It is called simply connected if the interior of any curve in G is also contained in G. Note that a patch can be open, and thus not include its boundary. As a consequence, any path in an open patch G cannot approach the boundary of G (which does not belong to the open patch itself) arbitrarily close.

3.2 Holomorphic functions

A complex function can be considered to be decomposed as

$$f(z = x + iy) = u(x, y) + iv(x, y), \tag{3.1}$$

and thus two functions of two variables. This will be a very useful decomposition for many topics. This allows to define many properties of complex functions in terms of real functions of two variables. E. g. (steady) continuity of the complex function is then equivalent to both real functions being (steadily) continous in both arguments.

One interesting question is on differentiation of complex functions. Because complex numbers do not have the usual concept of ordering, the usual definition of differentiation does not work. Rather, the derivative of a complex function df/dz with respect to its variable z at a fixed value of this variable Z is defined to be the unique quantity, which obeys

$$\left| \frac{f(z) - f(Z)}{z - Z} - \left. \frac{df}{dz} \right|_{z=Z} \right| \leq \epsilon \tag{3.2}$$

and taking the limit of ϵ to zero while taking $|z - Z|$ to zero. If this is possible for some range of values Z, the function is said to be differentiable on this range. If this is true for the whole domain of definition of the function, the function is called differentiable.

This definition has the advantage to reduce to the usual definition for real functions of real variables, if $f(z) = u(x)$. Furthermore, the derivative of most complex functions can be obtained by taking the results for real functions of real variables, and replace all quantities by complex ones. This is more involved if the function depends on both z and z^*. In this case, things become more subtle, and looking at the formulation (3.2) is often necessary.

Investigating along these lines leads to the very important statement that if a function $f(z)$ is continuously differentiable, this requires that for the form (3.1) the so-called Cauchy-Riemann partial differential equations

$$\partial_x u(x,y) = \partial_y v(x,y) \tag{3.3}$$
$$\partial_y u(x,y) = -\partial_x v(x,y) \tag{3.4}$$

are obeyed, where the derivatives need to be continous. Conversely, if the Cauchy-Riemann partial differential equations are not fulfilled, the function will not be differentiable. This will play an important role in the following. The proof follows from explicitly evaluating (3.2) for the form (3.1).

This now yields the interesting result that the function $f(z) = (z + z^*)/2 = \Re(z)$ is actually not differentiable, as this explicitly violates (3.3-3.4). Conversely, this implies that the condition (3.2) cannot be satisfied for any z. This can be seen by the fact that the imaginary part in the denominator can be varied independently from the nominator, leading to the contradiction. While this may seem at first surprising, as one often is used to do calculations with complex numbers just as with ordinary numbers, there is a reason for that. Effectively, differentiation of a complex function as if it would be a function of a single variable implies that the approach to the point where the derivative should be evaluated should be independent from the direction on how to approach it. That is a relatively powerful constraint, and hence this is much harder to satisfy as if there is only one direction, as is the case for real functions of real variables[1].

[1]It is also not a coincidence that (3.3-3.4) are reminiscent of rotation or Hamilton's equations. In these cases lines within a higher-dimensional setting play a significant role as well.

Because of the conditions (3.3-3.4), it is now possible to determine the derivative of arbitrary complex functions obeying the Cauchy-Riemann differential equations in terms of the derivatives of real functions of real variables as

$$d_z f(z) = \partial_x u + i\partial_x v = -i(\partial_y u + i\partial_y v).$$

Also, any function not obeying the Cauchy-Riemann differential equations cannot be differentiated, and especially not in this way. Because these are special additional conditions, being a differentiable complex function of a complex variable earned a special name, they are called holomorphic[2] . As will be seen being holomorphic entails a large number of consequences. The downside is that being holomorphic is much more restrictive than being differentiable for a real function of a real variable. Hence, much less functions fall into this category. In physics, holomorphic functions appear quite often, but likewise being not holomorphic can be a hallmark of non-trivial physics.

An important feature of holomorphic functions is that they cannot depend simultaneously on z and z^*, because this would violate immediately the Cauchy-Riemann differential equations (3.3-3.4). This has far-reaching consequences.

If the component functions are twice continuously differentiable it follows from (3.3-3.4) that

$$\partial_x^2 u + \partial_y^2 u = 0 \tag{3.5}$$
$$\partial_x^2 v + \partial_y^2 v = 0, \tag{3.6}$$

i. e. the real and imaginary components of the function fulfill the so-called partial differential equations of Laplace. Essentially, for every holomorphic function, u and v need to fulfill Euclidean wave equations. Thus, holomorphic functions can only differ by harmonic functions, i. e. functions fulfilling (3.5-3.6) which vanish at the boundaries, and by the boundary conditions themselves. This again shows how highly restricted holomorphic functions are.

Furthermore, any combination of holomorphic functions in a holomorphic way, i. e. by differentiable operations, yields again holomorphic functions. This includes composition and inversion. As a consequence, any polynomial and any rational function of a single complex variable are holomorphic. Especially, this includes (convergent) polynomials of the form

$$f(z) = \sum a_k (z - z_0)^k, \tag{3.7}$$

with some fixed z_0, i. e. polynomials of the same structure as obtained in Taylor expansions. An important definition in this context is the radius of convergence ρ, which is the largest value of $|z - z_0|$ for which (3.7) is convergent. If convergence is only obtained for $z = z_0$, which is guaranteed, the radius of convergence is zero. If the series converges for any value of z, the radius of convergence is infinite. The radius of convergence plays an important role in physics e. g. in the context of (thermodynamic) phases. It can be shown that the

[2]Note that the word 'analytic' is also used, and is synonymous in the present context for holomorphic.

radius of convergence can be obtained from the (complex) coefficients of (3.7), and is given by

$$\frac{1}{\rho} = \lim_{k\to\infty} \sup|a_k|^{\frac{1}{k}},$$

with the usual caveats for ρ being zero and infinite. It can be shown that the polynomial is then a continous function within its radius of convergence. Furthermore, it follows that the radius of convergence of derivatives of a polynomial is the same as the radius of convergence of the original polynomial, or larger. Hence, they are infinitely often differentiable, and their derivatives are obtained by deriving every term individually.

A special example is the complex exponential. Due to the Euler formula

$$e^z = e^{\Re z}\left(\cos \Im z + i \sin \Im z\right),$$

the complex exponential takes all its possible value within a finite strip of the imaginary part of z, and is then periodically continued. It is useful for the following to define the fundamental strip to be the strip satisfying $-\pi < \Im z \leq \pi$. In this fundamental strip the exponential function can be unambiguously inverted for $z \neq 0$, and takes actually all values in $\mathbb{C}$, except zero. This is a rather surprising insight: A proper subdomain of $\mathbb{C}$ is mapped to (almost) the entirety of $\mathbb{C}$ by the exponential function. While it is not the only function with this property, this neatly exemplifies the concept of infinities.

On the other hand, the inversion of the exponential is the logarithm, but it now becomes a multivalued function, due to the periodicity of the exponential,

$$\ln e^z = z + 2k\pi i, \tag{3.8}$$

for any integer number k. By definition, the principal value of the logarithm is given by the case $k = 0$, i. e. where the result is in the fundamental strip. However, it is not possible to obtain a value for $\Re z \leq 0$. Thus, the principal value of the logarithm yields the fundamental strip without the negative real axis (and zero).

It is useful to note that for $|z| < 1$ there is a series representation of the logarithm as

$$\ln(1+z) = \sum_{i=1} (-1)^{k+1} \frac{z^k}{k}.$$

The radius of convergence is thus arbitrarily close to one, but excluding one.

3.3 Cauchy's integration formula

3.3.1 Integrals along curves

Defining an integral of a function of a complex variable is not entirely obvious, as, like for differentials, it is not directly obvious what are the relevant quantities.

One way is to define such integrals using the concept of curves. Given some function $f(z)$ and a curve $z(t)$, say on an interval $[0,1]$, the integral can be defined as

$$\int_{z(0)}^{z(1)} f(z)dz = \lim_{n\to\infty} \sum_{k}^{n} (z(t_k) - z(t_{k-1}))f(z(t_\kappa)),$$

with $t_\kappa \in [t_k, t_{k+1}]$ arbitrary and the t_k are a discretization of the parameter range, such that in the limit the difference between any two consecutive values becomes arbitrarily small. This is the most direct generalization of the usual Riemann integral. It can then be proven, essentially as for ordinary Riemann integrals, that the resulting integral is independent of the parametrization of the curve as well as on how the interval is split.

Furthermore, this allows to use substitution to rewrite the curve integral into an ordinary integral as

$$\int_{z(0)}^{z(t)} f(z)dz = \int_{0}^{1} f(z(t))\frac{dz(t)}{dz}dt, \tag{3.9}$$

which can be turned into separate integrals for the real part and imaginary part of the integrand. Note that this implies that the value changes sign if the curve is traversed in opposite direction. Furthermore, the integral can be decomposed in a sum of integrals over parts of the curve, provided the total curve is regained, without doubling, from the individual curves of each term. Finally, if the function $|f(z)|$ is bounded by a constant c, then the integral is bounded by $cl_{\mathcal{C}}$

The result will depend, in general, on the chosen curve. Consider e. g. the function $f(z) = z^*$ and the two paths $(1+i)t$ and $t^2 + it$. Both paths coincide for the initial and final values $z(0)$ and $z(t)$. The results are

$$\begin{aligned} \textstyle\int_0^1 (1-i)t \times (1+i)dt = 2\int_0^1 tdt = & \quad 1 \\ \textstyle\int_0^1 (t^2 - it) \times (2t+i)dt = \int_0^1 (2t^3 + 3it^2 - t) = & \quad i, \end{aligned}$$

which are indeed different.

Interestingly, there also exists functions, for which the result is indeed independent of the curve. Consider $f(z) = z$. Then

$$\int_{z(0)}^{z(1)} z\frac{dz}{dt}dt = \int_{z(0)}^{z(1)} \frac{1}{2}\frac{dz^2}{dz}\frac{dz}{dt}dt = \frac{1}{2}\int_{z(0)}^{z(1)} dz^2 = \frac{1}{2}\left(z(1)^2 - z(0)^2\right),$$

in which it was used that the derivative was known. Thus, it appears to be an interesting question, under which conditions such a curve integral becomes independent of the path.

It becomes even more interesting when considering the case $f(z) = (z - z_0)^m$ with m integer. Take as a path a circle around z_0, $C(t) = z_0 + r\exp(it)$. This yields

$$\int_{\mathcal{C}} (z - z_0)^m dt = \int_0^{2\pi} r^m e^{imt}(ire^{it}) = ir^{m+1}\int_0^{2\pi} e^{i(m+1)t}dt = \begin{cases} 2\pi i \text{ for } m = -1 \\ 0 \text{ otherwise} \end{cases}. \tag{3.10}$$

Thus, the integral vanishes except if the function behaves like $1/z$ at z_0. This is in so far as surprising as for $m < -1$ it again vanishes, even though the function becomes more singular. Compared to the case of a real function $1/x$, this is odd. Also, the result is independent of both z_0 and r, and thus the same for any such curves. This is also odd, and hints to the fact that global structures of the curves are more important than its details. This is reminiscent of the homotopy classes, and will indeed be related to it.

3.3.2 Closed curves and holomorphic functions

It turns out that (3.10) is actually not an oddity, but is a particular example of a generic statement about integration of complex functions with singularities along closed curves. This is the celebrated theorem of Cauchy.

The first step is that for a holomorphic function $f(z)$ it follows that

$$\int\limits_{\mathcal{C}_1} f(z)dz = \int\limits_{\mathcal{C}_2} f(z)dz, \tag{3.11}$$

provided the two curves have the same starting point and end point. This implies for the integral of a holomorphic function inside a patch along any closed curve bounding this patch

$$\int\limits_{\mathcal{C}_{\text{closed}}} f(z)dz = 0. \tag{3.12}$$

As the former version follows from cutting the curve in the second version, it suffices to prove this one.

As any patch, which is enclosed by the curve, can be approximated arbitrarily well due to triangulation by triangles, it is sufficient to discuss the situation if the curve encloses a triangle. Then any other case can be build up from there by triangulation[3]. Given any triangle, it can be decomposed into more triangles. For the following, it is convenient to decompose it into four triangles enclosed by curves $\mathcal{C}_i$. Selecting the inner boundaries such that they are traversed in each direction twice with opposite orientation, the union of the four curves $\mathcal{C}_i$ is then $\mathcal{C}$, and the total integral, due to the composition rules, obeys

$$\int\limits_{\mathcal{C}} f(z)dz = \sum_i \int\limits_{\mathcal{C}_i} f(z)dz.$$

This implies

$$\left|\int\limits_{\mathcal{C}} f(z)dz\right| \leq 4 \max_i \left|\int\limits_{\mathcal{C}_i} f(z)dz\right| \tag{3.13}$$

as an upper bound to the integral. In particular, the triangles can always be chosen such that they have equal curve length, and thus $l_{\mathcal{C}_i} = l_{\mathcal{C}}/2$.

[3]Fractal boundaries make things cumbersome, but can be done.

This procedure can be repeated, by deconstructing always the triangle identified by the maximization in (3.13) again into four more. In this procedure, there will always remain at least one point in the patch, call it z_0. Because $f(z)$ is holomorphic, it is necessarily differentiable in z_0, and thus in some patch including z_0 follows that there exists some arbitrarily small ϵ such that

$$|f(z) - f(z_0) - (z - z_0)d_z f(z_0)| \leq \epsilon |z - z_0|$$

is true for all z in the remaining triangle. If it is not yet true, sufficient further subdivisions of the triangle will make it so. Especially, this implies that there is some bounded function $|\eta| \leq \epsilon$ such that

$$f(z) = f(z_0) + (z - z_0)d_z f(z_0) + \eta(z - z_0),$$

which is not coincidentally similar to a Taylor series.

By continuing the partitioning of the triangles sufficiently often to a level k, this implies that for the selected triangle bounded by the curve $\mathcal{C}_k$

$$\int\limits_{\mathcal{C}_k} f(z)dz = f(z_0)\int\limits_{\mathcal{C}_k} dz + d_z f(z_0)\int\limits_{\mathcal{C}_k}(z - z_0)dz + \int\limits_{\mathcal{C}_k}\eta(z - z_0)dz$$

follows. The first three terms can be calculated explicitly, and all yield zero, according to (3.10). Hence,

$$\left|\int\limits_{\mathcal{C}_k} f(z)dz\right| = \left|\int\limits_{\mathcal{C}_k}\eta(z - z_0)dz\right| \leq \epsilon\frac{l_{\mathcal{C}_k}}{2} = \epsilon\frac{l_{\mathcal{C}}}{2 \times 4^k}$$

But this implies

$$\left|\int\limits_{\mathcal{C}} f(z)dz\right| \leq 4^k\left|\int\limits_{\mathcal{C}_k} f(z)dz\right| \leq \frac{\epsilon}{2}l_{\mathcal{C}}$$

Since $l_{\mathcal{C}}$ is fixed, and ϵ can be made arbitrarily small, this implies that the absolute value of the integral is bounded from above with an arbitrarily small number, and thus the integral indeed vanishes, as claimed. Note that (3.10) is not in contradiction to this. $1/(z - z_0)$ is not holomorphic in z_0, and thus the proof does not apply. Conversely, $(z - z_0)^{m<-1}$ may also be not holomorphic in z_0, but the statement does not exclude the contrary, that integrals may vanish along a closed curve even if the function to be integrated is not holomorphic.

An interesting consequence is the following. Consider now a patch, which is not bounded by a single curve, but two disjoint curves $\mathcal{C}_1$ and $\mathcal{C}_2$, where one is fully contained in the other, and traversed in the opposite direction. E. g. a ring-like patch with a whole in the middle is of this type. It is then possible to create two new curves $\mathcal{C}'$ and $\mathcal{C}''$, by inserting two connecting pieces between both curves $\mathcal{C}_a$ and $\mathcal{C}_b$ in the points a and b at the outer curves joining the inner curve at points c and d, which are traversed in opposite directions. Because of (3.11), this can be used to separate the original patch into

two patches, each bounded by a single curve, yielding

$$0 = \int_{\mathcal{C}'} f(z)dz = \left(\int_{\mathcal{C}_1:a\to b} + \int_{\mathcal{C}_a:b\to c} + \int_{\mathcal{C}_2:c\to d} + \int_{\mathcal{C}_b:d\to a} \right) f(z)dz \tag{3.14}$$

$$0 = \int_{\mathcal{C}''} f(z)dz = \left(\int_{\mathcal{C}_1:b\to a} - \int_{\mathcal{C}_a:b\to c} + \int_{\mathcal{C}_2:c\to d} - \int_{\mathcal{C}_b:d\to a} \right) f(z)dz \tag{3.15}$$

where the integral again vanishes by virtue of (3.12), since the function is holomorphic on all involved patches. This is a first example of how deformations of curves maintain the integration result for holomorphic functions. Even more, this implies, by adding (3.14) and (3.15),

$$\int_{\mathcal{C}_1} f(z)dz = \int_{\mathcal{C}_2} f(z)dz$$

and thus it does not matter which of the two curves are used to calculate the result. This implies the interesting result that the two curve integrals, which are traversing their curves in opposite order, are actually not very sensitive to what lies in between or inside of both. Note that it was not necessary that $f(z)$ is holomorphic inside $\mathcal{C}_2$.

This statement can be extended to the situation that there are several closed curves $\mathcal{C}_i$, which are not intersecting, and are all contained within one curve $\mathcal{C}$, but not within each other. In analogy it can then be shown that

$$\int_{\mathcal{C}} f(z)dz = \sum_i \int_{\mathcal{C}_i} f(z)dz.$$

Again, it is sufficient that $f(z)$ is holomorphic only in the patch defined by all curves, and on the curves themselves. Cases, in which there are curves within curves can then be reduced recursively in this way.

3.3.3 Integration and differentiation

In similarity to the main theorem of integration and differentiation of real functions, a similar statement holds for functions of complex variables. However, the curve-dependence yields both a constraint and an additional consequence.

Given a continous function $f(z)$ in some (singly-connected) patch, for which any curve integral connecting two points inside the patch does not depend on the curve, there is a holomorphic primitive $F(z)$ satisfying

$$F(z) = \int_{\mathcal{C}_{z_0\to z}} f(z')dz' \tag{3.16}$$

$$f(z) = \frac{dF}{dz}. \tag{3.17}$$

If the integral would be curve-dependent, the resulting primitive would not be unique in satisfying the derivative condition. A dependence on z_0 is giving, as usual, only an additive constant for the primitive, which does not alter its derivative.

This can be seen as follow. The definition for derivatives of holomorphic functions (3.2) implies

$$\lim_{z\to z'}\left|\frac{F(z)-F(z')}{z-z'}-f(z)\right|=0,$$

where the claim (3.17) has already been inserted. Because of the required curve-independence, the curve $\mathcal{C}_{z_0\to z}$ can be decomposed into two consecutive curves $\mathcal{C}_{z_0\to z'}$ and $\mathcal{C}_{z'\to z}$. Because of the arbitrariness of the path, the second one can be chosen to be a straight line[4], with $l_{\mathcal{C}_{z\to z'}}=|z-z'|$. Then

$$\begin{aligned}
&\lim_{z\to z'}\left|\frac{\int_{\mathcal{C}_{z_0\to z'}}f(z'')dz''-\int_{\mathcal{C}_{z'\to z}}f(z'')dz''}{z-z'}-f(z)\right| \\
=\;&\lim_{z\to z'}\left|\frac{\int_{\mathcal{C}_{z'\to z}}f(z'')dz''}{z-z'}-f(z)\right|=\lim_{z\to z'}\left|\frac{f(z)\int_{\mathcal{C}_{z'\to z}}dz''}{z-z'}-f(z)\right| \\
=\;&\lim_{z\to z'}\left|\frac{f(z)(z-z')}{z-z'}-f(z)\right|=0.
\end{aligned}$$

In step two it was used that the function is bounded, but because of the limit can be bounded by the value at coincidence. In the third step the absolute value for the length of the curve can be dropped, as the orientation of the curve ensures that it has the same sign as the numerator. This proofs the claimed definition of the primitive.

3.3.4 Cauchy's theorem

The situation with $m=-1$ in (3.10) is, as indicated, actually only a special case of a much more general statement, Cauchy's theorem. It states that for any closed curve $\mathcal{C}$ and a point z in the interior of the curve inside a patch in which a function $f(z)$ is holomorphic

$$f(z)=\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{f(z')}{z'-z}dz' \tag{3.18}$$

holds. This statement is often cast in the form that the value of the integral of the integrand $f(z')/(z'-z)$ takes the value of the so-called residuum $f(z)$ at the position of the pole z, when integrated around a curve containing the pole. Note that if the pole z is not in the interior of the curve, the complete integrand is holomorphic, and thus the integral then vanishes.

To proof this statement, note that because of (3.10)

$$\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{f(z')}{z'-z}dz'=\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{f(z)}{z'-z}dz'+\frac{1}{2\pi i}\int_{\mathcal{C}}\frac{f(z')-f(z)}{z'-z}dz' \tag{3.19}$$

[4]If the patch has a wedge here, this requires more steps, but works analogously.

holds with

$$f(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f(z)}{z' - z} dz' \tag{3.20}$$

being a very involved way of writing unity. Because it was shown in section 3.3.2 that the integration path can be deformed arbitrarily as long as it does not cross the singularity in z, it is possible to find a circular curve $\mathcal{C}_\rho$ of radius ρ, which lies entirely in $\mathcal{C}$, and is centered around z,

$$\frac{1}{2\pi i} \int_{\mathcal{C}} \frac{f(z') - f(z)}{z' - z} dz' = \frac{1}{2\pi i} \int_{\mathcal{C}_\rho} \frac{f(z') - f(z)}{z' - z} dz'.$$

Because $f(z)$ is holomorphic and thus bounded, this radius can be chosen such that $|f(z) - f(z')|$ becomes smaller than any desired ϵ. Thus

$$\left| \frac{1}{2\pi i} \int_{\mathcal{C}_\rho} \frac{f(z') - f(z)}{z' - z} dz' \right| \leq \left| \frac{1}{2\pi i} \int_{\mathcal{C}_\rho} \frac{\epsilon}{\rho} dz' \right| \leq \frac{1}{2\pi} \frac{\epsilon 2\pi\rho}{\rho} \leq \epsilon, \tag{3.21}$$

and thus this integral vanishes. Inserting this into (3.19) and using (3.20) proofs Cauchy's theorem (3.18).

From this theorem a number of further useful statements can be derived. First, any continous function $\phi(z)$ can be used to create a holomorphic function $f(z)$ by an integration over an arbitrary closed curve $\mathcal{C}$ around some point z as

$$f(z) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{\phi(z')}{z' - z} dz'. \tag{3.22}$$

The existence of this functions follows from the definition of the curve integration. To show that it is holomorphic it is useful to construct its derivative explicitly

$$\frac{df}{dz} = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{\phi(z')}{(z' - z)^2} dz',$$

which follows if differentiation and integration can be exchanged. The proof follows essentially the same steps as before, and will be here omitted. As a consequence, also

$$\frac{d^p f}{dz^p} = \frac{p!}{2\pi i} \int_{\mathcal{C}} \frac{\phi(z')}{(z' - z)^{p+1}} dz', \tag{3.23}$$

which follows by complete induction.

Combining this statement and (3.18) implies that for any holomorphic function its pth derivative is given by

$$\frac{d^p f}{dz^p} = \frac{p!}{2\pi i} \int_{\mathcal{C}} \frac{f(z')}{(z' - z)^{p+1}} dz', \tag{3.24}$$

if the point z is contained in the interior of the curve $\mathcal{C}$, which needs to be entirely within the patch in which $f(z)$ is holomorphic. This therefore implies that a continous function ϕ in (3.22) will reproduce itself, as well as generate all its derivatives by (3.23).

Finally, the so-called theorem of Morera states that holomorphy of $f(z)$ is equivalent to the statement that

$$\int_{\mathcal{C}} f(z)dz = 0$$

for any arbitrary closed curve inside the patch where $f(z)$ is holomorphic. This follows as the curve can then be split arbitrarily into two, defining a continous primitive $F(z)$ of $f(z)$. Because the result is path-independent, $F(z)$ is holomorphic, and thus infinitely often differentiable by (3.24). But then, so is its first derivative, and thus $f(z)$ is holomorphic as well.

Chapter 4

Singularities

It has been visible that holomorphic functions are highly constrained, not only concerning their curve integrals, but also by the fact that they need to be infinitely often differentiable, and their derivatives are linked by Cauchy's theorem and its consequences very tightly to its original form. These information is enough to make very strong statements on series representations of holomorphic functions. In turn, this allows to make powerful statements about the singularity structure of complex functions.

4.1 Series in complex variables

It has already been shown in section 3.3.2 that any series within its radius of convergence defines a holomorphic function, which after section 3.3.4 then has derivatives of arbitrary high order.

As a consequence, any convergent polynomial is a holomorphic function, and its integral over all closed curves in its radius of convergence vanishes. Note that because all uniformly convergent series define a continous function such series can also be used by (3.22) to define a holomorphic function. Furthermore, there are important statements about such polynomials when it comes to the exchange of the (infinite) summation and integration and differentiation.

The first is that for any uniformly convergent series of some holomorphic functions f_k, not necessarily polynomials,

$$\int_{\mathcal{C}} \sum_k f_k dz = \sum_k \int_{\mathcal{C}} f_k dz \tag{4.1}$$

holds, i. e. integration and summation can be freely exchanged. This does not hold generally for series of functions, as counter examples show. (4.1) follows from the fact that any subseries of convergent series can be bounded from a above by an arbitrarily small constant from a large enough index onward, due to the boundedness. As a consequence, for any polynomial

$$\int_{\mathcal{C}} \sum_k a_k (z - z_0)^k dz = \sum_k a_k \int_{\mathcal{C}} (z - z_0)^k dz.$$

Furthermore

$$\frac{df}{dz} = \sum_k \frac{df_k}{dz}.$$

To show that, use (3.24) for $p = 1$. Using then (4.1) and once more using (3.24) to again exchange the order is sufficient. This is, of course, only true if the derivative of the sum is still holomorphic. This can be shown by the theorem of Morera in section 3.3.4 and once more (4.1). Thus, for any polynomial (or series of holomorphic functions) integration and differentiation can be exchanged with the summation.

This implies now the very far-reaching statement that any holomorphic function can be expanded in a Taylor series. To be precise, if the function f is holomorphic in the patch $\mathcal{G}$, then for any $z_0 \in \mathcal{G}$ there exists a Taylor series such that

$$\begin{aligned} f(z) &= \sum_k a_k^{z_0} (z - z_0)^k \\ a_k^{z_0} &= \frac{1}{k!} \frac{d^k f}{dz^k}\bigg|_{z=z_0}, \end{aligned}$$

and the radius of convergence $\rho(z_0)$ obeys $\rho(z_0) > 0$. In particular, this implies that the series expansion is unique for every z_0, and the whole patch $\mathcal{G}$ can be covered by discs of finite radii such that the function can be expanded everywhere. Note that this may require for different values of z_0 different Taylor expansions may be needed.

To proof this, take as the curves $\mathcal{C}_{\rho(z_0)}$ a circle on the edge of the corresponding discs. Then, by (3.23),

$$\begin{aligned} &\sum_k \frac{1}{n!} \frac{df}{dz}\bigg|_{z=z_0} (z - z_0)^k = \sum_k \frac{1}{2\pi i} \left(\int_{\mathcal{C}_{\rho(z_0)}} \frac{f(z')}{(z' - z_0)^{k+1}} dz' \right) (z - z_0)^k \\ = \; & \frac{1}{2\pi i} \int_{\mathcal{C}_{\rho(z_0)}} \frac{f(z')}{(z' - z_0)} dz' = f(z) \end{aligned}$$

where it has been used that

$$\frac{1}{z' - z} = \sum_k \frac{1}{(z' - z_0)^{n+1}} (z - z_0)^n \tag{4.2}$$

is a suitable Taylor series on the curve, as the disc has a finite radius, and thus z_0 is safely inside the interior of the curve, and $z \neq z'$.

4.2 Properties of series and analytic continuation

A practically very interesting statement is that to make sure two series expansion around the same point with the same radius of convergence are identical not only if they have the same coefficients, but it is already sufficient that they agree on an infinite number of points.

Thus, this implies then that they agree on all points within the radius of convergence. As a consequence, if they agree on an infinite number of points it follows that their coefficients have to agree. This can be proven by induction, which allows to establish agreement power by power.

This has a further consequence, the double-series theorem of Weierstraß, which plays an important role when doing expansions in physics. It is possible to define a double series starting with convergent sums

$$f_n(z) = \sum_k a_k^n (z - z_0)^k.$$

All of these are holomorphic. If

$$f(z) = \sum_n f_n(z)$$

is uniformly convergent, it agrees with

$$f(z) = \sum_k \left(\sum_n a_k^n \right) (z - z_0)^k,$$

with the same radius of convergence. Thus, this statement is equivalent to saying that both infinite summations can be exchanged. It is this feature which is relevant in physics, as it allows to chose the way how to expand arbitrarily. This can again be proven by induction.

Taking these statements together with the result of section 4.1 that every holomorphic function can be written as a series, this has a very far reaching consequence, the identity theorem: If two holomorphic function agree on a infinite number of points near a culmination point z_0, then they are identical on the whole patch in which they are holomorphic, and thus the patch is necessarily also identical.

This trivially follows for the radius of convergence of the, necessarily existing, series expansion of both functions around z_0. The non-trivial part is now to extend it to the whole patch, and thus possibly beyond the radius of convergence of the series at z_0.

The proof proceeds as follows. For any point z inside the patch, define a curve $\mathcal{C}_{z_0 \to z}$. Then choose a point along this curve ζ which is still inside the radius of convergence for the expansion at z_0. For the function to be holomorphic in the patch requires that the radius of convergence is non-vanishing if expanding the function at the point ζ, if the original expansion is not valid up to the boundary of the disc. In the latter case nothing would have to be done. Thus, there are now two overlapping discs. An expansion at ζ is now possible. But this expansion will have an infinite number of points on which the expansion in z_0 and ζ agree for both functions, due to identical series expansion at z_0, as there are an infinite number of points within the radius of convergence around ζ overlapping with the first disc. Thus, they need to agree again in the new disc. Repeating this procedure, the whole curve, and ultimately all points in the patch can be covered by overlapping discs, thus establishing the identity of both functions. Building such a chain of discs is also known as analytic continuation. It is particularly useful to extract

information about a function which is originally known only in some patch, but could be evaluated in a different (series) representation elsewhere more easily.

Consider as an example

$$f(z) = \sum z^n,$$

with a radius of a convergence $|z| < 1$. The sum can be explicitly evaluated within the radius of convergence to give

$$f(z) = \frac{1}{1-z}. \tag{4.3}$$

To find an analytic continuation beyond $|z| < 1$, a function is required, which, at least for some small patch for $|z| < 1$ yields (4.3). Since

$$f_c(z) = \frac{1}{1-z} = \frac{3}{5}\frac{1}{1-\frac{3}{5}\left(z+\frac{2}{3}\right)}$$

holds for $|z+2/3| < 5/3$, this function has the required property. This is also immediately seen from its series expansion

$$f_c(z) = \frac{3}{5}\sum\left(\frac{3}{5}\left(z+\frac{2}{3}\right)\right)^n = \sum\left(\frac{3}{5}\right)^{n+1}\left(z+\frac{2}{3}\right)^n.$$

Thus, f_c and f have some range, where they coincide, but extend each other beyond. Thus, f_c analytically continues f beyond $|z| = 1$. In fact, the function (4.3) can in this way be analytically continued to the whole complex plane, except, of course, at the pole. It should be noted that, in a sense, analytic continuation is an implementation of the transition functions of section 2.1.

From this follows immediately that any function, which takes the same value on an infinite set of points around a cumulation point is necessarily the constant function of this value. The reverse is then necessarily also true that any non-constant function must take at least two different values on an infinite number of points around any cumulation point.

Moreover, by the way how derivatives are constructed, this implies that any functions, which agree for the function value as well as for all derivatives in a single point are necessarily identical. This implies local information imply global information. This is one of the reasons why holomorphism is on the one hand a powerful tool, but on the other hand also highly constraining.

It can also be used to guarantee the existence of a series expansion of some functions with a singularity. If a function $g(z)$ can be written as

$$g(z) = \frac{f(z)-a}{(z-z_0)^\kappa} \tag{4.4}$$

where $f(z)$ is a suitable holomorphic function satisfying $f(z_0) = a$, but not constant, and κ is a positive integer. Then $g(z)$ can be expanded in a power series around z_0 with a non-zero radius of convergence as

$$g(z) = \sum a_k (z-z_0)^k.$$

The suitable holomorphic function will need to have a series expansion

$$f(z) = a + \sum_{k \geq \kappa} a_k (z - z_0)^k,$$

and inserting this into (4.4) proofs the claim.

Another unexpected feature of holomorphic functions is that for a non-constant holomorphic function f the function $|f(z)|$ cannot have a (local) maximum. The proof follows essentially from

$$|f(z)| = \left| a_0 + a_\kappa (z - z_0)^\kappa + \mathcal{O}\left((z - z_0)^{\kappa+1}\right) \right|,$$

which will be true for z sufficiently close to z_0. Rewriting this as

$$\left| |a_0| e^{\arg a_0} + |a_\kappa| |(z - z_0)|^\kappa e^{\arg a_\kappa + \kappa \arg(z - z_0)} + \mathcal{O}\left((z - z_0)^{\kappa+1}\right) \right|$$

it follows that there is some z for which

$$\arg(a_\kappa) + \kappa \arg(z - z_0) = \arg a_0,$$

holds, irrespective of the value of $|z - z_0|$. Because of the convergence of the series, there is some z for which the remainder can be bounded by $|a_\kappa||z - z_0|^\kappa/2$. Thus at worst the last term can be negative, and thus

$$|f(z)| \geq \left| e^{\arg a_0} \left(|a_0| + \frac{|a_\kappa||z - z_0|^\kappa}{2} \right) \right| \geq |a_0| = |f(z_0)|,$$

completing the proof. Thus, for any holomorphic function there is at any point always a direction in which its absolute value grows. It was decisive here that the series only depends on z, as this allowed to find such a direction. If it would have depended on z and z^*, this would not have been possible.

4.3 Laurent series

Most of the statements so far required the functions to be holomorphic throughout some patch. However, in many cases in physics functions are only holomorphic on patches, which have holes. The simplest example for such a case are two concentric discs of different radii at some point z_0, with a function $f(z)$ being only holomorphic outside their union and inside the larger disc, i. e. on a ring between their radii $r_1 > r_2$. This relatively special geometry can be relaxed for the larger circle, which can be some patch, which encloses the inner circle at least for a radius of size r_1, and may extend to infinity.

In such a case the preconditions for having a Taylor series no longer holds. It is, however, possible then to expand the holomorphic function into a Laurent series

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k. \tag{4.5}$$

To show this, create two concentric ring paths $\mathcal{C}_1$ and $\mathcal{C}_2$ in the ring with opposite directions, connecting them at two points by bridges and let z lie in one of the half rings. This yields two paths enclosing half rings, one of them containing z, $\mathcal{C}_z$, and the other, $\mathcal{C}$, does not. Because of section 3.3.4 follows

$$\frac{1}{2\pi i}\int\limits_{\mathcal{C}_z}\frac{f(z')}{z'-z}dz' = f(z)$$

and

$$\frac{1}{2\pi i}\int\limits_{\mathcal{C}}\frac{f(z')}{z'-z}dz' = 0.$$

Adding both yields

$$f(z) = \frac{1}{2\pi i}\left(\int\limits_{\mathcal{C}_1} - \int\limits_{\mathcal{C}_2}\right)\frac{f(z')}{z-z'}dz'$$

where for definiteness $\mathcal{C}_2$ is the inner ring. Using twice (4.2), once for z and once for z', yields

$$f(z) = \frac{1}{2\pi i}\sum_{k>0}\left(\int\limits_{\mathcal{C}_1}\frac{f(z')(z-z_0)^k}{(z'-z_0)^{k+1}}dz' - \int\limits_{\mathcal{C}_2}\frac{f(z')(z'-z_0)^k}{(z-z_0)^{k+1}}dz'\right).$$

Defining

$$\begin{aligned} a_{k\geq 0} &= \frac{1}{2\pi i}\int\limits_{\mathcal{C}_1}\frac{f(z')}{(z'-z_0)^{k+1}}dz' \\ a_{k<0} &= -\frac{1}{2\pi i}\int\limits_{\mathcal{C}_2} f(z')(z'-z_0)^k dz' \end{aligned} \tag{4.6}$$

identifies the Laurent series and even provides a constructive way of obtaining the coefficients. Note that, because $f(z)$ is holomorphic inside the ring, the result will necessarily not depend on the details of the path.

Obviously, the subseries with non-negative exponents form a usual holomorphic function, which can be written as

$$g(z) = \frac{1}{2\pi i}\int\limits_{\mathcal{C}_1}\frac{f(z')}{z'-z}dz',$$

very much in the same way as in section 3.3.4 already discussed. It is, however, quite useful to notice that also the series with negative exponents can be rewritten. For this, perform the substitution

$$\zeta = \frac{1}{z-z_0}.$$

This substitution is a first example of a conformal map, to be discussed in more detail in section 4.12. This creates a power series in ζ. It is, by construction, convergent in the ring. Because

$$\phi(\zeta) = f(z) = f\left(z_0 + \frac{1}{\zeta}\right)$$

ϕ is then necessarily holomorphic on the ring. Especially,

$$\sum_{k>0} a_{-k}\zeta^k = \frac{1}{2\pi i}\int\limits_{\mathcal{C}'} \frac{\phi(\zeta')}{\zeta' - \zeta} d\zeta$$

defines a holomorphic function, where the curve is defined by $|\zeta'| = |1/(\zeta - z_0)| = 1/r_2$ where r_2 was the radius of the curve $\mathcal{C}_2$. If that one was not circular, it is necessary to modify the condition correspondingly. Because of the possibility to rewrite both parts of the series as holomorphic functions it follows that the Laurent series of a function is unique.

As an example, consider the function

$$f(z) = \frac{1}{(z-1)(z-2)},$$

which is holomorphic for $1 < |z| < 2$. Using $z_0 = 0$ and the, by now, standard approach of rewriting

$$\frac{1}{1-z} = \frac{1}{z}\frac{1}{1-\frac{1}{z}} = \frac{1}{z}\sum_{k>0} z^{-k}$$

yields immediately the Laurent series

$$f(z) = -\frac{1}{2} - \sum_{k>0}\left(\frac{z^k}{2^{k+1}} + \frac{1}{z^k}\right).$$

This says nothing about the expansion of $f(z)$ inside or outside the ring. Inside, it is a normal holomorphic function, needing only a Taylor series, while outside it is on the outside area again a holomorphic function, as discussed in section 3.3.4 for functions with a hole in their patch of holomorphy.

Note, however, that the Laurent series changes when a different point z_0 is used for the expansion, though it of course agrees in its final value at every point. Take, e. g. $z_0 = i$ for the same function. This yields as Laurent series in the same ring

$$f(z) = -\frac{1}{2-i} - \sum_{k>0}\left(\frac{(z-i)^k}{(2-i)^{k+1}} + \frac{(1-i)^{k-1}}{(z-i)^k}\right).$$

Though some resemblance is visible, it is by far not obvious that this will yield the same results. Note, however, that here the condition is $\sqrt{2} < |z-i| < \sqrt{5}$ to stay inside the ring.

4.4 Isolated poles

If a function $f(z)$ is holomorphic within a certain patch except at some point z_0, then z_0 is called an isolated singularity. If not, that is the function $f(z)$ is holomorphic everywhere, it is called an entire function.

Because the holomorphic function can be expanded in a Laurent series everywhere else in the patch this necessarily implies that at least one a_k for $k < 0$ needs to be non-zero. Isolated singularities are called a pole of order k if there is some $k > 0$ such that $a_{-i} = 0$ for all $i > k$, i. e. the Laurent series does not extend to negative infinity but only until $-k$. If there are an infinite number of $a_{k<0} \neq 0$ it is called an essential singularity. Note that no order can be assigned to an essential singularity because otherwise only a finite number of the $a_{k<0}$ would not vanish.

That this is close to the concept of singularities for real functions follows from the fact that $|f(z)|$ will become arbitrarily large when approaching z_0. More specifically, for any γ there is some z with $|z - z_0| < \delta$ such that $|f(z)| > \gamma$, and δ is the radius of a disc centered at z_0 in which $f(z)$ is holomorphic. Note that this does not necessarily imply that $|f(z)|$ is monotonically increasing from all directions towards z_0. This can be proven using the Laurent series, as some of its terms will necessarily diverge faster than all terms of lower order and the absolute value will bound the function by it.

Consider as an example of these concepts a rational function

$$f(z) = \frac{p_m(z)}{q_n(z)},$$

where p_m and q_n are polynomials of order m and n, respectively. This function has potentially isolated poles at the zeros of q_n. Such a function, a function with a finite number of isolated poles, is called a meromorphic function. The name is also given to any function which is holomorphic except at isolated poles.

If z_0 is a pole of order $k_0 \leq n$, then

$$q_n(z) = (z - z_0)^{k_0} r_{n-k_0}(z)$$

where r_{n-k_0} is a polynomial without zero in z_0 of order $n - k_0$. Then p_m / r_{n-k_0} is holomorphic in a patch around z_0 in which no further zeros of q_n are located, which by assumption exists, and can thus be rewritten as a Taylor series with some new coefficients c_k, and thus

$$f(z) = \sum_{k=-k_0} c_{k+k_0} (z - z_0)^k,$$

making the existence of a pole of order k_0 explicitly in this form.

Essential singularities are a quite different story. Probably the most (in)famous essential singularity in physics is that of the function $e^{1/z}$ at $z_0 = 0$, which has the Laurent series

$$e^{\frac{1}{z}} = \sum_{k \leq 0} \frac{1}{k! z^k},$$

which therefore has no terms with positive coefficients, but is holomorphic otherwise.

A surprising, but technically involved to prove, consequence for essential singularities is the theorem of Casorati and Weierstraß. It states that if a function has an essential singularity at z_0, the function $f(z)$ approaches every possible number c arbitrarily close within a disc around z_0. Thus, for any c there is a z with $|z-z_0| < \delta$ such that $|f(z)-c| < \epsilon$ arbitrarily small.

The fact that holomorphic functions with isolated poles can be expanded in Taylor series times a singular prefactor has far-reaching consequences in physics. Especially, a perturbative series will converge if, up to isolated poles, the object to be expanded is a holomorphic function. If not, any perturbative expansion may at most be a quantitative good approximation, but will fail qualitatively to reproduce the original object, which thus requires non-perturbative methods. Due to the appearance of essential singularities in many quantities in physics, this implies that perturbative series are often at best an approximation, and can be completely misleading. Analyzing the holomorphy of a quantity in question, especially if statements using the Cauchy-Riemann differential equation (3.3-3.4) can be made, is therefore often an important step.

4.5 Residua

Given a function $f(z)$ with an isolated singularity of order 1 at z_0, the definition (4.6) implies for a closed curve $\mathcal{C}$ within a suitable patch of holomorphy of the function enclosing the isolated singularity

$$r_f(z_0) = a_{-1} = \frac{1}{2\pi i}\int\limits_{\mathcal{C}} f(z)dz = (z-z_0)f(z)|_{z=z_0}\,.$$

In this context, a_{-1} is called the residuum $r_f(z_0)$ of $f(z)$ at z_0.

This can be generalized. Consider some patch in which a function is holomorphic except at a set z_i of isolated poles of order one. Given a closed curve $\mathcal{C}$ enclosing all poles it follows that

$$\frac{1}{2\pi i}\int\limits_{\mathcal{C}} f(z)dz = \frac{1}{2\pi i}\sum_i \int\limits_{\mathcal{C}_i} f(z)dz = \sum_i r_f(z_i), \tag{4.7}$$

where the curves $\mathcal{C}_i$ enclose only one of the poles. This follows directly from combining (4.6) and section 3.3.4. Note that the details of the paths do not matter. However, to determine the residua $f(z_i)$ it is necessary to obtain either the Laurent series for each of the poles, or at least compute (4.6).

There is a particular useful case if $f(z)$ is holomorphic in the whole complex plane, except for isolated poles, and decays 'quickly' enough when $|z| \to \infty$ such that for a curve $\mathcal{C}_\infty$ running around the complex plane 'at infinity'

$$\int\limits_{\mathcal{C}_\infty} f(z)dz = 0$$

holds. This implies that the sum of residua is necessarily zero. This allows to determine one of the residua by the other ones. The path independence also allows another trick,

often useful in physics. If a function is holomorphic on the whole complex plane any integral over a closed curve vanishes. If there are isolated poles, the integral over closed curves can still be deformed arbitrarily, as long as an isolated pole does not pass through it, and it is fully determined by (4.7). This often allows to solve an integral by deforming its path such that it becomes 'simple' enough to be done explicitly.

Consider as an example again

$$f(z) = \frac{1}{(z-1)(z-2)}$$

which is holomorphic except for the isolated poles at $z = 1$ and $z = 2$. Its Laurent series at both poles can be obtained by rewriting,

$$\begin{aligned} f(z) &= \frac{1}{z-1}\frac{-1}{1-(z-1)} = -\sum_{k=-1}(z-1)^k \\ f(z) &= \frac{1}{z-2}\frac{1}{1+(z-2)} = \sum_{k=-1}(-1)^{k+1}(z-2)^k. \end{aligned}$$

Thus, the residua are $r_f(1) = -1$ and $r_f(2) = 1$. Any curve enclosing both poles thus vanish, as $r_f(1) + r_f(2) = 0$. This could have conversely be predicting by noticing that for $|z| \to \infty$ $f(z) \sim 1/z^2$, and thus the integral over $\mathcal{C}_\infty$ necessarily vanishes.

4.6 Zeros of polynomials

While the statement (4.7) is useful to perform integrals, it can also be used to derive other helpful results. Especially, it is useful to make statements about the zeros of functions.

Given any holomorphic function $f(z)$. If it has zeros z_i of order k_i within its patch of holomorphy then for a curve $\mathcal{C}$ enclosing them all follows

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{\frac{df(z)}{dz}}{f(z)} dz = \sum_i k_i. \tag{4.8}$$

To see this, note that around any of the z_i the function $f(z)$ can be written as

$$f(z) = (z - z_i)^{k_i} q_i(z)$$

where the q_i are polynomials which are non-vanishing at the z_i. Then the integral kernel in the vicinity of the z_i can be rewritten as

$$\frac{\frac{df(z)}{dz}}{f(z)} = \frac{k_i}{z - z_i} + \frac{\frac{dq_i(z)}{dz}}{q_i(z)} \tag{4.9}$$

which therefore has an isolated pole of order one at z_i by construction. Thus, the integrand has exactly the form suitable for (4.7) with residua k_i. Thus (4.8) follows.

The same argument can be done to show that if the function has in addition poles of order κ_i within the enclosing curve then

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \frac{\frac{df(z)}{dz}}{f(z)}dz = \sum_i k_i - \sum_i \kappa_i.$$

Especially, in the absence of zeros, this counts the sum of the order of the poles.

This approach also allows to proof that any polynomial $f(z)$ of a complex variable has as many zeros as its order is, counting multiplicities. Any polynomial of finite order is necessarily holomorphic in the whole complex plane. For any zero with $|z_i| > 1$, it follows that

$$z_i = -\sum_{k=0}^{n-1} \frac{a_k}{a_n z_i^{n-1-k}}$$

by solving the condition of being a zero. This implies

$$|z_i| \leq \sum_{k=0}^{n-1} \left|\frac{a_k}{a_n z_i^{n-1-k}}\right| \leq \sum_{k=0}^{n-1} \left|\frac{a_k}{a_n}\right|$$

and thus all zeros can be enclosed by a curve of sufficiently large radius, which can be set to one if all satisfy $|z_i| < 1$. Since it is a finite polynomial, it is holomorphic in the whole complex plane. Thus, it can be expanded for every zero around the same point[1] $z_0 = 0$. Then use

$$\begin{aligned} f(z) &= z^n q(z) \\ \frac{df}{dz} &= nz^{n-1}q(z) + z^n \frac{dq(z)}{dz} \\ \frac{\frac{df(z)}{dz}}{f(z)} &= \frac{nz^{n-1}q(z) + z^n \frac{dq(z)}{dz}}{z^n q(z)} = \frac{n}{z} + \frac{\frac{dq(z)}{dz}}{q(z)} \end{aligned}$$

where $q(z)$ is some polynomial. The second term has no poles close to zero, and can thus be expanded in a Taylor series around zero. But then the curve can be deformed sufficiently to use (4.8), yielding the claim.

This also implies that the number of zeros of a real polynomial is bounded by its order. After all, it can always be extended to the complex plane, and then will have at most as many zeros by the proof beforehand. However, no such simple statements exist to find the number of real zeros, unfortunately. Real polynomials are harder than complex polynomials.

4.7 Cuts and Riemann sheets

A particular specialty arises in cases where singularities are not isolated. To discuss them, it is useful to introduce the concept of Riemann sheets. At first sight, Riemann sheets are

[1] If $z_0 = 0$ should be a zero this can always be remedied by a shift of z.

just copies of the complex plane. In fact, they are a bookkeeping device. It is first useful to address how they serve for multivalued function.

Consider the nth root of a complex number z of unit magnitude $|z| = 1$,

$$z^{\frac{1}{n}} = e^{\frac{i \arg z}{n}} = e^{\frac{i(\arg z + 2k\pi)}{n}}$$

where k is an integer number between 0 and $n-1$. Thus, there are n possible solutions. The idea of Riemann sheets is now that k labels the Riemann sheet, and thus the root has a unique solution on each Riemann sheet. Considering this now as a function of $\arg z + 2k\pi$, this can be viewed as a function which performs a full circle on a given Riemann sheet, then jumps to the next Riemann sheet and so on until after the nth revolution the original Riemann sheet is reached again. This creates a new space, on which the nth root has actually a unique solution. This space can be considered to either have two continous and one discrete direction and is flat, or just a continous space but not flat. In the later case, the total set of Riemann sheets is the Riemann surface. Only on the Riemann surface the nth root becomes a single-valued, and thus invertible, function.

The actual geometric structure of the Riemann surface leads to the idea of a cut. Consider for simplicity the 2nd root of a general complex number z. The result can be parametrized as

$$\sqrt{z} = +\sqrt{|z|}\left(\cos(2t) + i\sin(2t)\right).$$

This is hence a different parametrization than the usual real part and imaginary part or absolute value and argument. It is implicitly defined that at $t = \pi$ the solution changes the Riemann sheet. Thus, the Riemann surface is parametrized by a radius and an angle t, ranging from 0 to 2π, but each sheet is covered by only one of the half-intervals. The Riemann surface can thus be considered a manifold, but with a highly non-trivial geometric structure.

In fact, as this happens for all lengths $|z|$, this looks like the first Riemann sheet is cut along the negative real axis, and the second Riemann sheet is cut as well, and glued to it such that $\sqrt{z}$ continuously moves on to the second sheet. It then moves again a full circle on the second Riemann sheet. The other edges of the Riemann sheet are also glued, such that the function can then move 'down' again to the first Riemann sheet when t reaches $3\pi/2$ and complete the path again to the real axis of the first sheet at $t = 2\pi$. While this cannot be visualized in three dimensions without the intersection of the Riemann sheets, no such intersection actually exists. On this Riemann surface the function is single-valued, as desired. Note that the Riemann surface depends on the function. In contrast to the complex plane it is not an entity existing in the same way for all functions. Rather, this should be understood as that for any multi-valued function of a complex variable a manifold exists, in which it is single-valued.

The edges where the Riemann sheets are glued together to form the Riemann surface is called the cut of the function. A cut has necessarily a starting point and an endpoint, in the present case at $z = 0$ and at $z = \infty$. Functions can have multiple cuts. The points where a cut starts or ends are also called branching points. It should be noted that while the Riemann surface is unique, its disassembling in terms of sheets and cuts is not. The

cut can be put everywhere, provided the function is multi-valued along it to allow for switching Riemann sheets.

To see how this works, consider the function $f(z) = 1 + \epsilon \exp(i \arg(z))$ with ϵ infinitesimally small. It then follows that

$$\begin{aligned} |f| &\approx 1 + \epsilon \cos \arg z \\ \arg f &\approx \epsilon \sin \arg z. \end{aligned}$$

Making this now a multivalued function by taking the root yields

$$\begin{aligned} \left|\sqrt{f}\right| &= \sqrt{|f|} \\ \arg \sqrt{f} &= \frac{1}{2} \arg f \approx \frac{\epsilon}{2} \sin \arg z. \end{aligned}$$

It is visible how the argument of $\sqrt{f}$ goes twice around when the argument of z is going around once. The branching point is indeed at $z = 0$ and at $z = \infty$.

For practical purposes, it is useful to catalog the properties of a couple of standard functions. Arguably most important is

$$f(z) = (z - z_0)^{\frac{m}{n}}.$$

This function has n Riemann sheets with branching points starting at $z = z_0$ to $z = \infty$. If the exponent would be irrational, the number of Riemann sheets becomes infinite. As an example where the second branching point is not at infinity, consider

$$f(z) = \sqrt{z^2 - 1} = \sqrt{(z+1)(z-1)}$$

In this case the cut is between the points $z = \pm 1$. If $|z| > 1$ the function is single-valued again, as the sign always factors out. Note that the cuts are really arbitrary positioned, as this is only used to introduce a tool to organize the multivaluedness. Thus, while it is often standard that the cut runs from $z = -1$ to $z = 1$ along the real line, nothing prevents one to have it run along a curve from $z = -1$ to $i\infty$ and then back to $z = 1$. Choosing a convenient cut is often key for making calculations accessible. Note that quite often it is said colloquial that the path is deformed, like in the case of the curve-independent integration in chapter 3, but something different happens here.

A particularly annoying case of a cut is the one of the logarithm (3.8). It has a cut starting at zero, running to negative infinity, with a countable infinite number of sheets. Here the additional complication arises that one of the branching points is simultaneously also a singularity. Thus, forming curves following along the Riemann surface possibly encompass this singularity multiple times, and this has to be taken into account. Trigonometric functions behave similarly. This can be seen explicitly by considering their inverse functions, e. g.

$$\sin^{-1} z = -i \ln\left(iz \pm \sqrt{1 - z^2}\right) = \arg\left(iz \pm \sqrt{1 - z^2}\right) + 2k\pi - i \ln\left|iz \pm \sqrt{1 - z^2}\right|.$$

The same multivaluedness with integer k as for the logarithm is visible. Thus, there are two cuts, starting from $z = \pm 1$ and running to $\pm\infty$.

4.8 Cuts and integrals

It is in physics often useful to determine what happens when integrating a function with a cut. Consider a holomorphic function $f(z)$, except for a cut between two points z_1 and z_2. Given any curve $\mathcal{C}$, which not encloses the cut, it follows that

$$\int_{\mathcal{C}} dz f(z) = 0.$$

Deform now this curve such that there are two curves. One, $\mathcal{C}_1$ encloses the cut at some distance, and the second one $\mathcal{C}_2$ runs along the cut z_c with an extension ϵ. Both curves are connected by a line. This joining is traversed once in each direction, and thus will not contribute to the integral. Thus

$$\int_{\mathcal{C}} dz f(z) = \int_{\mathcal{C}_1} dz f(z) + \int_{z_1}^{z_2} dz f(z + i\epsilon) + \int_{z_2}^{z_1} dz f(z - i\epsilon) + \mathcal{O}(\epsilon)$$

and the remainder term are the pieces needed for connecting the two lines along the cut. Taking the limit $\epsilon \to 0$ yields

$$\int_{\mathcal{C}_1} dz f(z) = -\lim_{\epsilon \to 0} \left(\int_{z_1}^{z_2} dz f(z + i\epsilon) + \int_{z_2}^{z_1} dz f(z - i\epsilon) \right). \tag{4.10}$$

Without cut, the right-hand side would be zero, because the values above and below the cut would be arbitrarily close, due to continuity. Because of the cut, this is no longer true, as these values can be arbitrarily different. Hence, the integral along a curve of a holomorphic function enclosing a cut depends only on the discontinuity across the cut. This quantity is therefore often formally denoted as $\mathrm{disc} f$,

$$\int_{\mathcal{C}_1} dz f(z) = -\lim_{\epsilon \to 0} \int_{z_1}^{z_2} dz \left(f(z + i\epsilon) - f(z - i\epsilon) \right) = -\int_{z_1}^{z_2} dz \left(\mathrm{disc} f(z) \right),$$

where one should be wary in the last step, as it may be necessary to first perform the integral and then take the limit.

Combining the results on residua from section 4.5 and the present insights implies that any closed integral of a holomorphic function is either zero or entirely given by the residua and the discontinuities enclosed by the curve. Conversely, this implies any closed curve can be deformed arbitrarily, as long as the deformation does not pass through singularities or cuts. Quite useful, this implies that it can also be deformed to infinity, which simplifies many actual calculations in practice.

As the determination of cuts and discontinuities is not simple, and to emphasize the use, consider the following example. The integral

$$\int_a^\infty \frac{(x-a)^c}{x^2 + b^2} dx,$$

is typical for many cases in physics, with x being a real variable, and with $0 < c < 1$. The following will show how the use of cuts and poles allows to calculate it. It has poles at $z = \pm ib$ and a cut starting at a. This cut extend to a branch point at negative infinity, and the cut will be put for the moment along the negative axis. Thus, the cut starts to the left of the integration contour. The discontinuity can be obtained by considering

$$\begin{aligned}\lim_{\epsilon\to 0}(x+i\epsilon-a)^c &= |x-a|^c \\ \lim_{\epsilon\to 0}(x-i\epsilon-a)^c &= |x-a|^c e^{2c\pi i}\end{aligned} \tag{4.11}$$

where the Riemann sheets have been arranged in a particular way by choice.

To determine the integral, a curve $\mathcal{C}$ is introduced, which makes the curve closed and contains the poles, and closes at infinity. There, the function falls off quickly enough. This then does not include the cut but only the poles, and thus by the theorem of residua yields

$$\int\limits_{\mathcal{C}} dx \frac{(x-a)^c}{x^2+b^2} = \frac{\pi}{b}((ib-a)^c - (-ib-a)^c). \tag{4.12}$$

At infinity, the function drops to zero. Thus, it needs to be also true by direct evaluation is possible by using (4.10) that

$$\int\limits_{\mathcal{C}} dx \frac{(x-a)^c}{x^2+b^2} = \int\limits_a^\infty dx \left(\frac{|x-a|^c}{x^2+b^2} - \frac{|x-a|^c e^{2c\pi i}}{x^2+b^2}\right) = \left(1-e^{2c\pi i}\right)\int\limits_a^\infty dx \frac{|x-a|^c}{x^2+b^2}. \tag{4.13}$$

By comparison of (4.12) and (4.13) this finally yields

$$\int\limits_x^a \frac{(x-a)^c}{x^2+b^2} = \frac{\pi}{b\left(1-e^{2c\pi i}\right)}((ib-a)^c - (-ib-a)^c).$$

Thus, here a number of intricate steps was necessary to perform the total integral. However, it was possible without finding the primitive. Since the latter is often a major problem, this emphasises how useful such investigations are, if the function in question is of a suitable type. As especially in perturbative expansions integrals are very often holomorphic with cuts and poles, this is surprisingly frequent in physics.

4.9 Dispersion relations

A special subclass of functions appearing in physics are so called real holomorphic functions, which satisfy

$$f(z) = f(z^*)^*, \tag{4.14}$$

and it is required that at least for some interval on the real axis the function is real. Examples are especially scattering amplitudes.

Indeed, because of (4.14), it follows

$$f(x+i\epsilon) - f(x-i\epsilon) = f(x+i\epsilon) - f(x+i\epsilon)^* = 2i\Im f(x+i\epsilon).$$

Thus, a real holomorphic function is either real for real argument or has a cut with a discontinuity of $2i\Im f(x+i\epsilon) = 2i\Im f(x-i\epsilon)^*$. Note that by taking the limit of $\epsilon \to 0$ this implies that the discontinuity at the cut is really just the imaginary part.

This leads to an interesting relation for 'typical' scattering amplitudes in quantum physics. These are often real holomorphic, but do not have further singularities, and decay quickly towards infinity. They have a single cut on the real axis from a to b, either being possibly infinity. Using the same setup as for deriving (4.10), i. e. two curves moving in counterclockwise direction around the cut, yields first that

$$f(w) = \frac{1}{2\pi i}\left(\int\limits_{\mathcal{C}_1} dz + \int\limits_{\mathcal{C}_2} dz\right)\frac{f(z)}{z-w}.$$

If w is not on the cut, this yields using (4.10)

$$f(w) = \frac{1}{2\pi i}\int\limits_a^b dx\left(\frac{f(x+i\epsilon}{x+i\epsilon-w} - \frac{f(x-i\epsilon}{x-i\epsilon-w}\right) = \frac{1}{\pi}\int\limits_a^b dx \lim_{\epsilon\to 0}\frac{\Im f(x+i\epsilon)}{x-w}.$$

This is especially true even if $\Im w = 0$, and thus establishes a relation on the real axis outside the cut between the real part and the imaginary part of f on the real axis. Such a relation is called dispersion relation or Kramers-Kronig relation. It is a decisive step in deriving the optical theorem.

4.10 Principal part integration

In a similar manner also a relation can be established for poles on the real axis. Consider an otherwise holomorphic function $f(z)$ with a single isolated pole on the real axis, e. g. $1/z$. Considering the integral

$$\int\limits_{-\infty}^{\infty} dz f(z)$$

the question arises, what is its value, even though formally it does not exist. In the case of $1/z$ the answer seems to be just zero, as the same negative value appears to arise for any positive value. This can be formalized using the so-called principal part integral. The principal part integral is defined as

$$P\int\limits_{-\infty}^{\infty} dz f(z) = \frac{1}{2}\lim_{\epsilon\to 0}\left(\int\limits_{-\infty+i\epsilon}^{\infty+i\epsilon} + \int\limits_{-\infty-i\epsilon}^{\infty-i\epsilon}\right) dz f(z),$$

i. e. it averages between the integrals taken over the curves infinitesimally displaced above and below the real axis. Because of the infinitesimal displacement the integral will be as convergent towards infinity as the original one.

Because of the theorems of residua this yields for a single enclosed pole at z_0

$$\lim_{\epsilon\to 0}\left(\int_{\infty+i\epsilon}^{\infty+i\epsilon} - \int_{\infty-i\epsilon}^{\infty-i\epsilon}\right) dz f(z) = 2\pi i \mathrm{res}(f(z_0)),$$

i. e. proportional to the residuum of f at z_0. This implies for the principal part integral

$$P\int_{\infty}^{\infty} dz f(z) = \lim_{\epsilon\to 0}\int_{\infty\pm i\epsilon}^{\infty\pm i\epsilon} dz f(z) \pm i\pi \mathrm{res}(f(z_0)).$$

which gives rise to the definition

$$\frac{1}{z - z_0 \pm i\epsilon} = P\frac{1}{z-z_0} \pm i\delta(z - z_0),$$

which, like the Dirac δ-function only makes sense as statement when performing an integral, applied to a (holomorphic) function.

This can be generalized for situations with multiple poles.

4.11 Γ function

As a general example of the concepts developed, and of particular use in physics, in the following the (generalized) Γ function will be discussed.

The Γ function is the generalization of the factorial $n!$ of an integer number. To define it, note that

$$\partial_\alpha^n \int_0^\infty e^{-\alpha t} dt = \partial_\alpha^n \frac{1}{\alpha} = \frac{(-1)^n n!}{\alpha^{n+1}} = \int_0^\infty (-1)^n t^n e^{-\alpha t} dt.$$

Thus, at $\alpha = 1$ this is an integral representation of $n!$. This is used to extend the definition of the faculty operation to arbitrary complex numbers as the Γ function by

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \tag{4.15}$$

for $\Re(z) > 0$. For integer numbers or zero this definition reverts automatically to the normal faculty,

$$\Gamma(n+1) = n!$$

though it does not work for $\Re(z) \leq 0$. This definition implies by partial integration

$$\Gamma(z+1) = z\Gamma(z) \tag{4.16}$$

as a generalization of $n! = n \times (n-1)!$.

The so defined function is actually a holomorphic function. This can be seen by differentiating (4.16) yielding

$$\frac{d\Gamma(z+1)}{dz} = \Gamma(z) + z\frac{d\Gamma(z)}{z}$$

Because the function can be bounded using its integral form, that is guaranteed. Hence, the derivative at least exists. The absence of singularities can be seen in the following way. It is possible to rewrite $\Gamma(z)$ as

$$\Gamma(z) = \frac{\Gamma(z+n)}{z(z+1)...(z+n-1)}, \tag{4.17}$$

and by the integral definition the numerator is an ordinary finite function if $\Re(z+n) > 0$. Hence, the only singularities, which exists, are single poles at z being zero or negative integers. It is thus the decisive possibility (4.16) to continue the Γ function always to a real argument which allows to determine the pole structure. This is an example of an analytic continuation. The definition of the Γ function is only valid at $\Re z > 0$. But (4.17) is now a form valid, and unique, in the whole complex plane. Note that as a consequence $\Gamma(z+1)^{-1}$ is an entire function, as it has no poles, but zeros at negative integer values of its arguments. It is moreover useful to note that

$$\frac{d\ln\Gamma(z)}{dz} = -\frac{1}{z} - \left.\frac{d\Gamma(z)}{\Gamma(z)dz}\right|_{z=1} + \sum_{k=1}^{\infty}\left(\frac{1}{k} - \frac{1}{z+k}\right).$$

The second term equals γ, the Euler constant, which therefore can be defined in this way. Thus, the logarithmic derivative of the Γ function is not meromorphic, as it has an infinite number of isolated poles at negative integers and zero.

4.12 Conformal maps

A conformal map is, in brief, a mapping of a variable such that the whole space is mapped into the whole space, but maintaining some structures invariant. Such conformal maps play a role in physics in the context of conformal symmetries, which are extremely constraining and therefore allow often exact results. For complex functions, conformal maps map the complex plane into the complex plane. An example has already been encountered in section 4.3. These maps are often useful in simplifying problems, especially in boundary-value problems.

Because the complex numbers are a plane, differentials dz are not numbers, but do have a direction. Consider a holomorphic map $w = f(z)$, i. e. $f(z)$ is holomorphic everywhere. At a given point z, differentials can be formed of the same magnitude but differing directions, say $dz_1 = \epsilon\exp(i\alpha_1)$ and $dz_2 = \epsilon\exp(i\alpha_2)$. Then it follows for the differentials of the map

$$\frac{dz_1}{dz_2} = \frac{\frac{df(z)}{dz}dz_1}{\frac{df(z)}{dz}dz_2} = \frac{dw_1}{dw_2},$$

where it was used that for a holomorphic function the derivative does not depend on the direction. Thus, the relative angel does not change by a holomorphic map, the map is angle-preserving. Note that it is possible to have also just mapping a patch into a patch.

Primitive maps, which do have the property to map the complex plane on itself are usually build from chaining a number of elementary transformations. The most relevant elementary ones are translations and rescalings, also called affine transformations,

$$w = az + b,$$

rotations

$$w = ze^{i\alpha}$$

with α real, inversion

$$w = \frac{1}{z}, \tag{4.18}$$

compression and extension

$$w = (z - a)^{\alpha}$$

with a real α, noting that this may require to cut the complex plane, and the Möbius transformation

$$w = \frac{az + b}{cz + d},$$

with $ad \neq bc$. The latter is actually already a combination of translation and inversion, which maps lines and circles into lines and circles.

To see that such a map may be highly non-trivial, consider the special Möbius transformation

$$w = \frac{a + z}{a - z}$$

which maps the imaginary axis on the unit circle. It is actually invertible by

$$z = a\frac{w - 1}{w + 1}$$

which may have not been anticipated by its action. However, this is an important feature of a map. In fact, for the inverse function to be differentiable it follows

$$\frac{df^{-1}(w)}{dw} = \frac{1}{\frac{df(z)}{dz}},$$

which implies that the derivative of the map may not have zeros to ensure the existence of the derivative.

In such maps the point zero and the circle at infinity often play an important role, e. g. in the inversion (4.18). It is often useful to define that zero is mapped to infinity, and because of the ill-defined nature of the direction of zero to the whole circle at infinity, and vice versa. The same is true for other singular points.

An interesting observation for the usefulness of maps is the Riemann mapping theorem. It states that for any simply-connected open patch, which is not the whole complex plane, there exists a holomorphic map $f(z)$ with $d_z f(z) \neq 0$ such that the patch is mapped to the unit disc, i. e. $|f(z)| < 1$. While the proof is somewhat involved, this insight is again technically very relevant.

4.13 Functions of multiple complex variables

In principle, a lot stays the same when moving to multiple variables. However, one important point is that directions of approach played already a central role. Generalizing the definition of what holomorphy is requires to translate the concept of direction independence to a higher-dimensional space. In particular, curves then no longer enclose patches. Since the lowest-dimensional system is already real four-dimensional the example of Klein's bottle already shows that then also orientability also becomes non-trivial.

Putting a precise definition on what a patch is and what the necessary discs are in the various steps are the most challenging parts of the enlargement of dimensions. It is relatively straightforward to construct conditions, which are sufficient to allow for the use of results on holomorphic functions, e. g. the generalization of Cauchy's theorem. The latter then, e. g., covers the possibility to have poles in different variables. However, phrasing minimal conditions for them to apply is difficult.

These subtleties arise rarely in physics, and thus in most cases it is sufficient to get away with requiring the patches to be replaced by convex volumes. Thus, these subtleties will not be treated here.

Chapter 5

Abstract groups

Groups are quite generic structures. A group is essentially any structure which fulfills a numbers of properties, and can be far away from anything one usually imagines under the name of group, e. g. the rotation group encountered in mechanics. It is therefore worthwhile to start out with a suitable abstract definition of what groups are, and what are generic traits of groups.

5.1 Groups

A group G is a structure on a set $\mathcal{G}$ of elements g_i. This set can be finite or infinite, leading to a finite or infinite group. The number of (distinct) elements is called the order of the group . Part of the structure must also be some map $\circ$ to combine two elements of the set. This structure is a group if it has has the following properties. Or, any structure fulfilling the following properties is a group, though it may also fulfill many more properties as well. These properties are called group axioms:

- The combination of any two elements is within the set of elements, $g_1 \circ g_2 = g_3$ with $g_i \in \mathcal{G}$ (Closure)
- The combination is associative, $(g_1 \circ g_2) \circ g_3 = g_1 \circ (g_2 \circ g_3)$
- There exists a (unique) element e with $e \circ g = g \circ e = g$, i. e. an identity element or unit element
- For every g there exists an element g^{-1} with $g \circ g^{-1} = g^{-1} \circ g = e$, i. e. an inverse element or just inverse

Note that the combination does not need to be commutative, i. e. $g_1 \circ g_2$ may or may not be the same as $g_2 \circ g_1$. If it is the same for all possible combinations of elements, the combination is commutative, and the group is called Abelian. If not, it is called non-Abelian.

5.2 Creating a standard example

Many of the concepts of group theory are quite abstract. It is therefore useful to create standard examples, which can be used to illustrate in a fixed setting all the concepts in a very explicit way. Of course, other examples will be used to highlight additional aspects.

The standard example to be used during the first part will be the rotations in two and three dimensions, $\mathcal{R}_2$ and $\mathcal{R}_3$, respectively. The elements are then just rotations. These will be characterized by rotation angles, one in two dimensions and three in three dimensions, where for now it does not matter what the precise rotation axes are. The combination of elements is to perform two rotations after each other.

These two sets and combinations are groups. Geometrically, a sequence of rotations remains a rotations, and therefore the set is closed under the combination. Rotations are also associative, as the combination of two rotations is again a uniquely defined rotation. The unit element is just no rotation at all, and the inverse element is the rotation backwards, which compensates a rotation.

From geometry, it follows that two rotations in two dimensions commute, and the group is Abelian. In three dimensions, however, rotations do not commute, and the group is non-Abelian.

Both groups have an infinite number of elements. To obtain a discrete rotation group, the simplest version is to use a set of rotation about a fixed angle (or angles in three dimensions), $\mathcal{R}_n^\alpha$, and its multiples, with the angle α being a rational factor of 2π. For the following, it is convenient to chose this angle to be $\pi/2$, providing the groups of discrete rotations in two and three dimensions. Due to the factor requirement, any combination of these group elements is again an element of the group, and the remainder of the group axioms are fulfilled in the same way. In this case, the group elements for the two-dimensional case can be labeled by a single integer n, $g(n)$, which counts the rotation angle α, $n\alpha$, and n runs from 0, the identity, to $2\pi/\alpha - 1$, the maximum rotation. In the three-dimensional case, this requires three indices.

Note that in contrast to what is usually done in linear algebra at no point any kind of matrix was written down, as such a realization - latter to be called a representation - is not necessary to describe the group nature of rotations.

5.3 Subgroups

If in a group G there exists a subset of the elements $\mathcal{G}_s \subset \mathcal{G}$ such that with the same combination this subset also forms a group, this is called a subgroup. A group can have multiple subgroups. As long as the map $\circ$ is not restricted in any way, the identity element and the corresponding inverse elements of the subset need to remain the same, and thus are necessarily part of the subgroups. If the map is restricted in any wa´y, this may change[1], though such cases have little relevance for physics. There are always the trivial subgroups of just the identity element and the whole group itself.

[1] E. g. for a two-dimensional matrix group a sub-group with a different identity element is obtained if simultaneously a restriction to a one-dimensional subspace is implemented.

A particularly interesting case occurs if a subgroup of a non-Abelian group exists, for which all elements commute with all elements under the combination, i. e.

$$g_s \in \mathcal{G}_s \text{ and } g \in \mathcal{G} \implies g_s \circ g = g \circ g_s.$$

Then the subgroup is called the center of the group. The center is necessarily an Abelian group. There is always the trivial center consisting only of the identity element.

In the standard example, two types of subgroups appear. One is that the rotation group in two dimensions is a sub-group of the rotations in three dimension. These are the rotations in a fixed plane, and hence $\mathcal{R}_2 \subset \mathcal{R}_3$, even when no particular plane is specified. The set of discrete rotations is also a sub-group of the corresponding rotations, as every element is also a group element of the full rotations, but the sub-group is closed, $\mathcal{R}_n^\alpha \subset \mathcal{R}_n$.

In the non-Abelian $\mathcal{R}_3$ the center is trivial. However, if parity would be added, the identity and parity together would form a non-trivial center.

5.4 Cosets

If there is a subgroup H of a group G, it is possible to define a right/left coset w. r. t. to an element g of a group as the set of all elements hg/gh with $h \in H$. This is in brief also written as Hg/gH. The set of all cosets is denoted as G/H, the coset space. Note that for $g \notin H$, it can be shown that every group element can appear at most in one coset. This implies that for finite groups the dimension of H and the cosets are together the total dimension of the group.

More importantly, if the left and right coset are identical, i. e. for every gh_1 there is some h_2 such that $gh_1 = h_2 g$, or in brief

$$gH = Hg, \tag{5.1}$$

the subgroup H is called a normal subgroup or invariant subgroup. Both trivial subgroups are invariant subgroups.

It becomes more interesting if the subgroup is non-trivial. If this is the case, the set of cosets also forms a group. To see this, define the combination of two cosets as

$$(Hg_1) \circ (Hg_2) = (Hg_1 Hg_1^{-1})(g_1 g_2) = (Hg_1)(Hg_2).$$

This combination is the combination of two cosets as the coset on the combination of the two group elements in the described way. But if the subgroup is invariant, then

$$Hg_1 Hg_1^{-1} = Hg_1 g_1^{-1} H = H,$$

where in the last step it was used that H is a subgroup, and the combination of two arbitrary elements of H is again in H. This implies that the combination of two cosets is again a coset. Since $He = H$, there is also a unit element, and thus also the inverse is included, since this is just $Hg_1 Hg_1^{-1} = He = H$. Thus, the coset space G/H forms itself a group, the so-called factor group of G by H or G/H.

Take for example the group consisting out of the fourth roots of unity, called Z_4 with map the multiplication. The set containing only the square-root is a subgroup, the group Z_2. Since the group is Abelian, the subgroup is an invariant subgroup, with the coset group Z_2, and thus $Z_4/Z_2 = Z_2$ is the factor group.

In the example of $\mathcal{R}_2^\alpha$, if 2α is still a factor of 2π, then $\mathcal{R}_2^{2\alpha}$ is a coset. E. g. for $\alpha = \pi/2$, these are the rotations under π. The two possible cosets then differ by the offset of $\pi/2$. These cosets are invariant cosets, because the combination of two rotations of π and $\pi/2$ can always be equally well represented by a combination of one or three rotations by $\pi/2$. This is also true, because the group is Abelian. Both together give therefore the factor group of $\mathcal{R}_2^{\pi/2}$.

Since cosets do not have many of the useful properties of groups, the existence of an invariant subgroup, and thus coset group, makes a fundamental difference in physics.

5.5 Conjugacy classes and automorphisms

The requirement for an invariant subgroup (5.1) can also be written as

$$gHg^{-1} = H,$$

for any element of the group g, i. e. the subgroup is invariant under a transformation with an element not part of the subgroup. It is useful to generalize this concept to so-called conjugacy classes. A conjugacy class S is a set of group elements, not necessarily a subgroup, which satisfies

$$gSg^{-1} = S$$

for every $g \in G$. Though not immediately obvious, this concept will be useful later on.

These conjugacy classes can also be formulated in a different context. If there exists a map $M : G \to G$ which is 1-to-1[2] and $M(g_1) \circ M(g_2) = M(g_1 \circ g_2)$, i. e. the map preserves the group combination, the map M is called an automorphism of the group.

A special case of an automorphism is

$$M_g(G) = gGg^{-1},$$

with g fixed. This special map is called an inner automorphism. Thus, a conjugacy class S is a set of group elements invariant under all inner automorphisms of the group,

$$M_{G/H}(H) = H,$$

where the subscript indicates the set of elements from which to construct the inner automorphisms.

In the case of $\mathcal{R}_2^\alpha$, every combination of elements form conjugacy classes, since the group is Abelian. For $\mathcal{R}_3^\alpha$, this is not the case. The only conjugacy class is the identity transformation, as there is no invariant subset of rotations under arbitrary rotations.

[2]Note that so far 'onto' is not a requirement, which can make a difference for infinite groups.

Chapter 6

Representations

6.1 Definition of representation

So far, groups have been completely abstract, and could be whatever entities they would be, even just the abstract objects, and everything there exists is a (finite of infinite) table declaring the results of any possible combination. Such an abstract group can then be realized by mapping the set of group elements on a set of linear operators S acting on a vector space V where also a combination of linear operators, usually chaining, is defined for which a unit element exists. A particularly important case of linear operators are matrices with matrix multiplication.

Thus, this requires a map $D : \mathcal{G} \to S$ with S being linear invertible endomorphisms $S : V \to V$. This mapping D is called a representation if it maintains the group structure. This is the case if it fulfills the conditions

- $D(e) = 1$, where $1 \in S$ is the identity linear operator, $1S = S1 = S$, e. g. the unit matrix
- $D(g_1)D(g_2) = D(g_1 \circ g_2)$, and thus there is a one-to-one relation between the combination in G and in S

Especially, this fulfills the existence of an inverse element for all group elements, as $D(g)D(g^{-1}) = D(g \circ g^{-1}) = D(e) = 1$. Hence, the set S must necessarily contain only invertible linear operators fulfilling $D(g^{-1}) = D(g)^{-1}$. This therefore implies that the $D(g) \in S$ form a group with the same structure as the original group $\mathcal{G}$.

For a representation it is not necessary to map every element g to a different element in S. The trivial representation $D(g) = 1$ certainly fulfills both conditions, and the resulting set also implements trivially a group structure in S. If every element is different in S, then it is called a faithful representation. Note that a faithful representation is not guaranteed to exist for any given pair of group and space. E. g. the set $\{\vec{0}\}$ is certainly a suitable zero-dimensional vector space, but except for the trivial group a faithful representation is impossible.

The dimension of the space is then also called the dimension of the representation. A representation is called unitary, if an adjoining operator † is defined on the linear operators,

and $D^\dagger(g) = D^{-1}(g) = D(g^{-1})$.

Of course, there are vectors $|v\rangle$ in the vector space V, and corresponding covectors $\langle v|$, on which the representation $D(g)$ can act as $|v'\rangle = D(g)|v\rangle$. For any basis $|e_i\rangle$ in the linear space there exist then (basis-dependent) matrix elements

$$D(g)_{ij} = \langle e_i|D(g)|e_j\rangle$$

which provide an explicit realization of the representation.

For the standard example, a possible representation are the well-known matrix representations. E. g. for the finite group $\mathcal{R}_2^\alpha$ a two-dimensional faithful representation is

$$D(g(n)) = \begin{pmatrix} \cos n\alpha & \sin n\alpha \\ -\sin n\alpha & \cos n\alpha \end{pmatrix}. \tag{6.1}$$

A possibility for $\mathcal{R}_3^\alpha$ would be parametrizations using the Euler angles in three dimensions, or just by the rotation along the three axes. This gives the faithful three-dimensional representation

$$\begin{aligned} D(g(n_1,n_2,n_3)) &= \begin{pmatrix} \cos n_1\alpha & \sin n_1\alpha & 0 \\ -\sin n_1\alpha & \cos n_1\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos n_2\alpha & 0 & \sin n_2\alpha \\ 0 & 1 & 0 \\ -\sin n_2\alpha & 0 & \cos n_2\alpha \end{pmatrix} \times \\ &\quad \times \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos n_3\alpha & \sin n_3\alpha \\ 0 & -\sin n_3\alpha & \cos n_3\alpha \end{pmatrix}. \end{aligned} \tag{6.2}$$

A trivial representation on $\mathbb{R}^1$ is $D(g) = 1$, the one-dimensional unit matrix. All of these representation fulfills $D(e)$ being the unit matrix.

In case of $\mathcal{R}_2^\alpha$, (6.1) implements the group combination just as $D(g(n))D(g(m)) = D(g(n+m))$, as two rotations in two dimensions is a rotation by the sum of the angles. In three dimensions, this is less obvious, and the group composition in terms of the indices is less trivial due to the non-Abelian nature, but can be implemented as well. Since these matrices are orthogonal, they are automatically also unitary, and therefore all three representation are unitary.

6.2 Relations between representations

Because the group is mapped on linear operators in a vector space, it is always possible to modify the representation by a linear transformation, i. e. a basis change, without altering the composition rule. Thus, two representations are called equivalent if it is possible to change from one to the other by applying to all group elements $D(g)$, the same (invertible) transformation S,

$$D(g)' = S^{-1}D(g)S,$$

since

$$D(g_1)'D(g_2)' = S^{-1}D(g_1)SS^{-1}D(g_2)S = S^{-1}D(g_1 \circ g_2)S = D(g_1 \circ g_2)',$$

as required. Of course, the unit element may differ in the new basis, but it is also related to the original one by the same similarity transformation.

In case of the standard example, this translates to the fact that similarity transformations can be applied, without changing the character of being a rotation.

6.3 Reducible and irreducible representations

A representation does not necessarily affect every element of the vector space V. If there is a subspace, constructed using some projector P, which remains invariant under the group action

$$PD(g)P = D(g)P^2 = D(g)P \tag{6.3}$$

for every group element g, the representation is said to be reducible. However, it should be noted that this can also be regarded as the fact that the group acts trivially on this subspace, i. e. $PD(g)P|_P = 1$. In a sense, the reducible representation is actually a direct sum $D(g)_{1-P} \oplus 1_P$, i. e. the group acts on the whole space, but with two different representations. In the subspace created by P this is then the trivial representation.

A generalization of this concept is given when the group acts with n, possibly different, representations in n subspaces of the whole space, such that the sum of the dimensions of the representations is the total dimensionality of the vector space. Especially, if the basis is such that this representation becomes block-diagonal for every group element g, this is called a completely reducible representation. In this case, the representation of any group element g forms a direct sum $D_1(g) \oplus ... \oplus D_n(g)$ in the vector space, with the D_i being the representations in the various subspaces.

That not every reducible representation is a completely reducible representation can be shown using a counter-example. This requires an infinite group. For finite groups, any reducible representation is also completely reducible[1]. To get a counter-example for infinite groups, consider the (Abelian) group of integers under addition, which form an example of a group. A (non-unitary) 2-dimensional representation is given by

$$D(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}, \tag{6.4}$$

since the unit element is the unit matrix and

$$D(x)D(y) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & y \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+y \\ 0 & 1 \end{pmatrix} = D(x+y).$$

However, there is an invariant subspace, obtained with the projector

$$P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \tag{6.5}$$

[1]This is stated here without proof. Proving it essentially involves first showing that all representations of finite groups are unitary.

Thus, this representation is reducible. It is, however, not completely reducible, as the complement subspace, constructed by $1 - P$, is not an invariant subspace. Hence, a block-diagonal decomposition is not possible in this case.

If there is no invariant subspace, and the representation is not block-diagonal for every group element, the representation is called irreducible. The requirement that a completely reducible representation is block-diagonal implies that the blocks must be irreducible representations within the corresponding subspaces. It is said that the reducible representation has been decomposed into its irreducible representations. This also implies that there can be irreducible representations of different dimensionality.

Especially, an irreducible and faithful representation of a non-Abelian group cannot exist in $\mathbb{R}^1$ and $\mathbb{C}^1$, as the non-Abelian nature of the group composition cannot be reproduced with numbers. This is also true for finite groups.

A trivial example of a completely reducible representation of the standard examples is given by

$$D^R(g) = \begin{pmatrix} D^I(g) & 0 \\ 0 & 1 \end{pmatrix}.$$

In this way the irreducible representation in the upper left corner is extended by a trivial representation. From this completely reducible representation reducible ones can be obtained by an arbitrary similarity transformation.

It should be noted that rewriting the two-dimensional representation of $\mathcal{R}_2$, which will be shown later to be irreducible, to the one-dimensional phase rotation is actually not a possibility to decompose a two-dimensional representation into a one-dimensional representation, and thus making it reducible, as this changes the vector space from a real one to a complex one. However, the two-dimensional representation in a complex vector space would be reducible to the one-dimensional one.

6.4 Regular representation

An important possibility of a (matrix) representation is the so-called regular representation or adjoint representation. It is the most straightforward possibility to generate a representation. It is defined by the action of each linear independent[2] element on the basis vectors of a vector space. Thus, for finite groups the dimensionality of the regular representation needs necessarily to coincide with the size of the group. The elements of the representation are then given by

$$D_{ij}(g) = \langle e_i | D(g) | e_j \rangle, \tag{6.6}$$

where g labels the group elements. In particular, in the adjoint representation there is always a choice of basis for finite groups such that $D^{\text{adj}}(g)_{ij} = D^{\text{adj}}(g)\delta_{ij}$. By insertion of a unity it can be shown directly that this maps the composition law of the group on the matrix multiplication in this space

$$D(g \circ h)_{ij} = (D(g)D(h))_{ij} = \langle e_i | D(g)D(h) | e_j \rangle = \langle e_i | D(g) | e_k \rangle \langle e_k | D(h) | e_j \rangle.$$

[2]What independent is for an infinite group will be defined in a more rigorous way in section 8.1.

Similarly, the other requirements for a representation can be shown.

It should be noted that the adjoint representation is not necessarily irreducible, but can be. For the standard-example of the rotations, the given two-dimensional representation and three-dimensional representation are the adjoint representations.

6.5 Some finite groups

In the following, some groups of a finite order will be discussed. These have simpler properties than those of infinite order, and are therefore useful to demonstrate many of the abstract properties. They are also of high physical relevance. E. g. there are groups for crystals which can be used to classify the crystals according to the discrete rotation groups they belong to. In particle physics, e. g. the discrete parity group plays an important role.

6.5.1 Cyclic group

6.5.1.1 Definition

A particular example of finite groups are the so-called cyclic groups or center groups Z_N of order N. These groups have the property that all group elements are created from one base element $g \neq e$, such that $g^N = e$. E. g. with three elements $\{e, g, g^2\}$. Here, g^2 is a shorthand for $g \circ g$, and likewise higher powers. This group is necessarily Abelian. These properties define the group, they cannot be derived.

This group can be represented by a complex phase with $D(e) = 1$, $D(a) = \exp(i2\pi/3)$, and $D(b) = \exp(i4\pi/3)$, which implements the group structure. The dimension of the (complex) space is thus 1. This representation is also faithful, and irreducible.

6.5.1.2 Parity

A physical example of a cyclic group is the parity transformation in quantum mechanics, which is the group $Z_2 = \{e, g_1\}$. For any Hilbert space, there are only two representations. One is the trivial one $D(g \in Z_2) = 1$. The other is $D(e) = 1$ and $D(g_1) = -1$. Thus, in the completely reducible basis, any one-dimensional subspace can be associated with either of the two representations. Since any element always commutes with the Hamiltonian, as they are always proportional to the unit element, there is a basis where every state has a definite parity. Those in sub-spaces with the trivial representation are said to have even parity, while those in which the other representation acts have odd parity. Thus, the parity of the states are in one-to-one correspondence to the (irreducible) representation of the parity group in their respective subspaces.

Note that this does not specify in any way which states are associated with which representation. This is a dynamical question, for which a Hamilton operator must be specified. In particular, eigenspaces of the Hamilton operator may not align with the parity subspaces. In this case, parity is broken. In that case, the Hamilton operator does not have the symmetry.

6.5.2 Permutation group

A permutation group S_n is the (finite) group of all possible sets of arrangements of n elements.

The permutation group S_2, e. g. contains the elements $e = (1,1)$, $g_1 = (2,1)$, i. e. two elements. g_1 is the possible interchange. A composition is defined by the repeated application. There is only one composition not involving the unit element, $g_1 g_1 = e$, and thus $g_1 = g_1^{-1}$. This result appears evident, as exchanging 1 and 2 and then 2 and 1 leaves things unchanged.

Things become a bit more interesting for the permutation group S_3. Here, there are six (3!) elements besides the unit element. There is the unit element $e = (1,2,3)$, the three transpositions $g_1 = (2,1,3)$, $g_2 = (1,3,2)$, and $g_3 = (3,2,1)$ and the two possible reorderings $g_4 = (3,1,2)$ and $g_5 = (2,3,1)$. These are compositions of the transpositions, e. g., $g_1 g_2 = g_4$, $g_4 g_3 = g_1$, and so on. However, the group is non-Abelian, as, e. g., $g_1 g_2 = g_4 \neq g_5 = g_2 g_1$. It is one of the smallest non-Abelian groups.

While the generalization to larger n is straight-forward, the group S_3 can serve as an example for the necessity of matrix representations. A non-trivial one-dimensional representation over the space of complex numbers is not possible, as then the non-Abelian nature cannot be maintained - $g_1 g_2 \neq g_2 g_1$ and $g_1 g_3 \neq g_3 g_1$ and $g_2 g_3 \neq g_3 g_2$ cannot be realized with numbers except if $g_i = 1$. However, there exists a two-dimensional, then necessarily irreducible, faithful representation of the group,

$$D(g_1) = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \qquad D(g_2) = \tfrac{1}{2}\begin{pmatrix} 1 & \sqrt{3} \\ \sqrt{3} & -1 \end{pmatrix} \qquad D(g_3) = \frac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ -\sqrt{3} & 1 \end{pmatrix}$$
$$D(g_4) = \tfrac{1}{2}\begin{pmatrix} -1 & -\sqrt{3} \\ 1 & -\sqrt{3} \end{pmatrix} \qquad D(g_5) = \frac{1}{2}\begin{pmatrix} -1 & \sqrt{3} \\ -\sqrt{3} & -1 \end{pmatrix}$$

and the trivial $e = 1$. The mapping can be proven, e. g., by direct computation. Also, not all of these six matrices can be simultaneously diagonalized, and this is a irreducible representation.

6.6 Characters

A useful tool to characterize representations of groups are characters. They are defined as

$$\chi_D(g) = \mathrm{tr} D(g) = D(g)_{ii},$$

where the last equality only holds for matrix representations. Assuming for the moment that $V = \mathbb{C}^n$, they are complex numbers, depending on the representation and the group element. For finite-dimensional representations[3], these are numbers, which are invariant under similarity transformations. Thus, all equivalent representations have the same characters. This allows to identify an equivalence class of representations. Note that $\chi_D(g^{-1}) = (\chi_D(g))^*$ for a unitary representation.

[3]To be more precise: representations in trace-classes.

They are also orthonormal in the following sense,

$$\sum_g \chi_{D_a}(g^{-1})\chi_{D_b}(g) = N\delta_{ab}, \tag{6.7}$$

where N is the size of the group. For infinite groups, this number will be a suitable measure of the size of the group to be defined later. The sum is over all group elements, while a and b refer to different representations. Hence, in this sense, representations are orthogonal.

This can be seen as follows. Consider the operator

$$A_{ab;ij} = \sum_g D_a(g^{-1})|a,j\rangle\langle b,i|D_b(g).$$

Note that this operator in general will be a map $A : V_b \to V_a$ between the spaces of different representations. This operator commutes with any $D_a(h)$, since (no summation implied)

$$\begin{aligned} D_a(h)A_{ab;ij} &= \sum_g D_a(hg^{-1})|a,j\rangle\langle b,i|D_b(g) = \sum_g D_a(hg^{-1})|a,j\rangle\langle b,i|D_b(gh^{-1})D_b(h) \\ &= A_{ab;ij}D_b(h). \end{aligned}$$

According to Schur's lemma[4], A is therefore proportional to $\delta_{ab}\lambda^a_{ij}1$, where the 1 acts in the space of the irreducible representation, as the operator does so, and λ^a_{ij} is the constant of proportionality. The still present δ_{ab} restricts the 1 to be active only in the subspace of the operators. The indices merely label the operator. Taking further the matrix elements yields

$$\sum_g \frac{n_a}{N}(D_a(g)^{-1})_{kj}D_b(g)_{im} = \delta_{ab}\delta_{ij}\delta_{km}, \tag{6.8}$$

where n_A is the size of the representation defined by the trace of unity in the space. This is a very important statement about representations. It states that the representation matrices are orthogonal in the space with coordinates the group elements.

An example is given by the Z_N groups. Its irreducible representations are given by

$$D^N_n(j) = e^{2\pi i\frac{nj}{N}}. \tag{6.9}$$

It can be proven that for Abelian groups there are only one-dimensional irreducible representations. Then the only possibilities are these, which differ by the number of times the unit circle is swept over when traversing the full group. However, this cannot be done arbitrarily often. At one point, all sweeps reduce again back to the original one, if the number n becomes again an integer multiple N. For $N = 2$, e. g., this is possible twice, since n can either be zero (trivial representation) or 1. If it would be two, this returns the trivial representation. For $N = 3$, the possibilities are $n = 0$, $n = 1$ (advance by

[4]Which essentially states that only matrices proportional to the unit matrix commute with any other matrix.

$2\pi/3$), and $n = 2$ (advance by $4\pi/3$) sweeps. $n = 3$ gives again back the trivial one, etc.. Thus, there are $N - 1$ independent non-trivial irreducible representations. In this case, the characters necessarily coincide with the corresponding representations, i. e. (6.9).

The relation (6.8) then takes the form

$$\frac{1}{N}\sum_{j=0}^{N-1} e^{2\pi i\frac{nj}{N}} e^{2\pi i\frac{n'j}{N}} = \delta_{n'n}$$

which is nothing but the statement that a Fourier decomposition of the Nth-root of unity is orthogonal when summed over with any possible number of sweeps of the unit circle. Note that the indices at the Ds are otherwise trivial, since it is a one-dimensional representation.

The pre-factor in the sum (6.8) combines the size of the irreducible representations n_a with a normalization factor N, which is the size for a finite group. Thus

$$\sum n_a^2 = N, \tag{6.10}$$

where n_a is the size of the representation and which can be seen after taking a trace over the indices, which will first combine both Ds to a unit matrix, which is then traced to yield a factor n_a. This can be seen as follows. Taking the trace of A, not with respect to the indices defining it, but to the outer Hilbert space, yields the following. Just taking the trace requires tr$A$ to be proportional to the size of non-vanishing elements of the block-diagonal, and thus yields $\delta_{ab}\lambda^a_{ij}n_a$, where n_a is the size of the representation in question. On the other hand, taking the trace over this full space is equivalent to taking

$$\begin{aligned}
\mathrm{tr}A_{ab;ij} &= \sum_k \left\langle k \left| \sum_g D_a(g^{-1}) \right| a,j \right\rangle \langle b,i|D_b(g)|k\rangle \\
&= \sum_{kg} \langle b,i|D(g)|k\rangle\langle k|D(g^{-1})|a,j\rangle = \sum_g \langle b,i|a,j\rangle = N\delta_{ab}\delta_{ij},
\end{aligned}$$

where N is some measure of the size of the group, and for finite groups the actual size. This identifies $\lambda^a_{ij} = n_a/N$, and therefore yields (6.8). The previous example implemented this already, since $1 + 1 = 2$ and $1 + 1 + 1 = 3$ etc..

This result shows again (6.7),

$$\frac{1}{N}\sum_g \chi_{D_a}(g^{-1})\chi_{D_b}(g) = \frac{1}{N}\sum_g D_{aii}(g^{-1}D_{bjj}(g) = \sum_g \frac{1}{n_a}\delta_{ab}\delta_{ij}\delta_{ji} = \delta_{ab}$$

and thus the characters of two different representations are necessarily different, since otherwise this is a sum of squares, and can therefore not be zero except when all characters would vanish.

Because of the trace nature, the characters are invariant under conjugation,

$$\mathrm{tr}D(h^{-1}gh) = \mathrm{tr}D(h^{-1})D(g)D(h) = \mathrm{tr}D(G), \tag{6.11}$$

and thus constant on conjugacy classes.

For finite groups, there is an important consequence. Take a function $f(g)$ of the group elements, which is constant on a conjugacy class. Any function of the group elements must be of the form

$$f(g) = c^a_{jk} D_a(g)_{jk} = \mathrm{tr} c^a D_a = \mathrm{tr} cD.$$

This is nothing but the statement that a map in the group must be representable. The limitation to finite groups comes from the desire to express this as a sum over irreducible representations, which is only possible if the reducible representations are always completely reducible, which is only true for finite groups.

If the function is constant on conjugacy classes, $f(h^{-1}gh) = f(g)$ then

$$f(h^{-1}gh) = \mathrm{tr} cD(h^{-1}gh) = \mathrm{tr} cD(h^{-1})D(g)D(h) = \mathrm{tr} cD(g) = f(g)$$

For this to be true, it must hold for any h. Thus, it must hold also for a sum over all h, appropriately normalized

$$f(g) = \frac{1}{N}\sum_h f(h^{-1}gh) = \frac{1}{N} c^a_{ij} D_a(h^{-1})_{jk} D_a(h)_{li} D_a(g)_{kl} = \sum_a \frac{1}{n_a} \mathrm{tr} c^a \chi_a(g),$$

where the orthogonality relation (6.8) has been used in the last step. This implies that such constant functions are only characterized by the traces of the matrices c, rather than by the full matrices.

The important insight to be gained from this follows from the fact that the sum is over the number of irreducible representations. All functions which are constant over conjugacy classes can be expressed like this, and since the characters are orthogonal, in general all irreducible representations are necessary. At the same time it would be possible to define functions, which are non-zero only on one conjugacy class, and zero elsewhere. This gives a complete basis of these functions. Since both ways of representing this class of functions must be equal, this implies the dimensionality of both bases must be equal, and hence the number of conjugacy classes must be the same as that of irreducible representations. This insight will play a crucial role later.

This has a further important consequence. Define the matrix

$$V_{ai} = \sqrt{\frac{k_i}{N}} \chi_{D_a}(g_i),$$

where k_i is inserted for later convenience. Then (6.11) implies $V^\dagger V = 1$. But since the matrix is square, this also implies $VV^\dagger = 1$, and thus the characters are also orthogonal over group elements

$$\chi_{D_a}(g_i^{-1}) \chi_{D_a}(g_j) = \frac{N}{k_i} \delta_{ij} \tag{6.12}$$

Which is again quite useful.

As an example for the usefulness of these manipulations, consider the regular representation. Due to the definition (6.6), the character $\chi_R(1) = N$, since the regular representation for a finite group is N-dimensional. At the same time, all of the other representation

matrices are necessarily traceless, since otherwise a group element would map some group elements into themselves, which is only possible for the unit element. Hence, the other characters are zero. Now, (6.11) implies that for a general reducible representation r

$$\frac{1}{N}\sum_g \chi_{D_a}(g^{-1})\chi_{D_r}(g) = \frac{1}{N}\sum_g \chi_{D_a}(g^{-1})\sum_{a\in R}\chi_{D_a}(g) = \sum_{a\in R} m_a^r,$$

since the trace of a block-diagonal matrix is just the sum of the traces of the blocks. Hence, m_a^r is the number of times an irreducible representation a appears in the reducible representation r. For the regular representation, it is also true that

$$\frac{1}{N}\sum_g \chi_{D_a}(g^{-1})\chi_R(g) = \chi_{D_a}(e) = n_a,$$

since $e^{-1} = e$ and $\chi_{D_A}(e) = \mathrm{tr}D_A(e) = n_a$, as it is just (unitarily equivalent to) the representation of the unit matrix in the representation a. Thus, $m_a^r = n_a$. Hence, every irreducible representation a appears in the regular representation exactly as often as its dimension.

For the standard example, the characters in the adjoint representations are given in two dimensions by $2\cos\alpha$, with α the rotation angle. For the subgroup with four rotations. Hence, they are 2, 0, -2 (because the group is actually a product group $Z_2 \times Z_2$, and thus there are two non-zero traces) and 0. Due to the cyclicity of the trace, these are invariant under any further rotations, which act as similarity transformations.

6.7 Abelian groups

Abelian groups have a number of further simplifying properties. First note that $h^{-1}gh = h^{-1}hg = g$, due to the Abelian nature. Hence, each element of an Abelian group is a conjugacy class itself. This implies that for finite groups the number of irreducible representation must be the same as the group order, due to (6.10). Hence, all irreducible representations of Abelian groups are one-dimensional. This statement is actually only true if the considered vector space is complex. For a real vector space, this is no longer necessarily true, as the standard example of rotations in two dimensions shows.

Especially, this implies that all irreducible representations of finite, Abelian groups are automatically diagonal. This has important consequences for a quantum system. If H has some finite, Abelian symmetry, e. g. parity and thus a Z_2 symmetry group, then all the appearing irreducible representations can be simultaneously diagonalized with H. Thus, since they are hermitian, the associated quantum numbers become observable. Note that this is a stronger statement than to have a symmetry. A symmetry can always have a unitary representation, but not every unitary representation is diagonal, and therefore not all eigenvalues are observables. E. g., in the case of spin only the total spin and one component can be measured, but not all components individually, since not all components are simultaneously diagonal. Thus, the diagonalizability of representations is important for the question of possible observables in quantum physics.

6.8 Tensor products

So far, most effort was invested into steering towards smaller, i. e., irreducible representations. Building tensor products is quite the opposite. Assume some representation D_1 on a vector space such that there are n base vectors $|i\rangle$, and a second representation D_2 on a second vector space having m base vectors $|j\rangle$.

It is possible to construct a combined vector space $|i,j\rangle$ just by combining both bases. This space has $n\times m$ base vectors. Since both representations act on individual sub-spaces, a tensor product $D = D_1 \otimes D_2$ of both representations can be formally defined as being just a representation acting on the two representations individually. Especially, its matrix elements are

$$D_{klij} = \langle k,l|D|i,j\rangle = \langle k|D_1|i\rangle\langle l|D_2|j\rangle, \tag{6.13}$$

which are by construction $(m+n)\times(n+m)$ elements. Of course, even if the original representations D_1 and D_2, which are $n\times n$ and $m\times m$ matrices in their originally subspaces, are irreducible, the tensor product representation D needs no longer be so. In fact, one of the most important tasks in many cases is to deconstruct a given representation in the relevant irreducible representations.

An important consequence of (6.13) is that characters of a tensor product are products of the characters of the individual characters,

$$\chi_D = \mathrm{tr}D = D_{kiki} = D_{1kk}D_{2ii} = \chi_{D_1}\chi_{D_2}$$

and thus the characters can be used to identify which representation is obtained from the tensor representation, as the characters uniquely characterize a representation.

For rotations, the tensorization is known from, e. g., the coupling of two particles with angular momentum. Different bases are then obtained from the Clebsh-Gordon procedure.

This general construction will be discussed in much greater detail in chapter 10.

6.9 Symmetry group of the $(2n+1)$gon

An example for the concepts so far is the symmetry group of the $(2n+1)$gons, i. e. the regular, closed polygons with $2n+1$ nodes. These objects are transformed into themselves, and therefore symmetric, under rotation by $\pm 2\pi j/(2n+1)$ for $j = 1...n$. The symmetry group furthermore contains, trivially, the identity, but in addition all $2n+1$ reflections by planes through one vertex and one edge. Thus, the size of the group is finite and $1+(2n)+(2n+1) = 2(2n+1)$.

This group has furthermore $n+2$ conjugacy classes. One is just given by the identity. Another one is the set of reflections, as any reflection can be transferred into any reflection by rotations. Thus, a reflection conjugate by a rotation will again yield a (in general different) reflection, and thus the set of all reflections is closed under rotations, and of course, the identity, and hence a conjugacy class. The final set are rotations at fixed j. A rotation by j will become under reflection a rotation by $-j$, and therefore both elements are necessarily in the same conjugacy class. The definition of conjugation requires $g^{-1}sg$

for any element s. Since the rotations in the plane are Abelian, this implies that such a prescription is: rotate by g, then by s, and then back by g, and thus only by s. Hence, any rotation, together with its mirrored companion, is a conjugacy class of its own, adding another n. Of course, since the vertices are identical, any rotation conjugacy class is equivalent.

This symmetry group has the identical representation. There is a second one-dimensional representation, which maps all rotations again to the identity and all reflections to the negative identity, and which is thus inequivalent. Thus, the character only for reflections are not 1 but -1. Both representations are not faithful. Since the number of conjugacy classes has to be the number of irreducible representations, there are n more necessary. These are given by two-dimensional matrices. There are three base matrices, when the 'odd' point is on the x-axis

$$\begin{aligned} D(e) &= 1 \\ D(P) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \\ D(R) &= \begin{pmatrix} \cos\frac{2\pi m}{2n+1} & \sin\frac{2\pi m}{2n+1} \\ -\sin\frac{2\pi m}{2n+1} & \cos\frac{2\pi m}{2n+1} \end{pmatrix} \end{aligned}$$

These are the unit matrix, a reflection through the x-axis, changing the sign of all y components, and the smallest possible rotation if the the maximal rotation should cover 2π m times. The latter creates n inequivalent representations, as they perform n different numbers of full rotations. Other elements are obtained by rotated reflection matrices and by replacing the angle by $2j\pi m/2n+1$. The characters are thus 2 for the unit matrix, 0 for all reflections, and twice the cosine of the angle for all rotations.

Chapter 7

Group actions and structure

This brief chapter will analyze some implications of groups and group structures when analyzing the cases of applying group elements to something, usually the vectors of the vector space in which a representation acts. This will be used here synonymous, but it should be noted that this can also be more abstract constructions, since the only requirement is to have some way to define the action of group elements.

7.1 Orbits

The basic concept is that of an orbit. Take a single vector v of the vector space in question. Then an orbit, sometimes also called group-orbit, $\mathcal{O} = \{D(g)v\}$ is defined as the set of all vectors obtained under application of all group elements from this single vector. An example are the hyperspheres which are obtained under the application of the rotation groups on a single vector pointing to the surface of the hypersphere, e. g. in $\mathbb{R}^2$ the orbit of $(1,0)^T$ would be $(\cos\alpha, \sin\alpha)^T$ under the two-dimensional rotation group.

This structure has a number of consequences. First, the group is itself an orbit, as group elements can be applied to group elements by the group composition, always starting from the unit element, and thus the whole group is a single group orbit. Second, any set on which a group acts can be decomposed into orbits, though the number of orbits does not need to be countable. The orbits are also called classes, and membership of an element of the set in an orbit is thus a class relation.

If an object is invariant with respect to the action of the group, it is called a group invariant. Thus, group invariants are trivial, i. e. one element, orbits. Fourth, cosets of a subgroup H of a group G, i. e. elements of type gh or hg, are orbits as well. Note that in general, except for invariant subgroups, left and right cosets are not identical, and thus the orbits of left cosets and right cosets are not identical.

From the point of physics, it is useful that orbits can also be classified by invariant tensors, which will be discussed in more detail in chapter 10. At the moment, it is sufficient that they are tensors in the vector space of representations. Consider in the simplest case of SO(n) and the representation in the n-dimensional real space. The only invariant tensor is proportional to the unit matrix. Therefore, the length of the vector, which

is invariant under rotation, characterizes the orbit. Since other representations can have more invariants, orbits may have more than one invariant to characterize them. Then, two orbits do only agree if they agree in all the quantities characterized by invariant tensors. E. g. for the representation of the group SU(3) on the vector space of three-dimensional, symmetric matrices, there are two invariant tensors. Therefore, there are two quantities for any orbit which are invariant under the action of the group. One is again the length. The other is a more complicated quantity, which follows from the fact that every matrix of this representation satisfies $d^c_{ab}T_{ab} = 0$, where d^c_{ab} is a special, totally symmetric rank three tensors, and is hence a traceless, symmetric tensor.

7.2 (Non-)linear representations

Though most of the previous has dealt with linear representations, i. e. representations where group elements g have been mapped to matrix representations $D(g)$ such that they act as

$$D(g)_{ab}x_b$$

on the elements x of the representation (vector) space, this is not necessary.

The only requirement for a representation is actually that it maintains the group composition law. Thus, it is possible to formulate a non-linear representation

$$y = f_a(x, g),$$

provided

$$\begin{aligned} y &= (f(x, g), h) = f(x, g \circ h) \\ y &= f(x, e) = x, \end{aligned}$$

where the group composition is maintained. Then, this map is a non-linear representation.

Though non-linear representation of groups are not too common, they are sometimes encountered in physics. An example is the Galileo group of rotations and translations. Under two consecutive Galileo transformations a vector transforms as

$$x \to \Lambda x + a \to \Xi\Lambda x + \Lambda a + b$$

and thus the linear operator behaves as

$$S(g = (\Lambda, a) \circ h = (\Xi, b)) = S(\Lambda\Xi, \Xi a + b) \neq S(\Lambda\Xi, a + b) = S(\Lambda, a)S(\Xi, b) = S(g)S(h).$$

Thus, the representation of translations is non-linear. But the group structure is still maintained, i. e. there is a unique map between composite group elements and composite operators.

7.3 The little group

Given some vector v, it is possible that there is a subgroup H of the original group G for which every subgroup element h satisfies

$$D(h)v = v$$

and thus this particular vector is invariant under a subgroup, and this identifies a suborbit. The subgroup H is then referred to as the little group, sometimes also stability group or isotropy group, of the vector v. Note that a little group of a continuous group can be both discrete and continuous. Orbits with the same subgroup are collected, and called stratum.

A straightforward example is the rotation group SO(3) in its three-dimensional representation. The vector $v = e_z$ has then as little group SO(2), all the rotations in the x-y plane. Note that the little groups are representation-dependent. E. g. in the trivial representation, all orbits have the full group as little group. For SU(2) in the two-dimensional representation, there is only one non-trivial stratum, with the little group being just the trivial group containing only the unit element. The reason is that for every vector in this representation, there is an SU(2) transformation which transforms it into a unit vector in one direction, and thus all of them are not invariant. In the three-dimensional representation, there is again only a single non-trivial stratum, but with little group U(1), as there is a phase free.

In physics, this becomes especially important if there is a physical reason, like an external magnetic field in a spin system, which is fixed. Then a spin system, which had previously some higher symmetry group, will have a lower symmetry group, which is the little group defined by the direction of the magnetic field. The little group is called in this context also the residual symmetry group.

In general, there is an infinite number of orbits, but only a finite number of subgroups. As an example, for the group SO(n) and representations as matrices on vector spaces there is a trivial stratum given by the orbit $v = 0$, which has as little group the original group. All vectors of non-zero length have only SO($n-1$) as little group, the rotations around them. Thus all vectors of non-zero, but fixed, length belong to the same stratum. For representations on a vector space of symmetric rank two tensors, the little groups of SO(n) are different, and the strata can be classified by the number of degenerate eigenvalues, as the spectrum is invariant under rotations, but the ordering is not. Similar, but more complicated, considerations apply to both more complex representations and other groups like SU(n).

Note that the orbits in a stratum for the little group H are in one-to-one correspondence to the coset G/H, as the remaining group elements will transform them into each other.

A general classification of strata and/or little groups is very complicated in general, especially for reducible representations.

The little groups can also be characterized by the invariants of an orbit. Given any orbit ϕ, then an invariant is defined by

$$I(D(g)\phi) = I(\phi),$$

where g is any group element. Thus, the action of an invariant on an orbit is invariant under any group transformation. E. g., once more the length of a vector for SO(n) is such an invariant. As noted, there can be multiple invariants, which can be labeled by an index.

It is possible to ask what are the extrema of the invariants with respect to the orbits,

$$\frac{\partial I(\phi)}{\partial \phi} = 0,$$

that is which orbits maximize the invariant. The solution to this question is known as Morse theory. E. g., for the length of a vector, $I(\phi) = \phi^\dagger \phi$, the only extremum is the trivial vector $\phi = 0$.

Since in physics potentials in the Lagrangian (or Hamiltonian) formulation are invariants of the symmetry groups, often but not always the second-order invariant, Morse theory is actually equivalent to looking for extrema, and thus (metastable) equilibrium of potentials. Since potentials often also break some larger group to a smaller group, the question thus turns into the question of finding the little groups, and strata, given some higher-order invariant.

7.4 Topological groups

To discuss more generally continuous groups, it is useful to introduce topological groups. A topological group is a group on which a topology is defined, see section 2.2, which gives the notation of closeness a realization. Especially, a topology should imply that if g and h are 'close' to g' and h', then so are gh and $g'h'$, as well as g^{-1} and $g^{'-1}$. Particular the last statement is non-trivial, as it requires to maintain the concept of closeness even for the inversion.

Besides introducing some kind of distance function between two group elements $d(g, h)$, not much is required for a topology. Alternatively, what is required is that for any group element it is possible to define a map which is locally isomorphic to Euclidean space. This will be used in chapter 8.

An important subclassification is that of compactness. Topologically compactness occurs if anything can be covered with a finite number of coverings. Local compactness occurs if this is true for a neighborhood. Lie groups in chapter 8 will be seen to be locally compact, as they can be expanded locally around unity, but they are not necessarily globally compact.

Once the group has been equipped with a topology, it is possible to define a path in the group. Since a distance between group elements can be measured, the question can be stated if a group element can be reached from some other group elements by moving along a continuous path, i. e., whether there is a path such that all group elements along the path are always close. If this is true, the group is connected. If not, then it decomposes into disconnected pieces. E. g. the group O(n) is not connected, as it is not possible to find a smooth path in the group between group elements with determinant 1 and -1. Thus, these groups decompose into two (or more) disconnected pieces.

Groups, which are not connected, have the property that the piece containing the identity is an invariant subgroup. This follows because any two group elements which are continuously connected to the unit element, are also connected via following the path with the unit element as an intermediate stop. On the other hand, for any element g connected to the unit element, all kgk^{-1} are connected to the unit element, since the element g can be reduced to the identity in a continuous path, but then $k^{-1}k = 1$. As a consequence, it may be possible that a not connected group is a (semi-direct) product group. E. g. $\mathrm{O}(n)$=$\mathrm{Z}_2\times\mathrm{SO}(n)$ for n odd.

The remaining piece disconnected from the invariant subgroup containing the unit element are cosets of the original group, since any two elements g and h from the same disconnected piece fulfill necessarily the requirement that $g^{-1}h$ is in the piece connected to the unit element. As a consequence, the components form a discrete quotient group, and the semi-direct product mentioned above is in many, though not all, cases, the semi-direct product of this quotient group and the invariant subgroup, as was the case for the $\mathrm{O}(2n+1)$ groups.

Groups which have just one component are called simply connected.

7.5 Group measures

One useful consequence of topological groups is that it is possible to construct a measure $\mu(g)$ of a group, which is essentially given by the mapping to Euclidean space[1]. In general, this measure will then be locally some curvilinear coordinates. The important point is that such a measure requires a generalization of the concept of translationally invariance of it, in the sense of

$$\int d\mu(g) f(gh) = \int d\mu(g) f(g),$$

i. e. it is invariant under any group transformation. However, since the group needs not be commutative, so neither must be $d\mu f(gh)$, and thus in general $d\mu(g)f(gh) \neq d\mu(g)f(hg)$. This defines left measures and right measures, respectively. In general the measure can be determined from a representation as

$$d\mu(g) = \frac{1}{\det M} \Pi dm_{ij}. \tag{7.1}$$

where the $M(g)$ are the representation matrices, and it is necessary to transfer the (continous) enumeration of matrices into the differentials.

An example where the so-defined left measure and right measure are different is the one defined by the three-dimensional real vectors in a two-dimensional matrix representation as

$$L = \begin{pmatrix} e^a & x \\ 0 & e^b \end{pmatrix},$$

[1] Essentially, it is some function which assigns to any set of group elements a number which describes the size of this subset.

with a, b, and x real numbers. The measures are obtained from (7.1) as the left measure

$$d\mu(g)_L = e^{-a} da db dx$$

and the right measure

$$d\mu(g)_R = e^{-b} da db dx.$$

The existence of such an invariant measure is not guaranteed, but can be shown to always exist for locally compact groups. In this case it is called a Haar measure. Furthermore, it is true for any compact groups that

$$\int d\mu(g) < \infty,$$

i. e. the so-defined group volume is finite.

A quite useful further example is the group SU(2). For the lowest-dimensional representation, its elements are given by

$$g(\theta, \eta, \phi) = \begin{pmatrix} \cos\frac{\theta}{2} + i \sin\frac{\theta}{2} \cos\eta & i e^{-i\phi} \sin\frac{\theta}{2} \sin\eta \\ i e^{i\phi} \sin\frac{\theta}{2} \sin\eta & \cos\frac{\theta}{2} - i \sin\frac{\theta}{2} \cos\eta \end{pmatrix}.$$

The resulting Haar measure using (7.1) is

$$\frac{1}{4\pi^2} \sin^2\frac{\theta}{2} \sin\eta d\theta d\eta d\phi$$

which was normalized such that the integral of the unit is one. Without this normalization, the result would be $16\pi^2$.

Chapter 8

Lie groups

Though discrete groups play an important role in many aspects of physics, an even more important role play continuous groups, as most discrete groups are in the end just restricted continuous groups. Therefore, the next step is to discuss continuous groups in more detail, and return to discrete groups when they appear as restriction of continuous groups, see section 9.9. Continuous groups, as their name indicates, are groups with a denumerable infinite number of group elements. A simple example are the real numbers with group combination the addition. Any sum of two real numbers is again a real number, the addition is associative, the zero is the unit element, and the negative of the real number is the inverse. Likewise, the real numbers without zero and infinity (or both included when defining $1/0 = \infty$) under multiplication satisfy the group axioms with the unit element now 1. The same is true for complex numbers. Thus, continuous groups are central. Note, however, that the integer numbers are a discrete group under addition, as there is just a denumerable infinite number of them. All of these are Abelian groups.

When it comes to continuous groups, the paradigmatic example, and probably the most important one for physics besides the real and complex numbers, is the class of Lie groups. These are a special class of continuous groups. They will be the central topic of this chapter.

Compared to the numbers, Lie groups have a number of differing properties. In fact, they are more rigid, as they fulfill additional constraints, and more flexible, as they also permit non-commutative structures.

The structure of continuous groups which will be central is that group elements g are no longer counted by an index, but by one or more parameters α, where α is a single, real-valued number or a vector of real-valued numbers, $g(\alpha)$. This relation needs to be one-to-one. Furthermore, any composition of group elements will yield a new group element chosen as a function f of the two parameters

$$g(\alpha) \circ g(\beta) = g(f(\alpha, \beta))$$

where the choice of the parameter space, which can be some patch or union of patches of the real-valued $\mathbb{R}^n$, together with the composition function f, determines the group.

Particular cases will be continuous groups where it is possible to define a notion of closeness of two group elements $g(\alpha)$ and $g(\beta)$, which is based on the closeness of α and

β. In this case, the group is a smooth topological group. However, this will not preclude the possibility that there are disjoint patches of group elements.

It is in principle possible to generalize the concept by not using real numbers to parameterize the group, to more abstract objects, in particular another group. We will return to this, if need be.

The standard example of rotations are continuous groups, and especially Lie groups. In two dimensions, the function f in the group composition is just the addition of the two angles. In three dimensions, the three angels form a three-dimensional vector. In general, the combination of two group elements will not be just by the addition of the angles, as this is not reflecting the non-Abelian nature of rotations in three dimensions. The composition function is therefore more complicated, and can be taken, e. g., from the theory of the Euler representation.

8.1 Generators

For the discrete numbers, the assignment of group element to index was arbitrary. Which element is number 1 and which number 2 did not matter. Essentially, this assignment is only an irrelevant choice of coordinate system. Likewise, the parameters α are vectors in an n-dimensional vector space. Thus, there is a freedom of choice in their basis. However, a useful convention is to define

$$g(0) = e,$$

i. e. the group element for all parameters vanishing is the unit element. Given its important role in the group axioms of section 5.1 this appears appropriate, and will indeed simplify many calculations considerably.

Any representation of a group will now also depend on the parameters. In fact, since there must exist a one-to-one map from the parameters to group elements, this can also be regarded as the representations being a function directly of the parameters, $D(g(\alpha)) = D(\alpha)$. However, since the representation needs not to be one-to-one, neither needs the relation of the representation elements to the parameters to be one-to-one. It is furthermore useful to transport the convention on the unit element from the group to representations of the group,

$$D(g(0)) = D(0) = 1,$$

where 1 is again the unit element.

A representation is now smooth in a patch of the parameter space, if inside this patch the representation is determined by an analytic function of its parameters, i. e. it can be written as a Taylor series. Especially, if the parameter vector δ is only infinitesimally

different from zero, it must be possible to just use the leading term[1]

$$D(\delta) = 1 + i\delta_a X^a + \mathcal{O}(\delta^2), \tag{8.1}$$

where the elements[2] X^a are also linear operators acting in the vector space of the representation, i. e. they are usually matrices as well, and depend on the representation. If the parameters are N-dimensional vectors, there are N elements X_a. This number is unrelated from the actual dimension d of the vector space in which the representation is defined, which can be larger, smaller, or equal to N. The X_a are called the generators of the group. The i appears by convention, and will yield useful hermiticity and unitary properties of various appearing matrices throughout. Moreover,

$$X_a = -i\frac{\partial}{\partial\delta_a}D(\delta)|_{\delta=0} = -i\partial_a D(0),$$

when considering the representation matrices as functions of the parameters. Finally, if the representation is such as that two different sets of linearly independent parameters yield two different representation elements, then the generators are necessarily linearly independent. Otherwise, there would exist a basis such that $D(\delta)$ would be independent of some δ_a (so-called not parsimonious), and then the element $D(\delta_a)$ would be linearly dependent on other group elements, and therefore the corresponding generator would be.

Continous groups are manifolds, as the parameters come from patches of $\mathbb{R}^n$ and the group composition ensures suitable overlaps. The required expansion (8.1) allows to identify the parameters as the coordinates in the group. Thus, groups are manifolds with additional structures, just like vector spaces are.

For unitary representations, because of

$$1 = D(\delta)D^{-1}(\delta) = D(\delta)D^\dagger(\delta) = (1+i\delta_a X^a)(1-i\delta_a X^{a\dagger})+\mathcal{O}(\delta^2) = 1+i\delta_a(X^a-X^{a\dagger})+\mathcal{O}(\delta^2)$$

it follows that the generators are Hermitian.

Note that, due to the defining requirement of being able to map continous group to some vector space, the topology for a given representation is automatically induced by the topology of the vector space where this representation is defined.

8.2 Algebras

Actually, the generators are more fundamental quantities than the group elements. Especially, it is possible to formally exponentiate the resulting series (8.1) describing a non-infinitesimal group element to obtain

$$D(\delta) = \lim_{N\to\infty}\left(1+\frac{i\delta_a}{N}X^a\right)^N = \exp(i\delta_a X^a). \tag{8.2}$$

[1] In general it may be that a Taylor series has as first term a non-trivial term, and it may also be that the second term is not the linear one. It can be shown that there exists always a parameter redefinition which yields a linear behavior to leading order. This is possible, as there are one-parameter subgroups based on a single generator, under which the composition rule becomes the ordinary addition. These coordinates are called normal.

[2] In most cases the position of the indices do not matter.

Formally, thus, a given group element is reached by performing an infinite number of infinitesimal steps. This is not a unique definition, as the definition (8.1) is compatible with any function which has a linear term in the Taylor series. However, it is the convenient one. Other definitions do not change anything on a qualitative level. Note, however, that it may still be necessary to chose for different group elements different points for the expansion. In particular, not simply connected groups will require for every disconnected part their own expansion point.

Hence, it would be possible to describe all of group theory on an abstract level only using the generators, and never mentioning the group elements or even their representation. The so-obtained structure is called the algebra, the set of all elements

$$X = \delta_a X^a$$

As a consequence, this algebra must be closed under addition, there must be a zero element and an inverse, and a multiplication by (real) numbers[3]. As these are also the criteria for a vector space, the generators do form a vector space, the generator space or algebra space. The generators can be considered to form a basis of this vector space, and the dimension is that of the number of generators.

That is different from the group or its representations. Especially multiplication by scalars in some representation is usually not mapping a group element into a group element. Thus, the relation (8.2) actually describes the map of a vector space to a group. As will be shown later, it is possible to characterize the algebra by just specifying an abstract relation

$$[X^a, X^b] = i f_c^{ab} X^c, \tag{8.3}$$

with $f_{ab}^c = -f_{ba}^c$ the structure constants. Algebras fulfilling the relation (8.3) are called Lie algebras, and the groups constructed from them are called Lie groups. Not every continous group is a Lie group, as it was required that the group is a smooth topological manifold. Note that since an algebra is a vector space of which the generators form a basis, they are not unique. Just with any other basis, they can be redefined into any other set of linearly independent basis vectors.

As can be shown that, for given structure constants f_c^{ab}, (8.3) is a unique characterization[4], this relation is also called itself algebra. Actually, not the values of the structure constants themselves are unique, as when rescaling both the generators and the structure constants the algebra remains invariant, but the relations between the structure constants.

A fact sometimes relevant in physics, to be discussed later in more detail, is that (8.2) is not a unique map from the algebra to the group. Due to the periodicity, it is possible to obtain different groups, depending on how the δ_a are defined. These groups can differ at most by elements which commute with all other elements, and thus only with respect

[3]It is also possible to enlarge this to complex numbers, but this does not yield anything relevant to physics.

[4]Note that sometimes the pre-factor in the algebra is different from i. Sometimes the i is taken as part of the structure constants, or additional factors like 2 or 1/2 appear, and thus the precise form of the algebra is subject to conventions.

to the center of the group. Until discussed, it will also be specified explicitly which group is relevant. Thus, the generators and the algebra are, in a sense, more fundamental.

Given this setup, there is an one-parameter family of group elements characterized by a fixed generator X

$$D(\lambda) = e^{i\lambda X},$$

in which the group composition is given by

$$D(\lambda_1)D(\lambda_2) = e^{i(\lambda_1+\lambda_2)X} = D(\lambda_1+\lambda_2) = D(\lambda_2)D(\lambda_1)$$

and thus by addition of the parameters λ_i. This sub-group, as it also includes always with $\lambda = 0$ the unit element and is closed under the group composition, is hence an Abelian subgroup of the group.

When considering now the composition of two different group elements

$$e^{i\alpha_a X^a} e^{i\beta_a X^a} = D(\alpha)D(\beta) = D(\gamma) = e^{i\gamma^a X_a} \tag{8.4}$$

the particular relevance of the algebra becomes manifest. Because the generators may not commute, γ will in general not just be the sum of α and β. For this the Baker-Campbell-Hausdorff formula

$$e^X e^Y = \exp\left(X + Y + \frac{1}{2}[X,Y] + \frac{1}{12}([X,[X,Y]] - [Y,[X,Y]]) - \frac{1}{24}[Y,[X,[X,Y]]] + ...\right)$$

becomes relevant. It shows that for the group composition rule (8.4) to be valid, the commutator of two generators must again be a generator, since otherwise products of generators would remain in contradiction to (8.4). Only an algebra like (8.3) ensures that this is the case. Put it differently, the necessity to implement the group composition forces the algebra to obey (8.3). Except for the antisymmetry of the first two indices, this does not yet constrain the structure constants other than to obey the group composition.

Of course, since this is an infinite series, it is not always trivial to calculate γ as a function of α and β. Furthermore, the anti-commutativity also implies that derivatives do act non-trivial on the α^a,

$$\partial_{\alpha^b} e^{i\alpha_a X^a} = i\int_0^1 ds e^{is\alpha_a X^a} X_b e^{(1-s)\alpha_c X_c}$$

which can be obtained similarly from the series expansion. Of course, in the infinitesimal case, this reduces to iX_b.

It should be noted that (8.3) does not take notice of the actual representation of the group, and especially not, whether it is reducible or irreducible, as long as there is at least one distinct element for every group element. But the values of the f_c^{ab} are nonetheless fixed by the group composition, as can be seen by evaluating (8.4) to second order,

$$\begin{aligned} e^{i\alpha_a X^a} e^{i\beta_a X^a} &= \exp\left(i(\alpha_a+\beta_a)X^a + \frac{\alpha_a\beta_b}{2}[X^a,X^b]\right) \\ &= \exp\left(i\left(\alpha_c + \beta_c + if_c^{ab}\frac{\alpha_a\beta_b}{2}\right)X^c\right), \end{aligned}$$

since the parameters determine uniquely the group element, which in turn is uniquely determined by the group composition, and which therefore determines the structure constants. Since in this calculation at no point the properties of the representation entered, the structure constants have, up to normalization, a unique value for all representations, and hence are a structural property of the algebra, and thus group. This is not true for the generators, which change depending on the representation.

There is one more statement, which can be made about the structure constants. If there exists at last one unitary representation, then because of the hermiticity of the generators,

$$-if_c^{ba}X^c = -[X^b, X^a] = -[X^{b\dagger}, X^{a\dagger}] = -[X^a, X^b]^\dagger = -i^* f_c^{ab*} X_c^\dagger = if_c^{ab*} X_c$$

and thus the Lie algebra relation

$$[X^a, X^b] = if_c^{ab} X^c, \tag{8.5}$$

can only be true for $f_c^{ab} = f_c^{ab*}$. Thus, when using this form of the algebra with the factor of i explicit, then the structure constants are necessarily purely real for any group which has a unitary representation. Since essentially all continuous groups in physics are of this type, this will be assumed henceforth, except when otherwise noted.

Another identity, which follows directly from the algebra (8.3), is the Jacobi identity

$$\begin{aligned} & [X^a, [X^b, X^c]] + [X^b, [X^c, X^a]] + [X^c, [X^a, X^b]] \\ = \; & if_d^{bc}[X^a, X^d] + if_d^{ca}[X^b, X^d] + if_d^{ab}[X^c, X^d] \\ = \; & i(f_d^{bc} f_e^{ad} + f_d^{ca} f_e^{bd} + f_d^{ab} f_e^{cd})X^e = 0. \end{aligned} \tag{8.6}$$

The last step is a consequence of the the antisymmetry of the structure constants in their first two indices, as can be shown by explicit computation, at least for the cases relevant in this lecture[5]. Using that the generators are independent, the last equation implies actually that already the sum of products of structure constants in the parentheses has to vanish. It is a very useful relation both in the form of the generators and the structure constants.

For the standard example of rotations, the situation is comparatively simple for rotations in two dimensions. Since there rotations commute, all structure constants vanish. For the rotations in three dimensions, this is no longer the case, as they are non-commutative.

A possible realization of the abstract generators of two-dimensional rotations is the single generator

$$R_1^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix},$$

which after exponentiation yields precisely the representation of group elements (6.1). An alternative is the one-dimensional generator 1, if the complex one-dimensional representation of two-dimensional rotations is used.

[5]For infinite-dimensional, or worse, cases, it is not entirely trivial.

A possible representation of the three-dimensional rotations to yield after exponentiation the group representation (6.2) is obtained by

$$R_1^3 = \begin{pmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{8.7}$$

$$R_2^3 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \tag{8.8}$$

$$R_3^3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{pmatrix} \tag{8.9}$$

which therefore is a direct extension of the two-dimensional case. From these, it is straightforward to calculate the structure constants, which turn out to be the anti-symmetric ϵ tensor, up to some constant. As will be seen later, it is no coincidence that these matrices and structure constants look reminiscent of the Pauli matrices, and that there is a deep relation, due to group theory, between the two groups describing spin 1/2 and spin 1 particles.

Note that there is an alternative way to derive the algebra: If a local representation of the group should exist which maintains the group composition, this implies certain analyticity constraints for the continuous representations. It can be shown that they can only be met with an algebra of this structure, where the structure constants satisfy the Jacobi identity.

8.3 Adjoint representation

As the next step it is useful to discuss a very special representation for any Lie algebra: The Jacobi identity can be used to construct the generalization of the adjoint representation of section 6.4 to Lie algebras. Define matrices

$$(t^a)^b_c = i f^{ab}_c. \tag{8.10}$$

Then the Jacobi identity can be rewritten, by judiciously replacing some of the structure constants, but not all, as

$$(t^b)^c_d (t^a)^d_e - (t^a)^c_d (t^b)^d_e - i f^{ab}_d (t^d)^c_e = 0,$$

but this is just the Lie algebra (8.3), and thus the t^a are a representation of the generators. For an n-dimensional[6] Lie algebra, these are therefore $n \times n$ matrices, and this is a representation on an n-dimensional, real vector space. Real, as the matrices are all purely

[6] As a vector space, the dimension of a Lie algebra is given by the number of linearly independent generators.

imaginary, since the structure constants are real. By exponentiation, which after multiplication with i are purely real matrices, this yields also an $n \times n$-dimensional representation of the group, thus also called the adjoint. Note that the so-obtained adjoint representation is not necessarily irreducible or faithful.

Since the generators form a vector space, it is possible to perform a linear transformation on them without changing the algebra relation,

$$X^{a'} = L_{ab} X^b.$$

However, such a change induces a change of the structure constants to maintain the form of the algebra

$$L_{ad} L_{be} [X^d, X^b] = [X^{a'}, X^{b'}] = i f_c^{'ab} X^{c'} = i f_c^{'ab} L_f^c X^f$$

implying

$$f_c^{'ab} = L_{ad} L_{be} (L^{-1})_f^c f_f^{de}.$$

Since any linear transformation must be invertible, the new structure constants exist. Note any rescalings, even by complex numbers, are acceptable linear transformations, implying that different prefactors can, and do in different conventions, appear in the structure constants.

It is an interesting fact that one of the indices does transform differently. This is not coincidental that this is reminiscent of covariant and contravariant vectors, as it is possible in general to extend the concept of Lie algebras to vector spaces with non-trivial metrics, where there are then indeed covariant and contravariant generators, and the structure constants are tensors of third rank, similar to the Christoffel symbols of general relativity. However, in almost all of physics such more general algebras, and resulting groups, do not play a role, and not at all in this lecture. Therefore, in the following mostly a flat, Euclidean metric in the vector space of generators will be assumed, except when noted otherwise. Note that this is a restriction on the possible group manifolds. Then, the position of indices does not matter, and will only be used to enhance readability.

As noted, the representations of the generators span a vector space. A useful scalar product for the adjoint, and later also for other, representation is defined by

$$\mathrm{tr} t^a t^b = d^{ab}, \tag{8.11}$$

which is symmetric. Due to the imaginarity of the t^a it yields a real number. In fact, d^{ab} is, by construction, a real and symmetric matrix.

For an arbitrary linear transformation, this changes to

$$L_c^a L_d^b \mathrm{tr} t^c t^d = L_c^a L_d^b d^{cd},$$

Since arbitrary linear transformations are possible, and d is a real and symmetric matrix, it is possible to diagonalize d, $d = \mathrm{diag} k_1, ..., k_n$, where the k_i are the eigenvalues of d. By choosing for the L a scale factor, it is possible to further transform the k_i always to either 1, 0 or -1. However, since the transformation occurs with the square of L, it is not possible to change the signs, or transform any away from zero.

The values of these eigenvalues actually classify the Lie algebras. The ones most relevant in most of physics are the so-called compact Lie algebras, with $k_i = 1$, i. e. all eigenvalues are positive. These will be the primary concern for now. Another one of particular interest to physics are those with one k_i negative, or equivalently, all but one k_i negative. These non-compact groups contain, e. g., the Lorentz group and the Poincaré group. Groups with any eigenvalue zero are special. Non-compact groups have their name as there exists no non-trivial finite-dimensional unitary representations, while such do always exist for compact groups. This will be seen later.

Returning to compact groups, it is possible to rescale all eigenvalues to some convenient number λ, and thus

$$\mathrm{tr} t^a t^b = \lambda \delta^{ab}.$$

This basis is particular useful, as the structure constants take by (8.10) the values

$$f^{abc} = \frac{1}{i\lambda}\mathrm{tr}([t^a, t^b]t^c) = \frac{f^{ab}_d}{\lambda}\mathrm{tr}(t^d t^c),$$

as obtained from rearranging the algebra. Because the trace is cyclic, the structure constants in this basis are now antisymmetric in all three indices, rather than only in the first two. Since this follows from the compactness of the algebra, this underlines once more the specialness of the last index of the structure constant in general. However, in the compact case, the scalar product (8.11) is positive definite, and thus the vector space has an Euclidean structure. Thus, contravariant vectors and covariant vectors are coinciding, and thus the last index does then no longer need a special treatment, and the position does not convey information. Thus, until turning away from compact group, all indices will be treated equally.

Note that because the generators are not only imaginary but also antisymmetric, this implies they are hermitian. Hence, the adjoint group representation obtained by exponentiation becomes unitary. Compact groups therefore have always at least one finite-dimensional, unitary representation. At least, if the number of generators is finite.

The standard example of the two-dimensional rotations has the trivial representation, i. e. $t^a = 0$ and thus all group elements being 1, as the adjoint representation, as it is an Abelian group. This implies that two-dimensional rotations are not compact, as the only eigenvalue vanishes. This is generic for Abelian algebras. However, in contrast to algebras of non-compact groups with one or more eigenvalues negative there still exist unitary representations.

For the rotations in three dimensions, the adjoint representation is three-dimensional, and coincides with (8.7-8.9).

8.4 Simple algebras and groups

The construction of subgroups from section 5.3 can now be extended to the algebra of Lie groups. This introduces the concept of subalgebras.

An invariant subalgebra, containing the generators $\{Y^a\}$, is defined by the fact that for any X^a in the whole algebra

$$[X^a, Y^b] \in \{Y^a\}, \tag{8.12}$$

and thus the invariant subalgebra is closed.

The invariant subalgebra also creates an invariant subgroup, i. e. for any X in the algebra there is a $Z \in \{Y^a\}$ such that

$$e^{iX}e^{iY} = e^{iZ}e^{iX}.$$

or, equivalently,

$$e^{iX}e^{iY}(e^{iX})^{-1} = e^{iZ}$$

This can be seen by expanding both sides to leading non-trivial order in X and Y, yielding

$$(1+iX)(1+iY)(1-iX) = 1+iY-XY+YX+\mathcal{O}(X^2,Y^2) = 1+iY-[X,Y]+\mathcal{O}(X^2,Y^2),$$

If the subalgebra is invariant, then the commutator is again an element of the algebra, and thus the sum of iY and the commutator is again an element of the subalgebra. Therefore, this is again a group element obtained from the invariant subalgebra, and therefore this is indeed an invariant subgroup.

By definition, both the trivial algebra containing only 0 and the full algebra are invariant subalgebras. These are called trivial subalgebras. If there are no non-trivial invariant subalgebras, the algebra, and group, are called simple. As it will turn out, most things about compact Lie groups and Lie algebras can be obtained from the study of simple Lie groups and simple Lie algebras.

Of course, even if there is no invariant subalgebra, there can still be non-trivial subalgebras, i. e. a set of generators $\{W^a\}$ satisfying

$$[W^a, W^b] \in \{W^a\}$$

but not (8.12). The existence of one or more non-invariant subalgebras is actually the normal case.

Selecting a convenient basis such that there are generators Y^a belonging to an invariant subalgebra and X^a which do not belong, then

$$[X^a, Y^b] = if^{abc}Y^c,$$

by construction. This implies that the f_{abc} have to vanish whenever the indices a and c are both from the algebra, but b is from an invariant subalgebra. Furthermore,

$$[X^a, X^b] \notin \{Y^c\}$$

by definition: (8.12) requires that any X^b which yields with X^a an element of the invariant subalgebra is itself part of the the invariant subalgebra. Thus, any structure constants where the indices mix the invariant subalgebra and the remainder of the algebra have to

vanish. It should be noted conversely that thus the remainder of the algebra necessarily also constitutes at least one other invariant subalgebra.

This result has implications for the reducibility of the adjoint representation. If the algebra would be non-simple, then this implies that generators have non-vanishing sub-matrices if the index a of the generator t^a belongs to a different subalgebra than the indices ij of the sub-matrices t^a_{ij} of t^a. After exponentiation, this implies a unit matrix in this submatrix. Thus, there would be invariant subspaces in this representation, and thus the representation would be reducible. However, if there is no invariant sub-space, such sub-matrices do not exist, and therefore the adjoint representation of a simple Lie algebra is irreducible. This can also be seen by the fact that invariant sub-algebras form distinct, invariant sub-spaces, and thus the algebra would not be simple.

An important case is where the is a an invariant sub-group, which consists only out of a single generator. This subalgebra is therefore necessarily Abelian. For an Abelian invariant subgroup the corresponding structure constants vanish. This implies that the generator of the Abelian subalgebra anticommutes with all other generators, and thus

$$\mathrm{tr} t^a t^b = -\mathrm{tr} t^b t^a = \mathrm{tr} t^a t^b$$

which can only be true if the corresponding k^a vanishes. Thus, the existence of an invariant Abelian subalgebra implies that the algebra is not compact. Thus, compact algebras cannot have invariant Abelian subalgebras. Algebras without such invariant Abelian subgroups are called semisimple. Note that a semisimple group is not necessarily simple, as it can have non-Abelian subgroups. Also, a compact group can have non-Abelian invariant subgroups, and may therefore be not simple. It is semisimple groups which are highly constrained by group theory. These are the elementary objects, and out of them others can be constructed. Thus, unless otherwise noted, in the following only compact, simple groups and algebras will be considered. Note that as a consequence the representations of the algebra will be usually (anti-)Hermitian, and of the group unitary.

The standard examples are quite different with respect to this. The two-dimensional rotation group is itself Abelian, and appears not compact. However, it has, strictly speaking, no non-trivial invariant subgroup, and therefore is both simple and semisimple. The three-dimensional rotations have no Abelian or non-Abelian invariant subgroups, though they have non-invariant subgroups. Thus, they are compact, simple, and semisimple.

8.5 The simplest case: su(2)

To start out, it is useful to consider the simplest possible compact simple and non-Abelian Lie algebra. This will not only illustrate many powerful concepts to be generalized later, but this algebra will also reappear many times over as a building block.

Since this requires to have a non-vanishing three-index antisymmetric tensor, this implies that at least three elements are necessary. Since in this case all other elements are fixed by the single element f_{123}, it is always possible to normalize this element to 1, and thus $f_{abc} = \epsilon_{abc}$, the 3-dimensional Levi-Civita tensor. The resulting algebra

$$[J^a, J^b] = i\epsilon^{abc} J^c$$

is not guaranteed to exist. But it does. It is well known in physics, giving the explicit proof that it exists: This is the algebra realized by the Pauli matrices, and the algebra is the su(2) algebra[7], which after exponentiation yields the SU(2) group, well-known from the physics of spin 1/2 particles.

The first step is to construct representations of the group and the algebra. Of course, these are well known in physics, but for the sake of systematics it is useful to follow here the particular route suitable for generalization in group theory.

To this end, the aim is therefore an explicit matrix realization. Since the representation should be either reducible or irreducible, the aim is to find a block-diagonal representation with the least possible elements of blocks. Since only finite-dimensional unitary representations, and thus Hermitiean generators, will be considered, it is always possible to diagonalize at least one generator. In fact, the number of diagonal generators is given by the number of generators which commute with each other. For su(2), this is at most one, given the values of the structures constants, which can be seen by explicit calculation. Let this be J^3.

As it can be chosen Hermitiean, the spectral theorem applies, and it has for an n-dimensional representation space n eigenvectors forming an orthonormal basis with eigenvalues j_3. These may or may not be degenerate, so they may be characterized by a second quantity α_{j_3}, lifting the degeneracy. Because of the hermiticity, the space spanned by the degenerate eigenvalues can be chosen such as that the vectors for the same j_3 and different α_{j_3} are orthogonal. Necessarily, as the representation is Hermitiean, and therefore all eigenvalues finite, there is a largest value of $j_3 = j$.

To proceed, raising and lowering operators $J_\pm$

$$J_\pm = \frac{1}{\sqrt{2}}(J_1 \pm iJ_2)$$

are introduced, just as in spin physics. This may now seem very specific, but it will be seen that this concept can be readily generalized later.

Explicit calculation yields

$$[J_3, J_\pm] = \pm J_\pm \tag{8.13}$$
$$[J_+, J_-] = J_3 \tag{8.14}$$

as the form of the algebra in this new basis. In fact, this is a special case of the general base transformation discussed in section 8.3. This implies

$$J_3 J_\pm |j_3, \alpha_{j_3}\rangle = (J_\pm J_3 \pm J_\pm)|j_3, \alpha_{j_3}\rangle = (j_3 \pm 1) J_\pm |j_3, \alpha_{j_3}\rangle, \tag{8.15}$$

and thus $J_\pm$ changes an eigenvector of J_3 into one of an eigenvalue differing by ± 1. So far this is the same as in spin physics.

However, now these operators will be used to construct the irreducible representations and to provide a constructive approach to completely reduce reducible ones, thus putting their action into the perspective of group theory.

[7]Note that lower case names will always signify the name of an algebra, while uppercase will signify the associated group.

Start with the irreducible, finite-dimensional representations. Since all steps made so far are true for both reducible and irreducible representations, nothing has to be adapted.

Since the representation is finite-dimensional, there is a lowest value, which can be achieved by applying J_-. After that, the only possibility is to generate the zero vector, as otherwise (8.15) would be violated, but since it follows directly from the algebra, it must be inviolable. Can therefore all eigenvectors be reached by applying J_- often enough to $|j_3, \alpha_{j_3}\rangle$? Assume that there would be an eigenstate of J_3, for which this is not the case

$$j_3|j_3, \alpha_{j_3}\rangle = J_3|j_3, \alpha_{j_3}\rangle \neq J_3 \sum_{n_i} J_-^{n_i}|j, \alpha_j\rangle,$$

but which is maximal, in the sense that acting with J^+ on it would yield zero. But such a vector would then also be linearly independent of all possible eigenvectors created from the other highest state. Since both sub-spaces remain disconnected, these two towers of states would form invariant subspaces, a contradiction to the assumption that this is an irreducible representation. Thus, this case can be disregarded.

Since the states $|j_3, \alpha_{j_3}\rangle$ therefore form a complete basis, which is traversable by $J_\pm$, it is sufficient to determine all matrix elements of the generators in this basis to obtain the explicit version of the representation. Already the requirement of the irreducibility implies that there cannot be any degeneracy, as no operator changes α_{j_3}, as can be seen from (8.15). Thus, the eigenstates cannot be degenerate, as otherwise there would be copies of the subspaces, and therefore the representation would not be irreducible. Therefore, there is no degeneracy, and α_{j_3} can be dropped as well.

It thus remains to determine the explicit representations. To start, note that due to (8.15)

$$J_-|j_3\rangle = N_{j_3}|j_3 - 1\rangle$$

where N_{j_3} is some j_3-dependent number. Now set $j_3 = j$. Then

$$|N_j|^2\langle j-1|j-1\rangle = \langle j|J_+J_-|j\rangle = \langle j|[J_+, J_-]|j\rangle = \langle j|J_3|j\rangle = j\langle j|j\rangle$$

where in the second step it was used that J_+ applied to the highest state yields zero. Since the eigenvectors of J_3 can be assumed to be normalized, this provides the value of N_j up to a phase. But since a rescaling by a phase is always possible without changing the normalization, it is admissible to set it by convention to one and thus

$$N_j = \sqrt{j}.$$

Thus, the action of J_- on the highest state is fully specified.

Conversely, this implies

$$J_+|j-1\rangle = \frac{1}{N_j}J_+J_-|j\rangle = \frac{1}{N_j}[J_+, J_-]|j\rangle = \frac{1}{N_j}J_3|j\rangle = N_j|j\rangle,$$

using the same trick. Since the only involved relation used in these calculation was the algebra, this would have remained identical, even if α_{j_3} would have been kept, showing

that no operator can move a state outside of the subspace created by the ladder starting from the highest state. This confirms the irreducibility argument above.

The same procedure can be used to create a recursion relation for the N_{j_3},

$$\begin{aligned} N_{j-k}^2 &= \langle j-k|J_+J-|j-k\rangle = \langle j-k|[J_+,J_-]+J_-J_+|j-k\rangle \\ &= \langle j-k|J_3+J_-N_{j-k+1}|j-k+1\rangle = j-k+N_{j-k+1}^2, \end{aligned}$$

where the freedom in the phase was already used to have real N_{j_3}. This recursion relation can be solved, e. g. using induction, to yield

$$N_{j_3} = \frac{1}{\sqrt{2}}\sqrt{(j+j_3)(j-j_3+1)}$$

and thus the N_{j_3} depend both on j_3 as well as the highest possible value j.

N_{j_3} vanishes for $j_3 = j-1$. Since J_- only lowers j by one, this j_3 is an integer. Since the only action of J_- is to replace a state by another state, the required, and assumed, vanishing of a state for a finite-dimensional representation can only occur if $j = l/2$, with l some positive integer or zero. In this case a solution is possible which is consistent with the assumptions. Of course, for $l = 0$, there is only a single state, and the representation is the one-dimensional trivial one. Note that this does not constitute a proof that there may not be more exotic other finite-dimensional representations than the one constructed here, though it may look obvious at the physical level. However, it can be proven, which will be skipped here. This will follow from the general constructions later.

The non-zero different values of l always give an irreducible representation. Since there are $(2j+1)$ states possibly in this way, these are the irreducible representations in $(2j+1)$ dimensions. Hence, the lowest-dimensional, non-trivial representation is two-dimensional.

The explicit matrix elements for these representations, called the spin representation, can now be constructed as

$$\begin{aligned} (J_a)^j_{kl} &= \langle j, j+1-k|J_a|j, j+1-l\rangle \\ \langle j, j_3|J_3|j, j_3'\rangle &= j_3\delta_{j_3j_3'} \\ \langle j, j_3|J_+|j, j_3'\rangle &= \sqrt{\frac{(j+j_3+1)(j-j_3)}{2}}\delta_{j_3',j_3+1} \\ \langle j, j_3|J_-|j, j_3'\rangle &= \sqrt{\frac{(j+j_3)(j-j_3+1)}{2}}\delta_{j_3',j_3-1} \end{aligned} \tag{8.16}$$

where the j are now made explicit to identify the dimensionality of the representations. Of course, in physics it just labels the spin of the particle described by this representation, hence also the name of spin representation.

Notoriously well known in physics is the lowest-dimensional representation with $j =$

1/2,

$$\begin{aligned} J_1^{\frac{1}{2}} &= \tfrac{1}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2}\sigma_1 \\ J_2^{\frac{1}{2}} &= \tfrac{1}{2}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{1}{2}\sigma_2 \\ J_3^{\frac{1}{2}} &= \tfrac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{1}{2}\sigma_3, \end{aligned}$$

where the special matrices σ_i are known as the Pauli matrices. They fulfill the useful relation

$$\sigma_a \sigma_b = \delta_{ab} + i\epsilon_{abc}\sigma_c.$$

Exponentiating yields the group elements

$$e^{\frac{i\alpha_a\sigma_a}{2}} = \cos|\vec{\alpha}| + i\frac{\alpha_a\sigma_a}{|\vec{\alpha}|}\sin|\vec{\alpha}| \tag{8.17}$$

where for this particular case the exponentiation is explicitly possible. This also shows that the fundamental representation is faithful, as different $\vec{\alpha}$ are mapped to different matrices.

Since the resulting two-dimensional matrices are unitary, as it is a unitary representation, and of determinant one, this is also called[8] the two-dimensional special unitary group, or SU(2), obtained from the algebra su(2).

Since this is the lowest-dimensional non-trivial representation, since the dimensionality is $2j+1$, this is also called the defining or fundamental representation of SU(2).

Irreducible representations of higher dimensionality can be constructed in the same way. E. g. for $j = 1$

$$J_1^1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad J_2^1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad J_3^1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Note that while in the two-dimensional case the resulting group is the basis for all two-dimensional special unitary matrices, this is no true in three dimensions. There, eight matrices would be necessary. This is only a subset of these. They will be constructed later from a different group.

It is also no coincidence that in three dimensions there are three matrices, reminiscent of the three Euler angles. There is a deep relation between the group SU(2) and the rotation group SO(3) in three dimensions. Essentially, SO(3) is part of SU(2) in a particular way, and actually their algebras su(2) and so(3) coincide. This will be returned to later, and has to do with discrete subgroups of continuous groups.

While the representation is irreducible with respect to the group itself, it is possible that the interesting Hilbert space in physics contains further degrees of freedom, and

[8] Note that the Pauli matrices are also encountered in the context of the so-called quaternions, which are a generalization of the complex numbers.

therefore every state can be further characterized by other observables simultaneously diagonalizable. In this case, any state is characterized by a state vector $|j, j_3, \vec{\beta}\rangle$, where $\vec{\beta}$ contains all other quantum numbers. These states can always be chosen normalized in some norm. This has the interesting consequence

$$\langle j', j_3', \vec{\beta}'|J_a|j, j_3, \vec{\beta}\rangle = (J_a^{j'})_{j_3' j_3''}\delta_{jj'}\langle j', j_3'', \vec{\beta}|j, j_3\vec{\beta}\rangle$$

which can be proven by inserting a complete set of states on either side of J_a and then using the orthogonality relation. The orthogonality relations then permit to explicitly evaluate the matrix elements. Since the matrix elements of J are fixed by the group structure, the group structure alone permits to evaluate all expectation values for states for which the expansion are known. This lies at the heart of the Wigner-Eckart theorem to be returned to later.

Since the construction principle for this irreducible representation can be generalized to any compact Lie algebra, it is worthwhile to summarize it:

1. Diagonalize one of the generators, here J_3
2. Find the state with the largest eigenvalue
3. Use the lowering operators to construct the other states from this state
4. If there are remaining orthogonal subspaces, repeat from 2.

The fact that several seemingly unique properties of su(2) have been used, like lowering operators, foreshadows the fact that similar structures will be found for other groups.

In this context, the values possible for j_3 are also called weights, and therefore the weights of su(2) in its fundamental representation are $\pm 1/2$. Since the construction starts with the maximal value for j_3, the highest weight, this is also called the highest weight construction. It can be proven that this procedure gives all representations, and that they are all finite-dimensional for compact Lie groups.

8.6 Weights and the Cartan subalgebra

Discussing more general groups will be essentially an extension of the previous discussion.

The starting point is to first identify a special subset of the generators of a group: Those which all commute with each others, i. e. those which in a suitable basis in a given representation will be diagonal. This set of commuting generators is called the Cartan subalgebra, or a subalgebra of the Cartan subalgebra if a non-maximal set is chosen. However, the maximal set will be shown to be unique, up to trivial redefinitions, and is therefore usually meant when speaking of a Cartan subalgebra. In the case of su(2), this was J_3.

For su(3), these are J_3 and J_8 in the usual Gell-Mann representation of the generators

$$
\begin{aligned}
2J^1 &= \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad 2J^2 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad 2J^3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \\
2J^4 &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad 2J^5 = \begin{pmatrix} 0 & 0 & -i \\ 0 & 0 & 0 \\ i & 0 & 0 \end{pmatrix} \qquad 2J^6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \\
2J^7 &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad 2J^8 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{pmatrix},
\end{aligned} \tag{8.18}
$$

which will be used in the following as a second example. The number of such generators is called the rank of the algebra.

Chose a representation with hermitian generators. Then the elements $H_i = H_i^\dagger$ of the Cartan subalgebra satisfy

$$[H_i, H_j] = 0.$$

This is again a vector space, as any linear combinations of the subalgebra is again in the subalgebra.

It is possible to chose a basis in this vector space satisfying

$$\mathrm{tr} H_i H_j = k_D \delta_{ij}.$$

Therefore the generators are again orthogonal in the sense of this scalar product.

Since by construction the Cartan subalgebra is the maximal set of commuting generators, it is always possible to find for a given representation a basis in which they are diagonal. As a consequence, any eigenvector of some H_i is also an eigenvector to any other H_j as well, tough for a possibly different eigenvalue. Thus, given any eigenstate, it is characterized by the eigenvalues for all the Cartan elements collected by a vector μ^j, where j runs over the dimensionality n of the Cartan subspace in which the representation is

$$H_i |\mu^j, \beta\rangle = \mu_i^j |\mu, \beta\rangle \tag{8.19}$$

and β collects all other quantities needed to specify the vector uniquely, e. g. if the representation in question is reducible. The eigenvalues can also be collected in n vectors of a dimensionality equal to the rank. The collection of μ_i^j are called the weights or weight vectors μ_i. Since the generators can always be chosen Hermitian, they can always be made real.

For su(2) with rank 1 in the two-dimensional representation, there are two one-dimensional weight vectors, (1) and (-1), or one two-dimensional vector $(1, -1)$.

For su(3) with rank 2 in the three-dimensional Gell-Mann representation the three eigenvectors are$(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$. There are three two-dimensional weight vectors, $(1/2, \sqrt{3})$, $(-1/2, \sqrt{3}/6)$, and $(0, -1/\sqrt{3})$. They can also be collected in two three-dimensional vectors, $(1/2, -1/2, 0)$ and $(\sqrt{3}, \sqrt{3}/6, -1/\sqrt{3})$.

It is often useful to construct a vector-of-vector notation, especially

$$\alpha\mu = \alpha_i\mu_i$$

to define a sum over a product of the weight vectors with another set of vectors α_i of dimensionality rank. The expression $\alpha_i\mu_i$ is then the ordinary scalar product in $\mathbb{R}^{\text{rank}}$.

8.7 Roots

For the adjoint representation the values of the weights are special. It will be seen that they can be used to uniquely specify the basis of the representation, which will be useful later on.

To calculate them, note that the adjoint representation has the same dimensionality d as the algebra itself. It is therefore possible to select for every generator X^a a unique base vector $|a\rangle$. Thus, in this basis a linear combination of two states correspond to a linear combination of the corresponding generators, $\alpha X^a + \beta X^\beta \sim \alpha|a\rangle + \beta|b\rangle$.

The action of a generator in this basis is therefore given by

$$X^a|b\rangle = |c\rangle\langle c|X_a|b\rangle = -if_{acb}|c\rangle \equiv |[a,b]\rangle \tag{8.20}$$

In the second step it was used that the adjoint representation has the structure constants as matrix elements. In the third step it was then defined that this particular linear combination of base vectors is denoted by the base vector of the commutator. This is convenient, but just a notation. By construction, (8.20) implies that the vectors $|[a,b]\rangle$ are zero vectors if both a and b are elements of the Cartan subalgebra. Hence, rank m of these vectors are zero vectors.

To give a particular example, choose once more su(2). A suitable basis is the Cartesian basis and associating $X^a \sim e_a$. The resulting matrix representation of the three generators becomes

$$X^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix} \qquad X^2 = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix} \qquad X^3 = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

which are surely hermitian and traceless, but none of them is diagonal in this basis. Note that the action of the generators in this basis on their corresponding state is to annihilate this state vector, $X^a|a\rangle = 0$. Note that the Cartan element is not diagonal in this basis.

Since the representation of the generators are by construction Hermitian, their eigenvectors form a complete basis. Thus, there exists a suitable basis $|\alpha\rangle$, in which every base vector is an eigenvector to every Cartan element,

$$H_i|\alpha\rangle = \alpha_i|\alpha\rangle.$$

By construction, the α are then the weight vectors in the adjoint representation. Here and hereafter the index j identifying the individual basis vectors $|\alpha^j\rangle$ is suppressed, if not needed explicitly.

For the su(2) case, there is only one Cartan element, which has eigenvalues 1, 0, and -1. Diagonalizing the Cartan element by a basis change, this becomes

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The base vectors $|\alpha\rangle$ in this new basis are again the conventional Cartesian unit vectors. However, they are no longer directly associated with the original generators. However, there are now three rank(=1)-dimensional weight vectors (1), (0), and (-1) or a rank(=1) three-dimensional weight vector $(1, 0, -1)$. This emphasizes that the number of weights depends on the representation.

Of course, this can be done fo any algebra. It is then possible to find linear combinations E_α of the non-Cartan generators such that

$$[H_i, E_\alpha] = \alpha_i E_\alpha \tag{8.21}$$

is satisfied, where the α are the weight vectors. Thus the other generators can be written such that they are characterized by the $d - m$ m-dimensional non-zero weight vectors, which specify their commutation relations with the Cartan subalgebra. Since the Cartan elements commute with each other, the remaining m weight vectors are necessarily zero vectors in this basis. To keep the association with the base vectors, this implies also a change of basis.

This permits still to chose a new normalization of the generators. It is convenient to select

$$\begin{aligned} \mathrm{tr} H_i^\dagger H_j &= \mathrm{tr} H_i H_j = \lambda \delta_{ij} \\ \mathrm{tr} E_\alpha^\dagger E_\beta &= \lambda \Pi \delta_{\alpha_i \beta_i}, \end{aligned} \tag{8.22}$$

where λ is some overall normalization constant.

As an example consider su(2). Given the diagonal form

$$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

the corresponding new basis then yields for the E_α

$$E_{\alpha=(-1)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad E_{\alpha=(1)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{8.23}$$

If V is the matrix to diagonalize, say, X^3, then for su(2)

$$E_{(\pm 1)} = \Lambda(iVX^1V^{-1} \pm VX^2V^{-1}),$$

where $\Lambda = \mathrm{diag}(-1,-1,1/2)$ is a normalization of the type discussed in section 8.3, ensuring (8.22). This implies, e. g.,

$$[H, E_{(-1)}] = -1 E_{(-1)},$$

making the appearance of the weight factors and the naming explicit.

Note that all matrices remain Hermitian and traceless, though they are no longer imaginary. Furthermore, they act now differently on the base vectors,

$$E_\alpha|\alpha\rangle = 0 \tag{8.24}$$
$$E_\alpha|0\rangle = |\alpha\rangle \tag{8.25}$$
$$E_\alpha|-\alpha\rangle = |0\rangle, \tag{8.26}$$

Thus the E_α move the base vectors around. Note that $|0\rangle$ is not the null-vector, 0, but indicates here the zero weight vector in the direction of the Cartan algebra. In the su(2) cases, these are the base vectors $|1\rangle$, $|0\rangle$ and $|-1\rangle$, and the E_α act like the conventional ladder operators.

It is not an accident that in the example $E_\alpha = E^\dagger_{-\alpha}$ holds. This applies generally. To see this, take the hermitian conjugate of (8.21),

$$-[H_i, E_\alpha]^\dagger = [H_i, E^\dagger_\alpha] = -\alpha_i E^\dagger_\alpha.$$

Here, it has been used that the weight vectors are real as eigenvalues of the Hermitian matrices of the Cartan subalgebra. This implies that

$$E^\dagger_\alpha = E_{-\alpha}$$

Thus, as long as $\alpha \neq 0$, this implies that there is in this basis always a negative weight corresponding to a Hermitian generator. The form (8.23) shows this explicitly, but the proof did not make reference anywhere to su(2), and is therefore generally valid. This also implies that in this basis not all generators are Hermitian. This is not a problem, as the explicit example of su(2) with the basis J_3 and $J_\pm$ shows, where the $J_\pm$ also are not Hermitian, but satisfy $J^\dagger_\pm = J_\mp$. More importantly, this implies that the number of non-Cartan generators is necessarily even for a Lie algebra, and that all non-zero weight vectors in the adjoint representation always come in pairs of opposite sign.

The weights in this basis are called roots, and their vectors hence root vectors. There are two interesting features of this particular way of choosing the representation.

One is that the operators $E_{\pm\alpha}$ raise and lower the weights μ of any given state by $\pm\alpha$, where γ are any other involved quantum numbers,

$$H_i E_{\pm\alpha}|\mu,\gamma\rangle = ([H_i, E_{\pm\alpha}] + E_{\pm\alpha}H_i)|\mu,\gamma\rangle = (\pm\alpha_i E_{\pm\alpha} + E_{\pm\alpha}H_i)|\mu,\gamma\rangle = (\mu_i \pm \alpha_i)E_{\pm\alpha}|\mu,\gamma\rangle. \tag{8.27}$$

This has already been seen in (8.24-8.26) for the particular case of the state with weights the roots. Especially, this implies that acting with $E_{\pm\alpha}$ on the respective other state $|\mp\alpha\rangle$, these states are annihilated, as $(\pm\alpha \mp \alpha) = 0$.

Due to (8.20), this has implications for the commutator of E_α and $E_{-\alpha}$. This can also be seen in a more explicit way. First, note that by explicit calculation using (8.21)

$$[H_i, [E_\alpha, E_{-\alpha}]] = 0$$

Hence, the commutator must be proportional some linear combination $\beta_i H_i$ of the Cartan generators. As usual, the coefficients can be obtained by tracing,

$$\lambda\beta_i = \mathrm{tr} H_i \beta_j H_j = \mathrm{tr} H_i [E_\alpha, E_{-\alpha}] = \mathrm{tr}[H_i, E_\alpha] E_{-\alpha} = \alpha_i \mathrm{tr} E_\alpha^\dagger E_\alpha = \lambda\alpha_i$$

and thus

$$[E_\alpha, E_{-\alpha}] = \alpha_i H_i. \tag{8.28}$$

Again, this directly visible from the explicit su(2) case,

$$\begin{aligned} E_{(1)}E_{(-1)} - E_{(-1)}E_{(1)} &= (1)_1 H_1 = \mathrm{diag}(1, -1, 0) \\ E_{(-1)}E_{(1)} - E_{(1)}E_{(-1)} &= (-1)_1 H_1 = \mathrm{diag}(-1, 1, 0). \end{aligned}$$

Note, how the choice of exchanging the two E_α does exchange the α_i on the right-hand side.

Though not explicitly written here, these results are encoding a lot more. In general, there are n generators of which m form the Cartan subalgebra. In the $d(= n$ in the adjoint representation)-dimensional representation, there are thus d weight vectors of dimensionality m. In the adjoint representation, these are either zero or come pairwise with opposite signs. Thus in statements like (8.28) actually α is also a counting index counting the weight vectors, and in a more explicit notation, α would receive an index j running from 1 to $d - n/2$, where the division of 2 comes from the $\pm$-degeneracy, with components α_i^j.

This is once more explicit in the su(2) case. In the 2-dimensional representation, there is one Cartan element, and thus there are two weight vectors which are one-dimensional. In the 3-dimensional adjoint representation, there are three one-dimensional weight vectors, one being zero, and the other two being the negative of each other. For su(3), with a two-dimensional Cartan algebra, the lowest-dimensional representation is three-dimensional. Thus, there are three two-dimensional weight vectors. The adjoint representation is eight-dimensional, and there are thus eight two-dimensional weight vectors. Of these two are zero, as they correspond to the two Cartan-elements. The other decompose into three pairs, each pair related by a sign.

The number of non-trivial weights is actually the same. In the two-dimensional su(2) case it is two and so it is in the adjoint representation. In the su(3) case, it is in the three-dimensional representation $3 \times 2 = 6$. In the adjoint representation, there are only three independent, non-zero weight vectors, and there are therefore again six independent weights. That does not seem to be a coincidence, and the weights indeed will play an important role in the following.

8.8 The su(2) substructure of arbitrary algebras

Now it turns out that any non-zero root vector actually selects an su(2) subalgebra of every algebra. To see this, define generators[9]

$$J_\pm = \frac{1}{|\alpha|} E_{\pm\alpha} \tag{8.29}$$

$$J_3 = \frac{\alpha_i H_i}{|\alpha|^2}. \tag{8.30}$$

Thus while the so-defined $J_\pm$ generators are simple generators and depend implicitly on the choice of α, J_3 is an object involving potentially the whole Cartan subalgebra, but still remains an element of the Cartan subalgebra. Note that in principle there are several such sets of generators, which should be distinguished by having an index α identifying them. If it is not necessary to distinguish two different sub-algebras, this will be suppressed.

In the su(2) case itself, there is only a single root vector, $\alpha = (1)$. Thus the sum for J_3 collapses, and $J_3 = H$, as $\alpha^2 = (1)^2 = 1$. Trivially then $J_\pm = E_{\pm\alpha}$.

In general,

$$[J_3, J_\pm] = \frac{\pm\alpha_i\alpha_i E_{\pm\alpha}}{|\alpha|^3} = \pm\frac{E_{\pm\alpha}}{|\alpha|} = \pm J_\pm \tag{8.31}$$

$$[J_+, J_-] = \frac{[E_\alpha, E_{-\alpha}]}{|\alpha|^2} = \frac{\alpha_i H_i}{|\alpha|^2} = J_3, \tag{8.32}$$

which, by comparison to (8.13-8.14), is an su(2) algebra. This identifies therefore an su(2) subalgebra in any Lie algebra. Moreover, because every generator has been associated with a root vector, this implies that every group has su(2) subalgebras of this form. This implies again that the number of non-Cartan generators of every group is even, as all non-zero root vectors appear pairwise. Especially, this also implies that every irreducible representation of any algebra can therefore be decomposed into irreducible representations of su(2). This will be a decisive insight in the following.

One question still arising is about the uniqueness of root vectors, i. e. can there be degenerate root vectors, which correspond to different E_α. Assume that there is a second generator E'_α to the same α. If it is not proportional to the original E_α, its corresponding base vector can always be decomposed into a part proportional and orthogonal to the one of E_α. But, by construction, the base vectors are eigenvectors of the Cartan, which form a full base, labeled by α. Thus, the orthogonal part is either an element of the Cartan, or belong to a different α, and therefore the operator cannot be different.

Furthermore, two different root vectors cannot be proportional to each other, except for sign reversal. Any such proportionality would rescale J_3 by its inverse. But since the representation is three-dimensional, since it is the adjoint representation of an su(2) subalgebra, the eigenvalues of J_3 must necessarily be ± 1 or zero. Thus, a constant of proportionality would yield the wrong behavior. Alternatively, because of (8.27), any α different would require further states than the three included in the adjoint representation.

[9] Note that the operators J_3 represent charge operators in physics.

Note that this construction also implies an important statement about generators. The point of the whole construction is that every generator can be rewritten as a sum of generators of su(2). These are traceless, since the raising/lowering operators are traceless to connect different states. The Cartans are also traceless, as any su(2) representation is symmetric, and therefore all Cartan elements have the same positive and negative eigenvalues, and are thus traceless as well. Hence, generators are always traceless. Note that the deconstruction has only be performed for compact Lie groups, so this statement also only applies to compact Lie groups.

8.9 Geometry of weights

The fact that it is possible to decompose the set of generators of every algebra according to (8.29-8.30) into sets of su(2) generators has very important consequences. Especially, since the algebra is a linear space, these combinations can also be used to define a new base of generators, which still satisfy the su(2) algebra (8.13-8.14), even when applied in a different representation.

Select some representation D and some state of the corresponding vector space having weight μ and some other quantum numbers γ. Then select one of the su(2) subalgebras, characterized by the root vector α, and apply its J_3 to the vector. This yields

$$J_3|\mu,\gamma\rangle_D = \frac{\alpha_i H_i}{|\alpha|^2}|\mu,\gamma\rangle_D = \frac{\alpha_i \mu_i}{|\alpha|^2}|\mu,\gamma\rangle_D. \tag{8.33}$$

Thus the eigenvalue of an arbitrary state of any representation of any J_3 is given by the projection of the weight vector on the root vector. Because J_3 is part of an su(2) algebra, it is necessary that its eigenvalues must be integer or half-integer, since this follows entirely from the algebra as shown in section 8.5.

Since the space is some representation space of the group, the vector is some linear combination of the basis vectors of the contained representation. Consider the case where the highest su(2) representation is j. Now, there is then some $p \geq 0$ such that

$$\begin{aligned}(J_+)^p|\mu,\gamma\rangle_D &\neq 0\\ (J_+)^{p+1}|\mu,\gamma\rangle_D &= 0.\end{aligned}$$

Applying J_3 yields

$$J_3(J_+)^p|\mu,\gamma\rangle_D = \frac{\alpha(\mu+p\alpha)}{\alpha^2}|\mu,\gamma\rangle_D$$

However, this must be an eigenstate of J_3, since all other contributions have been filtered out using the raising generator, and therefore the pre-factor

$$\frac{\alpha(\mu+p\alpha)}{\alpha^2} = j \tag{8.34}$$

must be proportional to j, and thus integer or half-integer, as otherwise another raising operator would not annihilate it.

In the same way the application of lowering operators yields

$$\begin{aligned}(J_-)^q|\mu,\gamma\rangle_D &\neq 0\\ (J_-)^{q+1}|\mu,\gamma\rangle_D &= 0,\end{aligned}$$

for some integer q, but which must then be an eigenvector of J_3 to eigenvector $-j$. Setting both results equal yields

$$2\frac{\alpha\mu}{\alpha^2}+p-q=0\rightarrow\frac{\alpha\mu}{\alpha^2}=-\frac{1}{2}(p-q), \tag{8.35}$$

which is a central statement in the following, as it will be the key for classifying Lie algebras. Note that j does no longer appear, and therefore the particular representation does not matter.

Now consider the case that μ is actually also a root vector, say β. This yields

$$\frac{\alpha\beta}{\alpha^2}=-\frac{1}{2}(p-q).$$

But the vector $|\mu,\gamma\rangle_D$ was general. Selecting instead the subalgebra characterized by the root vector β and using $\mu=\alpha$ instead yields

$$\frac{\beta\alpha}{\beta^2}=-\frac{1}{2}(p'-q'), \tag{8.36}$$

where p' and q' are again some other integers.

The true power of this result is found by multiplying both equations yielding

$$\frac{(p-q)(p'-q')}{4}=\frac{(\alpha\beta)^2}{\alpha^2\beta^2}=\cos^2\theta_{\alpha\beta} \tag{8.37}$$

This implies that the angle between the root vectors cannot be arbitrary, since the expression on the left-hand-side involves only integers. In fact, $\theta_{\alpha\beta}$ can only have the non-trivial values $\pi/2$, $3\pi/2$, $\pi/4$ or $5\pi/6$ - zero or π only occur if $\alpha=\pm\beta$, and thus correspond to the same su(2) subalgebra. Hence, the root vectors can only appear in very specific geometrical combinations.

8.10 The space of weights for su(2) and su(3)

To make the two previous, rather abstract, constructions more tangible, it is useful to consider examples. For the su(2) case itself, with only one root vector $(1,-1,0)$, there can only be two angles in (8.37), 0 or π. In the one-dimensional space of weight, they are symmetric around zero, being in the lowest-dimensional representation (1) and (-1) and in the three-dimensional adjoint representation (1), (0), and (-1). Thus, one of the two angles only describes the case of the negative root vector. For the su(3) case it is more interesting.

For su(3) the Cartan subalgebra is given by $J_3 = H_1$ and $J_8 = H_2$. In the lowest(= 3)-dimensional representation these have the eigenvalues $\{1/2, -1/2, 0\}$ and $\{\sqrt{3}/6, \sqrt{3}/6, -\sqrt{3}/3\}$, as can be read off from (8.18). Thus there are three weight vectors, combining two components each. If plotted in a plane of eigenvalues of H_1 and H_2, the weight plane, they form an equilateral triangle, at the corresponding coordinates.

From the remaining 6 generators it is then possible to construct three su(2) subalgebra pairs of raising and lowering operators. They can be enumerated by the root vectors. These root vectors must be the differences of weights, since the raising and lowering operators must move from one root vector to another by virtue of (8.27), and thus shift the eigenvalue by one. Using (8.18), it can be shown that the combinations

$$J^{\pm}_{(\pm 1,0)} = \frac{1}{\sqrt{2}}(J_1 \pm iJ_2) \tag{8.38}$$

$$J^{\pm}_{(\pm 1/2,\pm\sqrt{3}/2)} = \frac{1}{\sqrt{2}}(J_4 \pm iJ_5) \tag{8.39}$$

$$J^{\pm}_{(\mp 1/2,\pm\sqrt{3}/2)} = \frac{1}{\sqrt{2}}(J_6 \pm iJ_7). \tag{8.40}$$

fulfill these requirements. Note that the three J_3 generators belonging to the three algebras are given by

$$\begin{aligned} J^3_{\pm 1,0} &= J_3 \\ J^3_{\pm 1/2,\pm\sqrt{3}/2} &= \frac{1}{2}J_3 + \frac{\sqrt{3}}{2}J_8 \\ J^3_{\mp 1/2,\pm\sqrt{3}/2} &= -\frac{1}{2}J_3 + \frac{\sqrt{3}}{2}J_8. \end{aligned}$$

Thus, only for the first root vector they are given by one of the original Cartan generator like in the su(2) case, and are otherwise linear combinations. But these three are not linearly independent, and are just different linear combinations of the Cartan elements.

The adjoint representation of su(3) is eight-dimensional, as there are eight generators. Each element of the Cartan has therefore eight eigenvalues, and thus there are eight two-dimensional root vectors, which can be drawn in the same weight plane as the triangle. Two of them are necessarily zero vectors, leaving six non-trivial ones. These are given explicitly by the structure (8.38-8.40) due to (8.33). In the end, the non-zero root vectors are found to form a regular hexagon, with $\pi/6$ angels. Note that these eight vectors are not linearly independent, as the space is still the two-dimensional weight space of the two-dimensional Cartan subalgebra. This ambiguity shows that somehow weights and roots contain still superfluous information. This can be changed by requiring ordering, as will be discussed next.

8.11 Simple roots

Somehow, weight vectors in general, and root vectors in particular, contain too much dependent information. While eventually it is arbitrary how to deal with this, there is

one approach which turns out to be particularly suitable, which proceeds by introducing an ordering principle in the weight vectors. The thing which is needed to be implemented is thus some mechanism that a statement that a given weight is positive makes sense, so that raising and lowering operators due to (8.27) actually do raise or lower. Especially, this will require to define what is meant by highest weight.

In general, there is no ordering relation for vectors. However, in any fixed basis, it is possible to order by components. Introducing a fixed basis may seem suspect at first, as basis invariance is one of the greatest achievements of mathematics. But this will be rectified later, and the results will actually be basis invariant. For the practical calculations, however, it is as often useful to work in a fixed basis.

Once agreed, and given the weights in some basis, the convention will be the following: If the first non-zero component, counting from the top, is positive, then a weight or root is called positive. If it is negative, it is called negative. This also pertains to combinations. E. g. $\mu - \nu$ for two weight vectors μ and ν is positive if the first non-zero component of the result is positive.

In case of the two one-dimensional weight vectors in the lowest-dimensional representation of su(2), (1) and (−1), the first is thus positive and the second negative. For su(3), the weight $(1/2, \sqrt{3}/6)$ is positive and the weights $(-1/2, \sqrt{3}/6)$ and $(0, -\sqrt{3}/3)$ are negative. Of the root vectors, three are positive and three are negative, as can be read from (8.38-8.40). The two zero roots are, of course, neither positive nor negative. Hence, for an arbitrary representation the number of positive and negative weights do not need to match, as is visible for su(3) in the lowest-dimensional representation. But because of the relation that for every positive root α there is the reversed root $-\alpha$, there is always the same number of positive as negative roots in the adjoint representation. Finally, a change of basis may change the relative number of positive and negative weights, but not of roots.

As is visible from these examples, not all roots (or weights) are independent, as soon as the representation has a higher dimension than the Cartan subalgebra. This is automatically the case for the the adjoint representation of any non-Abelian algebra. In fact, at most as many weights can be linearly independent as is the dimensionality of the Cartan. Thus, there must be a minimal set of roots from which all others can be constructed. Since the roots can be used to construct the group, it would be good to identify this minimal set in terms of roots. So the question is, which of them are the independent ones. Well, any set which is linearly independent. But this leaves the problem of identifying a useful one. It turns out that a particular useful statement is that all roots can be created from a set of basic roots, the so-called simple roots. These are defined as all positive roots which cannot be written as sums of other roots. That this is sufficient for any Lie algebra requires a proof, whose construction will also provide further interesting insights. This is necessarily a stronger requirement than just linear independence, as the case of su(3) shows explicitly.

It is best to do so step by step.

First note that the difference $\alpha - \beta$ of two simple roots α and β cannot be a simple root. If $\beta > \alpha$, then $\beta - \alpha$ is positive. But then is $-\beta$ a sum of two positive roots, $\alpha - (\beta - \alpha)$, and can therefore be not simple. In the reverse case so is α, and also $\beta \neq \alpha$. If the first non-zero components match, the ordering can always be done according to the second components, and so on, since not all components can be identical. Visually, this

can be directly seen in the case of su(2) and su(3).

Because $\alpha - \beta$ is not a simple root, the corresponding base vectors have to vanish under the action of the corresponding raising and lowering operators,

$$J_\alpha^- |\beta\rangle = 0 = J_\beta^- |\alpha\rangle$$

as they act in the subspace of different su(2) subgroups. In terms of the adjoint representation of su(3), which is eight-dimensional, there are two dimensions being the Cartan, and three pairs which combine with some combination of the Cartans to form three su(2) subspaces of three dimensions each. The lowering and raising operators only act on states in these subspaces.

Using (8.35), this implies for two integer p and p'

$$\begin{aligned} \cos\theta_{\alpha\beta} &= -\frac{\sqrt{pp'}}{2} \\ \frac{\beta^2}{\alpha^2} &= \frac{p}{p'}, \end{aligned}$$

and thus yield an even stronger constraint for simple roots. This implies that

$$\frac{\pi}{2} \leq \theta_{\alpha\beta} < \pi, \tag{8.41}$$

where the upper bound stems from the fact that both roots must be positive.

This helps in establishing that simple roots are linearly independent. Consider an arbitrary linear combination γ of simple roots

$$\gamma = \mu - \nu = \sum x_\alpha \alpha - \sum y_\beta \beta,$$

where x_α and y_β are all positive. Since both sums run over all roots, it is always possible to split the sum into those with positive coefficients and negative coefficients and isolate the sign as a prefactor as done here. If the simple roots would be linearly dependent, then there would be a non-trivial possibility to choose all prefactors such that γ vanishes. However, the norm satisfies

$$(\mu - \nu)^2 = \mu^2 + \nu^2 - 2(\mu\nu) > 0$$

due to (8.41) - the cosine of the angle in this range is always negative or zero, and thus the last term is never negative. Since not both μ and ν can have zero norm at the same time, the norm of γ is always non-zero, and hence the simple roots are linearly independent.

On the other hand, by definition, all other positive roots can be obtained by sums of the simple roots with positive, integer coefficient, since negative coefficients are not permitted as this would yield a difference and hence not a positive root.

This leads to the conclusion that the number of simple roots is actually the dimensionality of the root space, 1 for su(2) and 2 for su(3), and thus there are as many simple roots as the dimensionality of the Cartan. If this would not be the case, there would be some root δ which cannot be determined by the linearly independent roots. But the

corresponding Cartan element δH would then commute with all $E_{\pm\alpha}$ due to (8.21). Thus, it would belong to a different disconnected subalgebra then those present, but this is only possible if the group is not simple, in contradiction to the starting point.

hence, the simple roots span the space of roots. Therefore, there should be a constructive way to obtain the remaining roots. The negative roots are trivially obtained by a multiplication from the positive ones, and the zero roots are also known. It thus requires only to obtain the remaining non-simple positive roots. These are sums of the simple roots, but which?

Due to (8.27), it is sufficient to act with the E_α on the simple roots, until evaluation of (8.35) yields a contradiction. The only question to answer is, if this procedure could miss some positive root γ. This does not happen, since this would imply that acting on any such state with $E_{-\alpha}$ must annihilate it. However, it is still a linear combination of the simple roots, thus

$$\gamma\gamma = \left(\sum k_\alpha \alpha\right)\gamma$$

But because of (8.33), this is just the eigenvalue of the state γ, and this needs to be negative. At the same time, this is the norm of the state, which must be positive, leading to a contradiction. Hence, the described procedure generates all positive, and thus all, roots.

Still, this has not yet identified the simple roots. For su(2), this is just (1). For su(3), these are the two roots $(1/(2\sqrt{3}), \pm 1/2)$, since $(1/\sqrt{3}, 0)$ can be written as their sum, and is therefore by definition not simple. Since there must also be two simple roots, and these are the only remaining positive roots, these two are the simple roots.

Note that (8.35) also implies that the lengths of different roots cannot be arbitrary. In fact, simple roots can have at most two different lengths. As a consequence, the ones with the smaller length are called short root, and the other long roots. If all simple roots have the same length the algebra is called simply laced.

8.12 (Re)constructing the algebra

Since the simple roots contain all information about the other roots, it must be possible to reconstruct the entire algebra from it. This is indeed possible. To see how this works, follow the example of the su(3) case, where there are two simple roots.

First, the number of simple roots gives the dimensionality of the Cartan algebra, n. For su(3), this is $n = 2$.

Then, every simple root encodes one su(2) subgroup, satisfying the corresponding algebras (8.31-8.32). So there are at least n, implying at least $3n$ generators. For su(3) $3n = 6$. In addition, every positive root obtained from the simple roots satisfying (8.35) adds another two generators, in total 2(#simple roots$-$dim Cartan). For su(3), there is only one more such root, yielding in total $6 + 2 = 8$. The remaining algebra relations can be obtained from (8.27), as well as from the relations leading up to (8.35).

To see how to proceed, note

$$\begin{aligned} HE_\alpha|\beta\rangle &= H|[\alpha,\beta]\rangle = \sqrt{2}\frac{\alpha(\alpha+\beta)}{\alpha^2}|\alpha+\beta\rangle \\ &\Rightarrow [E_\alpha, E_\beta] = \frac{\sqrt{2}\alpha(\alpha+\beta)}{\alpha^2}E_{\alpha+\beta}, \end{aligned}$$

up to a phase, chosen by convention. The second equality comes from (8.27), and the first equality from the definition of the adjoint representation. Thus the vector with the sum of roots must be the same as the vector of the commutator, uniquely giving the commutator. The factor $\sqrt{2}$ appears as $\mathrm{tr}[E_\alpha, E_\beta]^2 = 2\mathrm{tr}E_\gamma^2$, and is thus required to keep a consistent normalization. Thus, for α and β simple roots, this creates the remaining relations. Especially for su(3), this implies

$$[E_{\alpha^1}, E_{\alpha^2}] = \sqrt{2}\left(1-\frac{1}{2}\right)E_{\alpha^1+\alpha^2} = \frac{1}{\sqrt{2}}E_{\alpha^1+\alpha^2}.$$

Since these are the only non-trivial commutators, every other commutator can then be obtained by expanding and using the base commutators (8.31-8.32).

8.13 A non-trivial example: $\mathbf{g}_2$

To exemplify the concepts, another non-trivial example is useful. This will be the Lie algebra g_2, which is important for many reasons. For now, it just pops up out of nowhere, but it will become evident how it comes about later.

This algebra is also a rank 2 algebra, and has thus two simple roots $\alpha^1 = (0,1)$ and $\alpha^2 = (\sqrt{3}/2, -3/2)$. The angle between both simple roots is $5\pi/6$, since

$$\alpha^1\alpha^2 = -\frac{3}{2}.$$

Also, the second simple root has now a length of 3, rather than 1 as before.

Because

$$\frac{2\alpha^1\alpha^2}{(\alpha^1)^2} = -3 \tag{8.42}$$

$$\frac{2\alpha^2\alpha^1}{(\alpha^2)^2} = -1, \tag{8.43}$$

the lowest states are at magnetic quantum numbers $-3/2$ and $-1/2$. Thus, because of (8.36), the corresponding states can be increased up to 3 and 1, yielding 7 and 3 individual states. It is possible to act with E_{α_1} three times as a raising operator before hitting eventually zero weight, and thus a Cartan (and afterwards negative roots), while this is possible only once with E_{α_2}. Thus, p_1 can be at most 3 and p_2 can be at most 1. Hence, in the adjoint representation, the system contains a spin-3/2 and a spin-1/2 representation of

two su(2). If the adjoint representation would be completely reducible, this would imply 21 generators. But it is irreducible, thus reducing the number of generators.

This can be seen as follows. All other roots are given by linear combinations of the simple roots in the general form

$$q_1\alpha_1 + q_2\alpha_2$$

for some positive integers q_1 and $q_2 \neq q_1$ (because no multiple of a root is a root), but by virtue of (8.36)

$$\frac{2\alpha^i(q_1\alpha_1 + q_2\alpha_2)}{(\alpha^i)^2}$$

must be integer. Inserting (8.42-8.43), this yields that there are only four combinations

$$\alpha^1 + \alpha^2 \qquad 2\alpha^1 + \alpha^2 \qquad 3\alpha^1 + \alpha^2 \qquad 3\alpha^1 + 2\alpha^2$$

which satisfy the equation, and are thus roots. Note that either $q_i = 0$ will not move into different su(2) subalgebra, and therefore is also not an option. This is a corollary to the fact that (non-trivial) multiple of roots are not roots. Thus, there are in total 2 (Cartan)+4 (simple roots)+8 (other roots)=14 generators in g_2, with a thus not so empty root diagram.

8.14 The Cartan matrix

As the example of g_2 showed, the actual relevant information are encoded in the number of times simple roots p_i and other roots q_i appear. Thus in any representation for every weight vector the Cartan for a given simple root applied to its state yields twice its J_3 eigenvalue due to (8.35),

$$2J_3|\mu\rangle = \frac{2H\alpha^i}{(\alpha^i)^2}|\mu\rangle = \frac{2\mu\alpha^i}{(\alpha^i)^2}|\mu\rangle = (q^i - p^i)|\mu\rangle$$

If in the adjoint representation μ is a positive root it can be written as a linear combination of simple roots $\mu = k_i\alpha^i$,

$$q^i - p^i = \frac{2\mu\alpha^i}{(\alpha^i)^2} = k_j\frac{2\alpha^j\alpha^i}{(\alpha^i)^2} = k_jA_{ji} \tag{8.44}$$

holds. Thus the information on the eigenvalue is contained in the entries of the matrix

$$A_{ij} = \frac{2\alpha^i\alpha^j}{(\alpha^i)^2}$$

called the Cartan matrix. The vector k describes the composition of the root in terms of simple roots. Due to the factor two and (8.36) all entries of A are necessarily integers. By construction the diagonal entries are all 2, and thus the non-trivial information resides in the off-diagonal entries. Due to (8.35), the only possible values of $p^i - q^i$ for positive roots are 0, -1, -2, and -3, and thus only these values can appear in the off-diagonal elements.

This information, as it is a product between different roots, implies how simple roots fit into the subalgebras described by the other simple roots. Of course, since these are just projections of the simple roots, which are linearly independent, A is invertible.

For the cases so far, the Cartan matrix of su(2) is one-dimensional, and therefore can have only the element 2. The others are

$$A_{\mathrm{su}(3)} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} \qquad A_{\mathrm{g}_2} = \begin{pmatrix} 2 & -1 \\ -3 & 2 \end{pmatrix},$$

confirming once more that the two su(2) subgroups of su(3) are equivalent, but for g_2 they are not: It is the number of times the lowering/raising operators can be applied to the other subalgebras without hitting zero. Thus, in the su(3) case, both simple-root su(2)s are in the spin 1/2 of the other one, while in the g_2 case, they are in different representations.

The Cartan matrix can then be used to simplify calculating the other roots, rather than using every time a geometric check by (8.35). Applying a simple root raising operator J_{α^i} to a root will raise it by one element in this simple root direction, i. e. $k_i \to k_i + 1$. This implies the eigenvalue, according to (8.44) is changed to

$$q^j - p^j + A_{ij}, \tag{8.45}$$

as the corresponding term is increased.

Now, for the simple roots, there is always only one k_i which is different from zero, and has value 1. This is called the $k = 1$ layer. Applying the corresponding lowering operator will reduce this by one, and thus all k_i are zero, which is thus a zero root, and hence belongs to the Cartan. It is therefore more interesting what happens when a raising operators is applied. Then, there is only one possibility, i. e. one other k_i is increased by one, since otherwise a non-trivial multiple of a root is obtained, which is not a root. Thus, now two k_i values are non-zero.

Now, consider the following. Every simple root corresponds to a line of the Cartan matrix. The diagonal element is just normalization, and therefore does not matter. The other one represents the lowest value the spin of the su(2) algebra can have in the corresponding representation of the other su(2) subalgebras. Since the simplest root are at the same time constructed to give, according to the derivation of (8.35), the lowest possible values for the corresponding representations, this implies that no other root can exceed these values. Thus, using (8.45), this permits to construct all other roots.

For su(2), this is trivial, since there are no other roots.

For su(3), start with the two simple roots. They have $(2, -1)$ and $(-1, 2)$ as the corresponding matrix rows. Now, no multiple of roots are roots, so the only possibility is to add the roots, yielding $\alpha_1 + \alpha_2$. This implies, according to (8.45), the lines, yielding $(1, 1)$. Since the two -1 indicate that the possible range is $\{-1, 0, 1\}$, this is acceptable, and therefore another root. Forming $\alpha_1 + 2\alpha_2$ yields $(0, 3)$ and $2\alpha_2 + \alpha_1$ yields $(3, 0)$. In both cases, one of the components exceeds the allowed range, and thus neither is a root. Since there is no other possibility to proceed, as every step corresponds to applying another raising operator, this is finished, and all positive roots for su(3) are constructed,

and the negative ones are obtained by inflection. This gave three positive roots, three negative roots, and two zero roots, obtaining the eight generators again.

For g_2, this is more complicated. The matrix rows are $(2,-1)$ and $(-3,2)$, implying the allowed ranges to be different, $[-3,3]$ for the first element and $[-1,1]$ for the second. The first is to add both simple roots, yielding $(-1,1)$, which is valid, and thus another root has been gained. Adding α_2 yields $(-4,3)$, and therefore this is not a root, as it is outside the allowed interval. $(1,0)$ from $2\alpha^1+\alpha^2$ is. Going on yields only one possibility, as $2\alpha_1+2\alpha_2$ is a non-trivial multiple of a root, and thus not a root. Hence, only $3\alpha_1+\alpha_2$ with $(3,-1)$ is obtained, and is valid. Then, $4\alpha^1+\alpha^2$ is $(5,-2)$ is again invalid. $3\alpha^1+2\alpha^2$ has $(0,1)$ and is valid. Since the only remaining options $4\alpha^1+2\alpha^2$ and $3\alpha^1+3\alpha^2$ are again non-trivial multiples, this terminates, and therefore, this created all the positive roots previously found.

Of course, this can be extended to negative roots by subtracting rather than adding, or just by inflection.

8.15 The fundamental representations

So far, most of this applied to the adjoint representation. But the (simple) roots can also be used to make statements about any other (irreducible) representation.

Choose a representation D. Then there are states of highest-weight μ such that

$$E_{\alpha^i}|\mu\rangle = 0$$

for any simple root α^i, and thus for any positive roots, which are only linear combinations of simple roots. In fact, the whole representation will be later constructible from applying combinations of $E_{-\alpha^i}$ to these states.

In (8.35), which was derived for states in arbitrary representation, this implies to set $p=0$, and thus

$$\frac{2\alpha^i\mu}{(\alpha^i)^2} = l^i, \tag{8.46}$$

where the l^i are, because of (8.35), (non-negative) integers. Since the α^i form a complete basis, the μ are uniquely determined by these scalar products. Conversely, every set of l^i defines the representation for which the μ are highest weight states. Since there are rank simple roots, every representation of a simple Lie algebra can be labeled entirely by a set of rank non-negative integers. These integers are called Dynkin coefficients. For the adjoint representation, these are given by the projection of the simple roots onto each other. Therefore, the Cartan matrix contains the Dynkin coefficients of the adjoint representation.

There are weight vectors satisfying

$$\frac{2\alpha^i\mu^j}{(\alpha^i)^2} = \delta_{ij},$$

from which all other weight vectors can be reconstructed by

$$\mu = l^j\mu^j$$

They therefore represent, in an abstract way, the base of the weight space. Note that though the μ^i are highest weights of an irreducible representations, the so constructed weights may not necessarily be irreducible, as they have contributions from different irreducible representations. But they will always contain some irreducible representations, which can be identified using the methods to be discussed in chapter 10.

Since there are rank different such sets of base weight vectors, they correspond to rank different irreducible representations. These are called the fundamental representations of the group. These representations need not be of the same dimensionality, nor need they to be different from other representations. As will be seen, the two fundamental representations of su(3) have both the same dimensionality, and are the lowest ones. For g_2, however, they have different dimensionalities, and one is actually identical with the adjoint representation, while the other has the lowest-possible dimension.

For su(3), the fundamental weights are created from the two vectors (a^i, b^i). Inserting them into (8.46) yields

$$\alpha^j \mu^i = (a^i \mp \sqrt{3} b^i)/2$$

for both simple roots. Appropriately normalized to obtain the fundamental weights yields

$$\begin{aligned} \mu^1 &= (1/2, \sqrt{3}/6) \\ \mu^2 &= (1/2, -\sqrt{3}/6). \end{aligned}$$

To obtain the other weights in the corresponding representation it is sufficient to use the simple roots. Since, by construction, this is a highest weight state, in this case it is necessary to subtract, rather than add them.

These fundamental weights are by construction highest weights, and therefore they correspond in terms of the Cartan matrix to the vectors $(1, 0)$ and $(0, 1)$. To obtain the other weights, it is sufficient to apply thus the lowering operators to them. The only linearly-independent ones are the simple roots, by construction, and thus only they can be used to obtain new weights. Any such created weight is then unique, and therefore the number of these states gives the dimensionality of the representations.

Thus, for μ^1, subtracting α^1 yields $(-1, 1)$, which is an admissible state in su(3), since the components are within the $[-1, 1]$ range. Subtracting α^2 yields $(2, -2)$, which is not. Subtracting $\mu - \alpha^1 - \alpha^2$ yields $(0, -1)$, which is again admissible. Any further subtraction does not work, and therefore the system of weights in this representation is complete. It is hence three-dimensional. Likewise, for the other highest weight, the sequence is μ^2, $\mu^2 - \alpha^2$ and $\mu^2 - \alpha^2 - \alpha^1$, again of dimension three, which due to the symmetries of the simple roots was to be expected. Note that if the application with the other single root would have given an admissible state, the representations would have been four-dimensional. At this point, this would be the maximum possible, since any further application of simple roots would necessarily at most create a known state, if admissible at all, since everything else is obtained by some linear combination of the simple root lowering operators.

Note that once these weights are all known, the remainder to construct the representation explicitly can be done following the outlined procedure for the adjoint representation. This is in general tedious, but unique.

Interestingly, the pattern of weight vectors created from μ^1 go into the weight vectors created by μ^2 under sign reversal. It is said they are conjugated to each other. The reason is that if the generators T^a fulfill the algebra, so do $-T_a^*$,

$$if_{abc}(-T_c)^* = (if_{abc}T_c)^* = ([T_a, T_b])^* = -[-T_a^*, -T_b^*],$$

because the structure constants are real. Given that T^a are in the representation D, then the representation created by the $-T_a^*$ is called the complex conjugate representations, and often denoted by $\overline{D}$. These have the same dimensionality. Since the Cartan generators are Hermitian, this only reverses the signs of its real eigenvalues, and therefore, due to (8.19), the weight vectors reverse the sign.

If there is no similarity transformation which relates the T^a and $-T_a^*$, the representation is called complex, and otherwise real. The representations are then not equivalent, and hence genuinely distinct. The fundamental representation of su(3) is complex. Thus, both fundamental representations are distinct.

For real representations, there is a further distinction. The condition for real representation is that there is some S such that

$$T_a = -ST_a^*S^{-1}.$$

But S can have either the property $S^{-1} = -S^T$ or $S^{-1} = S^T$, since in either case the generators are transformed, up to a minus sign, in the same way, and an additional minus sign, which also reverses the sign of the structure constants, does not spoil the algebra. If $S^{-1} = -S^T$, the representation is called pseudo-real instead of real. The actual difference is that in case of a real representation, the generators are unitarily equivalent to antisymmetric, purely imaginary matrices. Such generators transform under conjugation into itself, since T^a and $-T^{a*}$ coincide. In the pseudo-real case, this is no longer the case, but there is still a similarity transformation relating T^a and $-T^{a*}$.

The fundamental representation of su(2) is pseudo-real, since the Pauli matrices are not equivalent to purely imaginary, antisymmetric matrices, but $\sigma_a = -\sigma_2\sigma_a\sigma_2$ holds, and thus there exists at least a similarity transformation. For g_2, the fundamental representation is real, which is stated here without proof.

Because in the adjoint representation for every root also its negative is present, the adjoint representation is always real. This can be generalized. A representation which has the same set of Dynkin coefficients contains all elements twice, and is therefore always real.

Sometimes the notation is used that a representation is denoted by its dimensionality and therefore the complex conjugate by its dimensions overbarred, e. g. 3 and $\overline{3}$ for su(3). For higher representations, where different representations can have the same dimensionality, this is not always possible. E. g., for su(4), the fundamental representations are 4, $\overline{4}$, but the third one is 6, which is not a complex representation. For su(5), these are again four falling into two pairs of complex conjugated ones, 5, $\overline{5}$, 10, and $\overline{10}$.

Since the conjugation corresponds to a mapping of group elements into itself, it is an automorphism as discussed in section 5.5. If the representation is not-real, this cannot be undone by a similarity transformation, and therefore is an outer automorphism, while in the real case it is an inner automorphism.

8.16 Weyl group

A seemingly surprising feature of the weight space has been that they show a high amount of symmetries under reflections. This is actually not a coincidence, but originates from the fact that su(2) algebras and groups have a build-in reflection symmetry. Since the eigenvalue spectrum is symmetric under reflection at 0 - $(-1, 0, 1)$ goes over into $(1, 0, -1)$ - every su(2) subalgebra yields a reflection symmetry. This, of course, is inherited by the weight space. The set of all reflections under which the weight space is invariant is called the Weyl group.

For su(2), this is just a single reflection. But for su(3) in the adjoint representation, there are several reflections, as there are for g_2. Note that if there are different reflection planes, reflections can be combined to generate a discrete rotation group with reflections. E. g., the adjoint weights of su(3) are invariant under rotations by $2\pi/3$.

These reflections also show whether a complex conjugate representation is present. This is the case if two representations are identical under reflection.

8.17 Constructing other representations

To construct other representations, the starting point are always the Dynkin coefficients $(n_1, n_2, ...)$. The values n_i are then chosen to give the highest weights in the corresponding su(2) subalgebras. The other weights are then constructed by applying the lowering operators, i. e. subtracting the simple roots.

E. g., for su(3) the case $(1, 1)$ is actually the adjoint representation, as has already been seen before. For g_2, it would be $(3, 1)$.

To give examples of other representations, consider again the example of su(3). Start with some arbitrary positive integers, e. g. $(2, 0)$. Note that now one of the coefficients is actually larger than the largest one in the fundamental or adjoint representation. Inserting this into (8.46) yields

$$2\mu^1 = \left(1, \frac{1}{\sqrt{3}}\right)$$

In the Cartan basis $2\mu^1$ is $(2, 0)$. Given that the α_i read from the Cartan matrix are $(2, -1)$ and $(-1, 2)$ The possible sequence of weights is then $2\mu_1 - \alpha^1 = (0, 1)$, $2\mu_1 - 2\alpha^1 = (-2, 2)$, $2\mu_1 - \alpha_1 - \alpha_2 = (1, -1)$, as well as two more, which are the corresponding ones obtained by adding simple roots. Thus, there are 6 weights, and the representation is 6-dimensional. The shape in the weight-plane is a pyramid, and there is hence one non-trivial reflection, implying the presence of the conjugate representation $(0, 2)$.

Since it is also very important in physics, another representation should be mentioned, the ten-dimensional $(3, 0)$, as well as its complex-conjugate $(0, 3)$. This representation also shows another interesting feature. There are two weights, $(2, -1)$ and $(-1, 2)$, which are reached as $3\mu_1 - \alpha_1 - \alpha_2$ and $3\mu_1 - 2\alpha_1$, which yields under further application a weight with $(0, 0)$, $(3\mu_1 - 2\alpha_1) - \alpha_2$ and $(3\mu_1 - \alpha_1 - \alpha_2) - \alpha_1$. Since, despite appearance, this is not just adding vectors, but just a shorthand for applying non-commuting operators, the question arises, whether the so-reached weight appears actually twice or just once.

To answer the question, it is is necessary to explicitly calculate the difference, which is essentially given by the necessary terms introduced to bring one sequence into the other form, while acting on the highest-weight state. In this particular case, both sequences turn out to be identical, since any additional terms turn out to be either identical zero, or vanish when applied to the highest-weight state. There is no general principle to see this, and thus if weights can be reached by different ways, it is necessary to check in every case, whether the so-created weights are linearly dependent or not.

8.18 Dynkin diagrams

As is seen, simple roots are enough to reconstruct any compact Lie algebra. To simplify notation, it is therefore useful to develop a graphical language to encode simple roots. These are the so-called Dynkin diagrams.

Denote a simple root by a dot. Then the angle between two roots is denoted by the connection between them. No connection implies $\pi/2 = 90°$, $2\pi/3 = 120°$ a single line, $3\pi/4 = 135°$ a double line, and for $5\pi/6 = 150°$ a triple line.

The Dynkin diagram for su(2) is then just a single dot, as there is only one root. For su(3), there are two dots, with an angle of $2\pi/3$, and thus a single line. G_2 would then be two dots with three lines.

Chapter 9

Classification of Lie groups

Probably the most important result in Lie groups is that there is only a denumerable infinite number of them, which can be completely classified by their Dynkin diagrams. The reason is that (8.35) is a very restrictive statement.

9.1 General construction

To show this, it is best to separate the proof into many smaller steps.

The first step is to introduce the concept of decomposability. A root system (or algebra) is called decomposable, if it separates into two (or more) orthogonal subsets, i. e. into elements which mutually commute with every element of a different subset. This is just the case of semisimple Lie algebras. As a consequence, this implies that the Dynkin diagram of a simple Lie algebra must be simply connected, and any root cannot be orthogonal to every other root. An algebra, which cannot be decomposed, is called indecomposable.

As a consequence, every simple Lie algebra is already known to satisfy three constraints

i) The simple roots are linearly independent, as discussed in section 8.11

ii) Because of (8.35), if α and β are two distinct simple roots, $2\alpha\beta/\alpha^2$ is a non-positive integer

iii) The simple root system is indecomposable, as otherwise the Lie algebra would not be simple

Any system of vectors satisfying these three constraints is also called a Π-system. From these three constraints the complete classification follows. This will be constructive, and thus start with the least number of simple roots, and thus dots in the Dynkin diagram.

One node is su(2). Since one dimension has only one possibility for a linear independent vector, this exhausts all possibilities. Thus, su(2) is the smallest possible Lie algebra. hence, the decomposition in su(2) algebras of any algebra in section 8.8 is then in a sense a decomposition into the most simple building blocks.

Two can have one, two, or three lines connecting the two dots in their Dynkin diagram. A single connected line is su(3), and that with a triply connected line g_2. Remains the

one with two lines. This turns out to be the Lie algebra so(3) of the rotation group SO(3). This could be obtained from reconstructing the Lie algebra from the simple roots, as done in section 8.12, and then the group. This will be skipped here.

The next step is to notice that if there are only three vectors, there are only two admissible Π systems, corresponding to Dynkin diagrams being chains with at most one double connection. The geometric reason is that the angles enclosed by three vectors need to be less than 2π, because otherwise they are coplanar. Since (8.35) only admits a certain number of discrete angles, there are only two possibilities with three vectors satisfying this condition:

$$\frac{\pi}{2} + 2\frac{2\pi}{3} = \frac{11\pi}{6} < 2\pi$$
$$\frac{\pi}{2} + \frac{2\pi}{3} + \frac{3\pi}{4} = \frac{23\pi}{12} < 2\pi.$$

The other three structure of Dynkin diagrams, which can be created, all have an angular sum of 2π or more, and thus are linear dependent systems, and thus are not Π-systems. That has very far-reaching consequences, as any indecomposable subdiagram of a Dynkin diagram is again an indecomposable Dynkin diagram, and therefore a subalgebra. Therefore, only Dynkin diagrams will correspond to simple Lie algebras which have no other structure then the above will be found to be possible. In particular, the same geometric argument implies that only in the two-dimensional case, where coplanarity is acceptable for all independent roots, an angle of $5\pi/6 = 150^\circ$ is possible. Hence, g_2 is actually the only simple Lie algebra with triply connected nodes, making it rather special, just like su(2). Furthermore, this implies there is no triangle possible, as this will also create an angle in excess of 2π.

The next step returns to the construction of smaller Dynkin diagrams from larger ones, which are required to be Π-systems. Take a Dynkin diagram with a singly connected line. Replacing the two nodes attached to the line and the line with a single node, inheriting the previous connections of the two nodes, creates also a Π-system.

Having only a single line between the two simple roots α and β has the following implications. A single line implies an angle of $2\pi/3$. Since furthermore all simple roots are linear independent, then any other simple root γ connected by a line to either α or β has to have a vanishing scalar product with either α or β, but not with both. But then either $\gamma(\alpha+\beta) = \gamma\alpha$ or $\gamma(\alpha+\beta) = \gamma\beta$. Likewise, any simple root not connected to either α or β has to have a vanishing scalar product with both, and thus also with their sum. Thus, replacing α and β by $\alpha+\beta$, and thus removing a node, creates a new set of root vectors with one root vector less, which again satisfies all constraints, and keeps all the angles as before. It is thus again a Π-system. This implies that by removing single lines, it is possible to shrink the system further and further, until having eliminated all single lines eventually. However, any subset of three vectors will become linearly dependent if it includes more than one double line or a cycle, according to the previous step. Hence, any Dynkin diagram representing a Π-system can have at most one double line (if it consists out of more than two nodes), and can have no cycles, as otherwise it would be created from a non-Π system in the beginning by reversing the process of deleting lines.

The next step is gaining insight into what kind of possible ends can be attached to a Π-system. The previous argument ensures the possibility that a single node can be attached, by reversing the process. Furthermore, since it is not possible to append a triple line, the only two other options are a branching into two lines and the addition of a double line.

Consider now the option to append a branch at the last node. Call the two added nodes α and β. Since they are not connected by a line, they are orthogonal and $\alpha\beta = 0$. They are also connected by a single line to their anchor point, called γ. Thus

$$\frac{2\alpha\gamma}{\alpha^2} = \frac{2\alpha\gamma}{\gamma^2} = \frac{2\beta\gamma}{\beta^2} = \frac{2\beta\gamma}{\gamma^2} = -1.$$

These relations can be used to show that

$$\begin{aligned} \frac{2\gamma(\alpha+\beta)}{\gamma^2} &= -2 \\ \frac{2\gamma(\alpha+\beta)}{(\alpha+\beta)^2} &= -1. \end{aligned}$$

Thus, replacing the branch by a double line with node $\alpha+\beta$ provides again correct values satisfying (8.35). Thus, if a Dynkin diagram ending in a branch is a Π-system then so is the Dynkin diagram with the branch joined to a double line a Π-system.

This also implies that there is no possibility to have branches with more than three twigs. If there would be, they could be contracted to two double lines, which does not form a Π-system. For the same reason can a branch not occur at a node connected by a double line. Furthermore, even if the two branches are starting from different nodes, the Dynkin diagram can be shrunk until an offending diagram is reached. Thus, only Dynkin diagrams with at most one branching can represent Π-systems.

The next step requires to eliminate some special cases, which are not covered by the previous steps. In fact, only four cases have to be treated separately. These four cases are

- A diagram with a central node with three branches of length two nodes each
- A seven-node chain with a branching at the center node with a twig of one node
- An eight-node chain with a single branching at the third-last (or first) node with a twig of length one node
- A five node chain with a double line between nodes 3 and 4 (or 2 and 3)

All of these special cases cannot be shrunk to a contradiction using the previous rules.

However, it turns out that the angles prescribed between the simple roots in these configurations can only occur if the simple roots would be linear dependent, which is a contradiction to the requirement of being simple roots. Thus, these Dynkin diagrams cannot be Π-systems. This can be shown by explicit calculation, which is skipped here.

These results are sufficient to construct now all possible Π systems, and therefore all simple Lie algebras. The best route of action is to construct all possible Dynkin diagrams,

starting from a single node, and then adding lines in allowed ways. Here, these diagrams will be constructed, and then afterwards discussed in separate sections in more detail.

The simplest possibility is to just create a chain of single nodes. This is the simply laced algebra of special unitary matrices. Since there is an infinite number of such diagrams, this is an infinite family of these. They will be discussed in section 9.2.

The next possibility is to have a chain of singly-connected lines, but with a single double connected line at the end. There are now two possibilities, depending on whether the single doubly-connected node at the end is shorter or the others are. If the single one is shorter, this will be the algebra of the special orthogonal rotations in odd dimensions. They will be discussed in section 9.4. If the opposite is the case, these are the so-called symplectic group to be discussed in section 9.5.

Since any double line can be turned into a branch, there is always a corresponding diagram with a branch, which turns out to be the group of special orthogonal rotations in even dimensions. However, this group is then also simply laced. It will be discussed in section 9.3. The fact that even dimensions and odd dimensions of spatial rotations are described by different Lie algebras falls in line with other knowledge from physics that both cases are qualitative different.

All of these are again infinite families, since the singly-connected chain can be extended arbitrarily. These are the diagrams possible of this type.

Finally, there are five diagrams, which are possible but do not belong to any of the four infinite families. One is the already known group g_2 with two nodes of differing length and a triple connection. Then there is a four-node chain with a double connection in the middle, with the two nodes on either side of the double connection have a different length then on the other side. This algebra is called f_4. Then there are three more simply laced algebras, called e_6, e_7, and e_8, consisting out of a chain with a single branch to a twig of a single node at the third-last node. These so-called exceptional groups will be discussed in section 9.6.

All of these group appear in or the other way in physics, though with vastly different rates of occurrences.

9.2 Special unitary groups

Probably the most important algebras in quantum physics are the special unitary algebras su(N) of rank $N-1$, which have as fundamental representations the $N \times N$-dimensional special unitary matrices. Their Dynkin diagram is just a singly-connected straight line of dots.

Using the same normalization of

$$\mathrm{tr} T^a T^b = \frac{1}{2}\delta_{ab}$$

as before for su(2) and su(3), they can be formulated as a straight-forward generalization. Especially, the $N-1$ Cartan generators as generalization from the Gell-Mann form (8.18)

are given by $N \times N$ matrices

$$H^m_{ij} = \frac{1}{\sqrt{2m(m+1)}} \left(\sum^m \delta_{ik}\delta_{jk} - m\delta_{i,m+1}\delta_{j,m+1} \right).$$

The remaining generators are as before also Hermitian traceless matrices, and altogether there are $N^2 - 1$ generators. As a consequence, the defining representation is N dimensional.

To construct the remainder requires the N $N-1$-dimensional weight vectors, i. e. vectors of the N eigenvalues of the $N-1$ Cartan generators, given by

$$\mu^j_m = H^m_{jj} = \frac{1}{\sqrt{2m(m+1)}} \left(\sum_{k=1}^m \delta_{jk} - m\delta_{j,m+1} \right),$$

with no summation implied. Each has norm $(N-1)/(2N)$ and relative scalar product $\mu^i\mu^j = \mu^i_k\mu^j_k = -1/2N$. Thus, they exhibit the same regularity as expected from the Dynkin diagram, forming a regular shape, the so-called $N-1$-simplex, in the $N-1$-dimensional weight space. Since in this counting the nth weight has $n-1$ leading zero entries, it is often useful to chose an inverted ordering where the last positive component is used.

The regularity implies that the $N-1$ roots are obtained as differences between weights,

$$\alpha^i = \mu^i - \mu^{i+1},$$

which then are all of unit lengths and satisfy

$$\alpha^i\alpha^j = \delta_{ij} - \frac{1}{2}\delta_{i,j\pm1},$$

as implied by the Dynkin diagram. The fundamental weights are then given by

$$\phi^j = \sum_{k=1}^j \mu^k.$$

Especially in this form is μ^1 the highest weight of the fundamental representation.

This group is sometimes also called A_n in mathematics.

Except for SU(2), whose fundamental representation is pseudo-real, all other groups have complex fundamental representations.

9.3 Special orthogonal groups of even dimension

This set of algebras so($2n$) and groups SO($2N$) are connected to the rotations in $2n$-dimensional real spaces, well-known in physics. They have a Dynkin diagram very similar to the ones of su(N), just that they fork at the end into two elements, rather than just being a chain. In mathematics, they are sometimes called D_n.

The generators in the adjoint representation are traceless, purely imaginary antisymmetric $2n \times 2n$ matrices, like the example of so(2) in section 8.2 showed. The Cartan generators are given by

$$(H_m)_{jk} = -i(\delta_{j,2m-1}\delta_{k,2m} - \delta_{k,2m-1}\delta_{j,2m}),$$

which is a 2×2 matrix in the form of the second Pauli matrix embedded. Thus, eigenvectors and eigenvalues are just the embedding of those of this Pauli-matrix, i. e. the eigenvalue pairs ±1, together with the corresponding eigenvectors

$$\pm e^k_j = \delta_{j,2k-1} \pm i\delta_{j,2k}.$$

From this it is possible to read off the corresponding weight vectors $e^k_m = \delta_{km}$, as usual.

This creates the root vectors $\pm e^j \pm e^k$ for $j \neq k$, the positive roots $e^j \pm e^k$ for $j < k$ and the n simple roots $e^j - e^{j+1}$ for $j = 1...n-1$ and the cyclic one $e^{n-1} + e^n$.

Interestingly, the Dynkin diagram of so($2n$) cannot be reduced arbitrarily. The algebra so(8) is the last one which has a genuine distinct Dynkin diagram. Removing a further node yields three simply-connected nodes, which is the same as su(4). Thus, the algebras of so(6) and su(4) coincide. Furthermore, an attempt to remove the node in the middle breaks up the diagram into two single nodes. hence, so(4) is not simple, but has the same algebra as su(2)×su(2), which is of fundamental importance in relativity.

9.4 Special orthogonal groups of odd dimension

9.4.1 Generalities

It is at first a little bit surprising that the special orthogonal groups of odd dimension should be different than those in even dimension, but as will be seen later even and odd dimensions often induce quite different structures. As a consequence, this group is called in mathematics sometimes B_n.

In the case of Lie groups, this becomes obvious in the form of the Dynkin diagrams, as this group corresponds to a Dynkin diagram with a double connection between the last two elements of the chain, rather than two connections as is the case for even dimension.

It also becomes more clear when considering the structure of the Cartan generators. Before, they decomposed into 2×2 blocks, which is not possible in odd dimensions. In fact, the last one has eigenvalue zero, and is a zero on the diagonal. This was already visible in the case of SO(3) in section 8.2, even though in this case all generators had zeros on the diagonal, since this was not a Cartan basis. As a consequence, for the first $2n$ dimensions, the same structure arises as previously. The difference is for the additional odd one. This yields a surplus weight vector $\pm e^x$, another positive root e^x and another simple root e^n. Note that the root $e^{n-1} + e^n$ is therefore not simple, and hence the size of the Cartan is the same as in the corresponding so($2n$) case. This was also seen already for so(2) and so(3) in section 8.2, as also there in both cases only two generators appeared.

Note that the Dynkin diagram of so(3) is the same as that of su(2), and thus is the same algebra.

9.4.2 Spinor representation

The true difference for odd-dimensional rotation groups is seen in the spinor representation, which will also illustrate more on the relation between su(2) and so(3).

The, in a sense, natural representation of these algebras are in a $2n+1$-dimensional real vector space. However, there also exists a representation in a 2^n-dimensional vector space, and this is actually one of the fundamental representations. The reason comes from the one special simple root. There are in total n different fundamental representations, with fundamental weights

$$\begin{aligned} \mu^j &= \sum_{k=1}^{j} e^k \\ \mu^n &= \frac{1}{2}\sum_{k=1}^{n} e^k \end{aligned}$$

where the first $n-1$ cases are generated from $j = 1...n-1$. The last one is different due to the normalization. Using Weyl symmetry, the weights of the last fundamental representation can be constructed

$$\frac{1}{2}\left(\pm e^1 \pm e^2 \pm ... \pm e^n\right).$$

These weights are, by construction, transformed into each other by mirroring, and therefore the spinor representation coincides with its complex conjugate.

Of the weights there are therefore 2^n, and thus the space is 2^n dimensional (e. g. 2-dimensional for so(3) and 4-dimensional for so(5)). Such a space can be characterize as a tensor product of n two-dimensional vector spaces. Then any hermitian generator can be created from a tensor product of Pauli matrices. The Cartan generators are then

$$H_j = \frac{1}{2}\sigma_3^j,$$

where j indicates the two-dimensional space in which the Pauli matrix acts. In a similar way all other generators can be constructed from the Pauli matrices.

These representations now have a few specialties. Since they are self-conjugate, they need to be real. However, there are still two different kinds of real representations, real and pseudo-real. Due to the structure in terms of Pauli-matrices, it depends on the actual number of involved spaces whether the representation are real or pseudo-real. Explicit calculation shows that the spinor representation of $so(8n+1)$ and $so(8n+7)$ are real, and of $so(8n+3)$ and $so(8n+5)$ are pseudo real. Especially, so(3) has a pseudo-real spinor representation, just like su(2).

It is possible to construct spinor representations also for even n. However, they do then not form a fundamental representation. In a very similar calculation, it can be shown that the spinor representations of $so(8n)$ are real and for $so(8n+4)$ pseudo-real, while the other ones $so(8n+2)$ and $so(8n+6)$ are complex. The latter can already been seen from the isomorphism between so(6) and su(4), with su(4) having complex representations. As a consequence of the existence of spinor representations, representations of the $SO(n)$ groups, depending on n, can be real, pseudo-real, or complex.

9.4.3 Connection to Clifford algebras

The fact that there is something called a spinor representation already indicates that there is a relation to spins, and thus also to the Clifford algebra.

A Clifford algebra is an algebra with n generators γ_i satisfying

$$\{\gamma_i, \gamma_j\} = 2\delta_{ij} \tag{9.1}$$

which should not be confused with the case of the usual Clifford algebra where the Minkowski metric appears on the right-hand side. This version will be discussed later, as it is not connected to a compact Lie algebra.

Given an implementation of the Clifford algebra with n elements, define the operators

$$M_{ij} = \frac{1}{4}\left[\gamma_i, \gamma_j\right]$$

and with

$$\gamma = (i)^n \gamma_1 ... \gamma_{2l} \tag{9.2}$$

define

$$M_i = \frac{1}{4}\left[\gamma_i, \gamma\right].$$

It can then be shown that these operators can be used to create spinor representations of the Lie algebras, especially for so$(2l+1)$ with M_i and M_{ij} and for so$(2l)$ the two inequivalent spinor representations

$$\frac{1}{2}(1 \pm \gamma) M_{ij}$$

Thus, there is an intrinsic relation to the (Euclidean) Clifford algebra.

An explicit example is given by the spinor representation of so(3), which, since $l = 1$, is two-dimensional. This requires three generators, M_1, M_2, and M_{12}. Thus, the necessary Clifford algebra needs two generators. They are obtained from the two Pauli matrices σ_1 and σ_3, as they satisfy (9.1). Then $\gamma = -i\sigma_2$, and

$$\begin{aligned} M_1 &= -\frac{1}{2}\sigma_3 \qquad (9.3)\\ M_2 &= \frac{1}{2}\sigma_1 \\ M_{12} &= \frac{i}{2}\sigma_2. \qquad (9.4) \end{aligned}$$

This shows that again so(3) is isomorphic to su(2), and the spinor representation of so(3) is the fundamental representation (8.17) of su(2).

However, this still leads to a difference in terms of the group. The one marked difference between (8.17) and (9.3-9.4) the decisive difference is that the parameters of so(3) are then mapped to half the one of su(2), and thus to every element of so(3) there corresponds two representation matrices, differing by a sign, showing that this (and actually all) spinor representation of so(3) are double-valued. This will be taken up again in section 11.3.

9.4.4 Crystallographic subgroups

One of the probably most important applications of special orthogonal groups of odd dimensions are the crystallographic groups, especially of so(3). These are discrete subgroups, which are obtained from so(3) by requiring that only certain rotations are still possible.

Some cases are provided by the previous standard example of discrete rotations. E. g. choosing the only permitted rotation angles to be $\pi/2$ and its direct multiple leads to the crystallographic group of the cube, choosing $2\pi/3$ to the one of the hexagon. There are only very few such subgroups in 3 dimensions, which are closed. This yield the platonic bodies, the tetrad, the cube, as well as the bodies with eight (octagon), twelve (ikosaeder), and 20 (dodecahedra) surfaces.

This is not not the only possibility to create subgroups. Another possibility is to permit different rotation angles in different directions. This yields eventually the so-called point groups of possible crystal structures.

The same is, of course, possible in lower or higher dimensions, and not restricted to odd dimensions. E. g., in numerical calculations in particle physics the discrete subgroups, and their faithful and reduced representations, of so(4) play an important role, especially for fixed angles of $\pi/2$, the so-called hypercubic groups.

9.5 The symplectic group

The symplectic algebras appear at first sight somewhat strange, but is actually not unknown in physics, merely well hidden. In fact, it is closely related to the phase space of Hamiltonian mechanics. There, it was seen that of the $2n$ dimensions of the phase space n had in the Hamiltonian equations a different sign than the other n. This implied certain transformation properties for rotations in the phase space, which are encoded in the so-called symplectic groups sp($2n$) of even dimensions. The Dynkin diagram is the same as for so($2n+1$), but the length of the simple roots are different. This group is in mathematics sometimes called C_n. In fact, it is possible to consider this group as the norm-preserving rotations in a space of quaternions[1], rather than the spaces of real or complex numbers where the SO(N) and SU(N) groups act.

As a consequence of this structure, every sp($2n$) has as a subalgebra su(n). These are embedded in the $2n$-dimensional representation space as

$$\begin{pmatrix} T^a & 0 \\ 0 & -T^*_a \end{pmatrix},$$

where the T^a are the $n-1$ Cartan generators of su(n). Then, only one more Cartan generator exists, which can be written as

$$H_n = \frac{1}{\sqrt{n}}\sigma_3 \otimes 1$$

[1]These are 'numbers' of the type $a1+\vec{b}\vec{\sigma}$, where 1 is the 2×2 unit matrices and the σ_i are proportional to the Pauli matrices. This fact is actually the reason why su(2) is so special, as the quaternions are connected to the group SU(2).

and therefore generates a block-diagonal string of the third Pauli matrix σ_3. This Cartan generator necessarily commutes with all the su(n) Cartan generators. Thus, the Cartan is n dimensional.

An alternative characterization of the symplectic group is that it covers all $2n$-dimensional matrices M such that

$$M^T J M = J$$

where J is a block-diagonal matrix made from 2×2 blocks j being the matrix

$$j = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Another way of regarding the symplectic group is to characterize them as unitary matrices with quaternionic instead of complex entries.

The fundamental representations of symplectic groups are pseudo-real. All other representations are either real or pseudo real. Note that sp(2) has the same Dynkin diagram as su(2), a single node, and therefore both algebras agree, and thus also to the one of so(3). That all three algebras agree, but not the groups has to do with the possibility to have with the same algebra groups with different discrete subgroups, which will be discussed in section 11.3.

9.6 Exceptional groups

There are also the five exceptional groups, of which g_2 has already been encountered in section 8.13. Like the previous cases, the exceptional groups can be related to certain norm-preserving rotations on certain spaces. These spaces are formed by so-called octonions, a generalization of the quaternions, and actually the last extension possible. The reason is that the requirement for fields in which also a division operation and a length can be defined, so-called division algebras, is very restrictive, and there is only a certain limited number of cases possible, which are the real numbers, the complex numbers, the quaternions, and the octonions. Note that the multiplication for octonions, in contrast to the previous cases, is actually not associative.

This gives another perspective of why there is only a limited set of compact Lie algebras: There is only a limited set of division algebras and therefore a limited set of norm-preserving rotations in vector spaces with division algebras as scalars, and compact Lie groups are just the groups implementing these rotations.

To understand better the properties of exceptional algebras, it is useful to discuss the octonions in some detail. Like complex numbers and quaternions, octonions are generated by additional elements,

$$a + b_\alpha i_\alpha$$

where a and the b_i with $i = 1...7$ are real numbers, and the seven i_α generalize the imaginary units, or the Pauli matrices in case of the quaternions. Similarly to the quaternions, they fulfill the multiplication table

$$i_\alpha i_\beta = -\delta_{\alpha\beta} + g_{\alpha\beta\gamma} i_\gamma,$$

where the tensor g is totally antisymmetric with

$$g_{123} = g_{247} = g_{451} = g_{562} = g_{634} = g_{375} = g_{716} = 1$$

in a suitable basis, and all other elements related by index permutations or otherwise are zero. Note that this rule is not associative, since it implies, e. g.,

$$(i_1 i_2) i_7 = i_5 \neq -i_5 = i_1 (i_2 i_7)$$

and thus the octonions are not a group.

The fact that it is a division algebra is signaled by the fact that for the norm

$$|a + b_\alpha i_\alpha| = \sqrt{a^2 + b_\alpha^2}$$

$|AB| = |A||B|$ follows for any two octonions A and B, just like for real and complex numbers and quaternions.

However, they are related to the exceptional Lie groups in a non-trivial way. It starts with the fact that $g_{\alpha\beta\gamma}$ is actually an invariant tensor of g_2, and, as noted above, the other exceptional groups can be linked to rotations in spaces of octonions. However, this will not be detailed further.

An interesting feature of e_8 is that it is the only simple, compact Lie algebra for which the adjoint representation is also the lowest-dimensional representation, and thus again one of the fundamental representations. Note that in g_2 also one of the fundamental representation coincides with the adjoint, but this is not the lowest-dimensional one.

Except for e_6, which has one complex fundamental representation, the fundamental representations of all exceptional groups are real. This has the consequence that e_6 plays a special role in many physics contexts, where complex representations are often required.

9.7 Subalgebras and Dynkin diagrams

The Dynkin diagrams also offer a straightforward procedure to identify subalgebras. In this context, it is useful to define regular subalgebras as subalgebras where the Cartan generators are a subset of (linear combinations of) the original Cartan algebra. A regular subalgebra is called maximal, if it contains the full Cartan algebra of the original algebra.

Subalgebras can be created from Dynkin diagrams in various ways. The simplest is using any of the the reduction rules found in section 9.1. Since this guarantees another simple Lie group, this yields a regular subalgebra. However, since the number of nodes is smaller, this cannot be a maximal one. Furthermore, removing a node connected to more than one node will create also a Lie algebra, but since this splits the Dynkin diagram into two, this is no longer a single simple algebra, but a combination of two. Furthermore, since a node is missing, there is besides the Cartan algebra of the two new subalgebras a single Cartan generator, and therefore another u(1) algebra. E. g. the Dynkin diagram of $\mathrm{su}(N+M)$ can be cut into two by removing the Nth node. This yields a subalgebra of type $\mathrm{su}(N)\times\mathrm{su}(M)\times\mathrm{u}(1)$. The u(1) comes from the removed node, and the total rank is $N-1+M-1+1 = N+M-1$, as it should be.

The construction of maximal regular subalgebras is a little bit more involved. This is obtained in the following way. Since the Dynkin diagrams can be grown from a single seed, the lowest root, it is possible to construct a unique extended Dynkin diagram by adding to each simple root this lowest simple root. This is no longer a Π-system, since this system has now a linear dependence, as there is one more simple root. However, the removal of any simple root creates a new Π system. In terms of the Dynkin diagrams, it can be shown that this corresponds to creating a cycle for A_n, adding a branch to the singly-connected end node of B_n, adding another doubly-connected node at the other end of C_n, adding a branch at the other end of D_n, symmetrizing e_6 and elongating all other exceptional groups. Now maximal subalgebras can be generated from removing nodes from these extended diagrams. This, e. g., shows that so($2N$+1) has a so($2N$) subgroup, as expected from physics.

9.8 General Lie groups

The consequence of the previous sections is that since there is only a countable infinite number of compact, simple Lie groups, and thus there is only a countable infinite number of compact Lie groups.

These are direct product groups involving any of the simple Lie groups, additional Abelian U(1) factors (which generate generators with zero length according to the scalar product of section 8.3), and possibly discrete factor groups in form of center groups. The algebras are combinations of simple groups and the Abelian algebra.

In physics, all gauge theories of Yang-Mills type are of this kind. The only other gauge group encountered in physics are from non-compact groups in form of general relativity. This requires a theory of non-compact groups to be discussed in chapter 11.

9.9 Consequences for discrete groups

As indicated in section 9.4.4, at least some Lie algebras/groups do have also discrete subalgebras/groups, which can necessarily be enumerated. This can be actually generalized. It can be shown that for compact, simple, discrete groups there are only two possibilites. Either they are a discrete subalgebra of a continous Lie algebra, or they belong to a very small, and finite, number of special cases. For those which are discrete subalgebras of the continous ones, similar consequences arise as in section 9.4.4. This yields little new in terms of group theory, and will thus not be pursued here, although these are in practice often very important.

The remaining are 26 so-called sporadic finite groups. They are not subgroups of continous Lie groups. Out of these, twenty are subgroups of a single group, the so-called Monster group. The other six, know as pariahs, are different. The monster group is the largest one, with about 10^{54} elements. This group has actually a connection to function theory, and a very special function, the so-called j function. This is known as the monstrous moonshine. This function is also connected to so-called conformal field theory,

which therefore establishes a connection to physics. There are indications that also the other five pariahs may be connected similarly to function theory.

The non-pariah sporadic groups are separated in three generations. The first generation, the Mathieu groups, are connected to the permutation groups of n points, and consists out of 5 groups. The second generation, called the Conway groups, contains 7 groups and is connected to the automorphisms of a special lattice, the so-called Leech lattice, in 24 dimensions. The third generation contains the other 7 groups and the monster itself, and are characterized by their relation to the monster.

Chapter 10

Tensor products

It is a rare situation in physics to have only a single symmetry with a single representation characterizing a system. If multiple symmetries start to play a role, this leads to the concepts of tensorizing representations, frequently of different groups. This is not limited to compact Lie groups, but it is possible for any groups, no matter whether continuous or discrete.

10.1 Tensorizing states

One particular important situation in physics is that some state is characterized by different representations of two different groups or even the same group. An example is given by atoms with electrons with spin. Then the electrons carry two representations, one of SU(2) and one of SO(3), together with corresponding quantum numbers. The SU(2) quantum number is related to the spin of the electron, and the SO(3) quantum number is related to its orbital angular momentum. However, in fact both originate with the Lorentz group, and eventually the problem is that the electron needs to be characterized by two different representations of the same group.

In a more general language, a state which belongs to two different representations, possibly of two different groups, can be considered as a product state

$$|j,k\rangle = |j\rangle|k\rangle,$$

where j and k may also be sets of quantum numbers of some group D_1 and D_2. This is called a tensor product of the states. If the representation of D_1 is n-dimensional and of D_2 is m-dimensional, the total dimensionality is nm. This can be thought of as that k runs over all possible values for any fixed value of j and vice versa.

The matrix representation of the groups act then in a tensor representation,

$$D^{1\otimes 2}|j,k\rangle = (D^{1\otimes 2})_{iljk}|j,k\rangle = (D^1)_{ij}|j\rangle(D^2)_{lk}|k\rangle. \tag{10.1}$$

Thus, this is the statement that each state transforms in its own representation.

Expanding for continuous groups close to the identity this yields

$$(1_{1\otimes 2} + i\alpha^a X_a^{1\otimes 2}) = (1_1 + i\alpha^a X_a^1)(1_2 + i\alpha^a X_a^2) = 1_{1\otimes 2} + i\alpha^a(X_a^1 1_2 + 1_1 X_a^2),$$

where the index at the 1-operator indicates in which space the unit operators act. Leaving the unit operators implicit, this amounts to $X_a^{1\otimes 2} = X_a^1 + X_a^2$, and thus the generators add in a tensor product. Especially, if the states are eigenstates to the generators, this implies that the eigenvalues add. It should be noted that this notation is very compact, as a has to run over the total number of generators $N_g^1 + N_g^2$ of both groups. Formally, the generators of the tensorized group are enumerated such that, e. g. for $N_g^1 = N_g^2 = 2$ this should be read as

$$X = (X_1^1 1_2, X_2^1 1_2, 1_1 X_1^2, 1_1 X_2^2),$$

and thus the explicit scalar product yields

$$\alpha^a (X_a^1 1_2 + 1_1 X_a^2) = \alpha_1 X_1^1 1_2 + \alpha_2 X_2^1 1_2 + \alpha_3 1_1 X_1^2 + \alpha_4 1_1 X_2^2,$$

such that the parameter vector is also $N_g^1 + N_g^2$ dimensional. Alternatively, this can be regarded as

$$e^{i\alpha_a X_a^1} \otimes e^{i\beta_b X_b^2} \approx 1_{1\otimes 2} + i\alpha_a X_a^1 1_2 + i 1_1 \beta_b X_b^2,$$

for α and β both infinitesimal, to make the relation better manifest. This generalizes to tensor products of more than two representations in a straightforward way.

10.2 Clebsch-Gordon construction

While if the generators stem from different groups, little further can be done. This changes, if the tensor product is created from two representations of the same group. Then a tensor product will create a, usually completely reducible, representation constructed from the representations of the lower-dimensional spaces. This not only appear similar to the reconstruction of the representation from the Cartan matrix. Indeed, taking this approach yields a complete highest-weight construction for higher-dimensional representations. This is called the Clebsch-Gordon construction. In fact, this starts again from tensor products of su(2).

For this note first that

$$J_3^{1\otimes 2} |j j_3, j' j_3'\rangle = (j_3 + j_3') |j j_3, j' j_3'\rangle. \tag{10.2}$$

Thus tensorization can give at most a representation with $j^{1\otimes 2} = j^1 + j^2$. Representations with even higher j can then be obtained by successive tensorization.

As an example, consider to get $j = 3/2$. Consequently, it should be possible to obtain it from tensorizing a $j = 1$ representation and a $j = 1/2$ representation. The $j = 1$ representation can itself be obtained by tensorizing two $j = 1/2$ representation, but since the $j = 1$ representations are known from orbital angular momentum in quantum mechanics this step will be skipped.

In the spirit of the highest-weight construction, the state with $j_3 = 3/2$ is, by virtue of (10.2), necessarily given by

$$|3/2, 3/2\rangle = |1, 1\rangle |1/2, 1/2\rangle = |1, 1/2; 1, 1/2\rangle,$$

where the second notation $|jj';j_3j_3'\rangle$ will be used for brevity.

To obtain the other states, it will be just proceeded as in the highest-weight construction by applying the tensorized lowering operators, creating

$$\begin{aligned}
|3/2,1/2\rangle &= \frac{1}{N_{1/2}}J_-^{1\otimes 2}|3/2,3/2\rangle = \frac{1}{\sqrt{3}}|1,1/2;1,-1/2\rangle + \sqrt{\frac{2}{3}}|1,1/2;0,1/2\rangle \\
|3/2,-1/2\rangle &= \frac{1}{N_{-1/2}}J_-^{1\otimes 2}|3/2,1/2\rangle = \sqrt{\frac{2}{3}}|1,1/2;0,-1/2\rangle + \sqrt{\frac{1}{3}}|1,1/2;-1,1/2\rangle \\
|3/2,-3/2\rangle &= \frac{1}{N_{-3/2}}J_-^{1\otimes 2}|3/2,1/2\rangle = |1,1/2;-1,-1/2\rangle,
\end{aligned}$$

where the N_i are the normalizations from the tensorized operators. This represents a four-dimensional sub-space of the six-dimensional space obtained by tensorizing the two representations. As this is a complete irreducible representation this shows that indeed the tensor product lead to a reducible representation. But this also implies that there is another subspace.

The basis for the other two directions is given by orthogonalization to these, e. g. using the Gram-Schmidt procedure. A possible set of base vectors are

$$\begin{aligned}
|1/2,1/2\rangle &= \sqrt{\frac{2}{3}}|1,1/2;1,-1/2\rangle - \sqrt{\frac{1}{3}}|1,1/2;0,1/2\rangle \\
|1/2,-1/2\rangle &= \sqrt{\frac{1}{3}}|1,1/2;0,-1/2\rangle - \sqrt{\frac{2}{3}}|1,1/2;-1,1/2\rangle,
\end{aligned}$$

where the naming is motivated by the fact that they can be obtained from each other by applying $J_\pm^{1\otimes 2}$, and form a representation to $j = 1/2$. Thus, the tensor space carries a completely reducible representation, and furnishes a $J^1 = 3/2$ and a $J^2 = 1/2$ irreducible representation. That the total representation is reducible was expected, as it is just the tensor product of two irreducible representations. However, this permitted nonetheless to construct a new irreducible representation of higher dimension, spin 3/2.

This situation is often denoted in short as

$$J^{1\otimes 2} = J^1 \otimes J^2 = j \oplus j'$$

or in the particular case above

$$1 \otimes 1/2 = 3/2 \oplus 1/2$$

to denote which kind of representations can be derived from tensorizing two representations.

Note that by taking the norm on both sides of the expressions, and using the fact that all states are normalized, their overlaps are given by

$$\langle J^1 J_3^1 | j, j'; j_3, j_3'\rangle = c_{J^1 jj' J_3^1 j_3 j_3'} \tag{10.3}$$

are obtained. These are called Clebsch-Gordan coefficients. They can be found tabulated for low-dimensional case as well as implemented in computer algebra systems, as they can be determined algorithmically. E. g. $c_{3/211/21/21-1/2} = \sqrt{1/3}$. Thus, these coefficients are encoding purely the group-theoretical structure of tensor products, and can therefore be algorithmically calculated. In particular, this is the situation to calculate the total spin of an atom with a spin 1 nucleus and an electron in s-wave, but did not require any knowledge of the electromagnetic interaction. But from this can be deduced that there are two possible states the atom can have, one in which the spins are aligned, and one, in which this is not the case. This will dictate its reaction, e. g. to an external magnetic field. Generalizing this insight will be done in section 10.4.

It is an interesting situation to consider the case of tensorizing twice the same representation, e. g. twice $s = 1/2$. Then

$$|1,1\rangle = |1/2,1/2;1/2,1/2\rangle.$$

This state is necessarily symmetric when the two representations making it up are exchanged. This is a generic feature. Since at the highest weight this is a combination of twice the same state, this is necessarily a symmetric combination. The raising and lowering operator commute with the exchange operator P_{12} for the two representations

$$P_{12}J_{\pm}^{a\otimes a} = P_{12}(J_{\pm}^{a}\otimes 1 + 1\otimes J_{\pm}^{a}) = (1\otimes J_{\pm}^{a} + J_{\pm}^{a}\otimes 1) = (J_{\pm}^{a}\otimes 1 + 1\otimes J_{\pm}^{a})P_{12} = J_{\pm}^{a\otimes a}P_{12},$$

and therefore the symmetry of the state created by the descent operation preserves the symmetry. Hence, all the states remain symmetric.

This is no longer true once the orthogonal other representation is constructed. E. g. for $J^1 = 1$ the orthogonal states form a $J^2 = 0$ representation,

$$|0,0\rangle = \frac{1}{\sqrt{2}}(|1/2,1/2;1/2,-1/2\rangle - |1/2,1/2;-1/2,1/2\rangle)$$

which changes sign under the exchange of the first and the second representation. Which symmetry it has is a less generic statement, as there may be several lower representations involved. However, they all keep the symmetry of their respective highest-weight state.

Note that the fact that this is a tensor product was important for this argument. If a $j = 1$ state is not made up as a tensor product, but has intrinsically this value of j, its symmetry properties may be different. In physics, this translates to the fact that composite states made from other particles inherit properties, but elementary particles of the same spin may not need to have the same intrinsic features.

It is quite instructive to understand how this procedure can be used to reconstruct representations of algebras other than su(2) from the fact that they are made from su(2) subalgebras. To this end, consider su(3), which has three su(2) subalgebras. But only two of them are complete, as there are only two Cartan generators. The remaining generators are thus necessarily mixing the two subalgebras. As was noted in section 8.14, the root system is created symmetrically from both su(2) subalgebras.

What happens is the following. In the fundamental representation, there is an su(2) representation in the first two dimensions made from the generators J_1, J_2, and J_3, with

J_3 has the corresponding eigenvalues. At the same time, J_3 is also the Cartan generator of su(2) in the adjoint representation. The other ladder operators will move always between two adjacent dimensions. Likewise, the last Cartan distinguishes the latter two dimensions. Thus, what has happened is that this is reducible, but not completely irreducible, representation of the tensor product of two $j = 1/2$ representations of su(2).

The concepts of the Clebsh-Gordan coefficients generalize. However, they become more involved. For each Cartan element a highest weight exists, with a separate counting variable each. Thus, a state of, say, su(3), will be identified as $|jj_3kk_3\rangle$. The j and k represent the highest weight for each of the Cartan elements. The three set of ladder operators will move the corresponding eigenvalues around.
The lowest dimensional representation is three dimensional, but have $j = 1/2$ and $k = 1/2$. Using the Gell-Mann matrices to define $H_1 = J_3$, $H_2 = J_3 - J_8/\sqrt{24}$ and the the other three pairs to create ladder operators, it follows that the three Cartesian vectors correspond to the states

$$\begin{aligned} e_1 &= \left|\frac{1}{2}0\right\rangle \\ e_2 &= \left|-\frac{1}{2}-\frac{1}{2}\right\rangle \\ e_3 &= \left|0\frac{1}{2}\right\rangle \end{aligned}$$

where the explicit reference to j and k have been dropped. Note that this implies that there are two subspaces of two dimensions, in which the H_i act as a $1/2$ representation, and in the orthogonal subspace they act trivially, indicated by the 0. Two of the ladder pairs act in these subspace, while the third pair connects both, mapping e_1 to e_3, and vice versa. They therefore exchange the highest weight states.

Constructing a tensor product now works as before, but has the additional complication that the actual tensorization is happening in the three-dimensional space, which in this case does not coincide with a single irreducible representation. Thus, consider the case of a tensor product of the two fundamental representations. It will be nine-dimensional. This can be decomposed, in the same way as before for the su(2) case, into an eight-dimensional adjoint representation and a one-dimensional trivial representation. The trivial ones has $j = k = 0$. The adjoint one has $j = k = 1$. Again, one of the states, needs to be shared, e. g. $j_3 = k_3 = -1$

$$|1 - 11 - 1\rangle = |1/2 - 1/21/2 - 1/2\rangle|1/2 - 1/21/2 - 1/2\rangle.$$

Thus, as before, starting from this shared state, all others can be constructed by application of the tensorized operators, $J_\pm^{i\otimes j} = j_\pm^i \otimes j_\pm^{j'}$, yielding the states, and likewise eventually the Clebsh-Gordan coefficients by taking scalar products. Note that out of these operators not all will act linearly independently on the vectors, due to the mixing of the su(2), such that in the end again only six will appear. Of course, as this the adjoint representation, the operators could also be directly reconstructed from the structure constants. Furthermore, the one connecting the two su(2) will now also create a connection not only for the highest weight states. However, the whole process becomes quickly very tedious.

10.3 Tensor operators

Of particular importance for physics are sets $l = 1...n$ of operators O_l^s which satisfy the algebra

$$[J_a, O_l^s] = O_m^s (J_a^s)_{ml}, \tag{10.4}$$

with no summation over s. This notation should be read as that the commutator of the group generator J_a with O_l^s yields a linear combination of the operators O_l^s weighted by the matrix elements of J_a in the representation s. That appears a first sight to be rather special. But these kind of operators are quite common in physics. Operators satisfying this are called tensor operators. Besides their importance for physics, they will be very useful in constructing further properties of groups. It is therefore worthwhile to investigate them a little closer. Note that the J^a can, e. g. be the linear operators acting in the whole vector space, which may carry multiple representations. More details of it requires specialization.

To give an explicit realization consider one-particle quantum mechanics. Angular momentum is then given by the $j = 1$ representation $J_a = \epsilon_{abc} r_b p_c$ of so(3), and take the position operator r_i. This yields

$$[J_a, r_b] = \epsilon_{acd}[r_b p_d, r_b] = -i\epsilon_{acb} r_c = r_c (J_a^1)_{cb}$$

where in the last step the adjoint representation was identified for the matrix elements $(J_a^s)_{ml}$. Thus, the vector r is forcing the $j = 1$ representation for the coefficients, and thus carries $s = 1$, $O_i^1 = r_i$.

It should be noted that the resulting matrices J_a^s are not necessarily the ones obtained using the highest-weight representation. For technical reasons, however, it is usually better to have the J^s in the explicit highest-weight representation. To achieve this, it is best to find some linear combination O_0^s of the O^s such that

$$[J_3, O_0^s] \sim O_0^s,$$

since then

$$[J_3, O_0^s] = O_l^s (J_3^s)_{lj},$$

and this collapses to a δ-function multiplied with a constant, since J_3 is diagonal. For the position operator, this is J_3 with $[J_3, r_3] = 0$, and thus $r_3 \sim r_0$. The remaining operators can then be constructed using multiple applications of the commutation relations, e. g.

$$[J^{\pm}, r_0] = \mp \frac{1}{\sqrt{2}} (r_1 \pm i r_2).$$

and thus the explicit form.

A word of caution here when indeed it turns out that the representation for J is found to be reducible. It is then possible to decompose the operators O_l also into sets of operators such that each correspond to an irreducible representation of J. However, there is no general construction principle how to do so, since for operators there is no notion of orthogonality. Thus, the best approach is to first find the linear combinations of operators which commute with J_3, to identify the present j. This is the complicated step, as there

is no construction principle available. Then the other operators can be obtained using the raising and lowering operators once more.

Finally, if the operator O is a tensorized operator itself, i. e. a product of two tensor operators in two representations, then it will create an algebra with the tensorized J. So, for O^s_m and O^r_i

$$[J_a, O^s_m O^r_i] = [J_a, O^s_m] O^i_n + O^s_m [J_a, O^i_n] = ((J^s_a)_{mn} \otimes 1_{ij} + 1_{mn} \otimes (J^r_a)_{ij}) O^s_n O^r_i,$$

with the special case

$$[J_3, O^s_m O^r_i] = (m+i) O^s_m O^r_i.$$

10.4 Wigner-Eckart theorem

This can now be used to construct the generalized Wigner-Eckart theorem, a central theorem for the separation of group structure and dynamics. It assumes that the tensor operator in the relevant basis is known. Then it states

$$\langle j', j'_3, \alpha | O^s_l | j, j_3, \beta \rangle = \delta_{j'_3, l+j_3} \langle j, l+j_3 | s, j, l, j_3 \rangle \langle j', \alpha | O^s | j, \beta \rangle. \tag{10.5}$$

where α and β are all other quantum numbers. Visible is the appearance of the Clebsch-Gordan coefficients (10.3). This states that the dependence of the matrix element on j_3, j'_3, and l is trivial and determined solely by the group structure. All the dynamics is completely encoded in the so-called reduced matrix elements $\langle j', \alpha | O^s | j, \beta \rangle$, which does not depend on j_3, j'_3, and l. Of course, this may not be too useful in practice for actually calculating the reduced matrix element, but separates conceptually kinematical effects from dynamical effects. Especially, it implies that if O is a tensor operator, its expectation values are $2s+1$-times degenerate, as there are $2s+1$ different matrix elements on the left-hand side, but only one on the right-hand side, up to cases where Clebsch-Gordan coefficients vanish.

To prove this theorem, notice that in the relation for tensor-operators (10.4) on the right-hand side only the operator itself appears, but the actually value of the J are matrix elements, and thus just numbers. Especially

$$\begin{aligned} J_3 O^s_l |j, j_3, \alpha\rangle &= [J_3, O^s_l] |j, j_3, \alpha\rangle + O^s_l J_3 |j, j_3, \alpha\rangle \\ &= O^s_{j_3 j'_3} (J^s_3)_{j_3 j'_3} |j, j_3, \alpha\rangle + O^s_l (J^j_3)_{j_3 j'_3} |j, j_3, \alpha\rangle = (l + j_3) O^s_l |j, j_3, \alpha\rangle \end{aligned}$$

Acting on this expression now from the left with the corresponding state, J_3 can be applied to it. However, since the representation of J_3 is not explicitly given, this still requires to transform it into an abstract basis. This is done by inserting a unity in the representation s. Then these are the expectation values of J_3 between the two, but they can be translated into each other using the Clebsch-Gordan coefficients, which then finally creates the Wigner-Eckart theorem (10.5).

10.5 Invariant tensors

It is possible to combine the structure of Lie groups based on weights with the tensor notation.

Since a representation has as many dimensions as its has rank-dimensional weight vectors, it is possible to find a basis in which the base vectors correspond to a weight vector. These are the eigenvectors to the Cartan generators, which agree since Cartan generators are simultaneously diagonalizable. The weight vectors are, after all, the eigenvalues of the Cartans for these eigenvectors. Thus, there is a one-to-one correspondence

$$\mu = e_\mu,$$

where the $e_\mu = e$ are the usual Cartesian unit vectors, where the index enumerating the associated weight vector will be suppressed.

An arbitrary generator T^a in this basis will have a matrix representation $(T^a)^i_j$. The position of the indices will become clear soon, but is not coincidentally reminding of the situation in relativity. It should be noted that the range of indices is that of the weights of the corresponding representation. E. g., if this is su(3) and the representation is the fundamental one, the indices run from 1 to 3, while in the adjoint one it would run from 1 to 8. For su(2), it would be 1 to 2 and 1 to 3, respectively. That is important, as this, e. g. limits the number of elements in anti-symmetric quantities in the indices. For su(2) in the fundamental representation, there can be no totally antisymmetric object with three or more indices, as with two different values for the indices there is no possibility to have all three indices different. Furthermore, indices belong to a fixed irreducible representation, and it is not possible to, say, contract indices of different irreducible representations, as they act in different subspaces. Acting now with T^a on a base vector yields

$$T_a e = (T^a)^j_i e_j,$$

implying a contraction over two indices.

Take now the complex conjugated representation. It lives in the same space, and it will have again a suitable basis of vectors corresponding to the weight vectors. These are in general different from the e. Denote these base vectors by $\bar{e}$ with components e^j to distinguish them from the e and their components e_i. In general, there is a base in which these are again the usual Cartesian unit vectors, but not in the one where these correspond to the original representation, except when both representations are identical. Then

$$T_a \bar{e} = (-T^a)^\dagger \bar{e} = -(T^a)^i_j e^j$$

and they therefore transform with a minus sign and a contraction over the other index.

Such states can be tensorized, yielding tensors $t^{ijk\ldots}_{abc\ldots} = A^i B^j \ldots T_a S_b \ldots$. Acting on them with a generator implies, just as in (10.1), an action on every index, by definition

$$T_z t^{ijk\ldots}_{abc\ldots} = \sum_l \left((T_z)^m_{n_l} t^{ijk\ldots}_{ab\ldots n_{l-1} m n_{l+1}\ldots} - (T_z)^{n_l}_m t^{ijk\ldots n_{l-1} m n_{l+1}\ldots}_{abc\ldots} \right)$$

and thus the generators act on all indices simultaneously. An object acting on only one index requires in this convention an explicit definition. The number of indices is the rank of the tensor, as usual. The tensor of highest weight for any rank is the one constructed from tensorizing the single states of highest weight. Such a state is necessarily symmetric in exchange of upper and lower indices within each other, as the combination is symmetric. It also vanishes for any contraction of a upper and lower index, because the necessary Kronecker-δ δ^j_i has the property $\delta^i_j = -\delta^j_i$, since its is also constructed from the (traceless) generators. The features are conserved under application of the raising and lowering operators, and are therefore shared by all states, which are created from this highest-weight state.

Note that it is possible to define tensors, on which the generators act to the left, in principle the usual bra-states. Since they are transformed with the inverse, this implies that they will have the position of indices reversed.

This construction can now be used to characterize the concept of invariant tensors t, i. e. tensors which are invariant under the action of the group/algebra,

$$T_a t = 0$$

Because of the tracelessness of the generators, all algebras have the unit tensor δ^j_i in common,

$$(T_a\delta)^i_j = (T_a)^i_k\delta^k_j - (T_a)^k_j\delta^i_k = (T_a)^i_j - (T_a)^i_j = 0$$

Another always present invariant tensor, again due to the tracelessness, is the Levi-Civita tensor

$$(T_a\epsilon)^{ijk} = (T_a)^i_l\epsilon^{ljk} + (T_a)^j_l\epsilon^{ilk} + (T_a)^l_k\epsilon^{ijl} = (T^a)^l_l\epsilon^{ijk}.$$

This follows, as always all three indices have to be different, and thus whenever there are two different indices, this yields the trace component, yielding thus zero in total. The same is true for ϵ with all indices down, for the same reason. Note that the contraction with an invariant tensor can be used to raise and lower indices, or to map a tensor of a certain rank to another rank.

There may also be more invariant tensors of different rank. If they cannot be decomposed in these invariant tensors, they are usually a characteristic property of the group. The first example is the Casimir operator defined as

$$T^2 = T^a T^a$$

a concept known from spins: This is the total spin. This operator commutes with all generators,

$$\left[T^2, T^a\right] = T^b\left[T^b, T^a\right] + \left[T^b, T^a\right]T^b = if^{abc}\left\{T^b, T^c\right\} = 0 \tag{10.6}$$

and is therefore an invariant tensor. Because it commutes with all the generators, it has to be proportional to the identity, and the coefficient is called the Casimir (operator). In a fixed representation j the Casimir operators is given

$$\begin{aligned} C_j &= g_{ab}\sigma^a\sigma^b \\ g_{bc} &= \mathrm{tr} f_b f_c \end{aligned}$$

where the σ_i are the generators in the representation j, and f are the generators in the fundamental representation, and the value therefore depends on the convention for the normalization of the scalar product of two generators in (8.11). An often useful normalization is $C_F = 1$. This normalization fixes the other Casimirs, but also often different values are assigned to the fundamental Casimirs for different fundamental representations.

An interesting quantity, which can be derived from the Casimirs, is called the index of a representation j, and defined as

$$I_2(j) = \frac{d(j)C_j}{d(F)C_F}$$

where $d(j)$ gives the dimension of the representation j. Except for the spinor representation of the special orthogonal groups of dimension 6 or less, the index is always an integer. This particular combinations appears in many practical calculations.

One further example is the totally antisymmetric tensor d^{abc} defined as

$$d^{abc} = \mathrm{tr} T^a \{T^b, T^c\} \tag{10.7}$$

which is an invariant tensor of su(3). However, this expression actually vanishes for both su(2) and g_2, and therefore this tensor does not exist in these two cases. It can be shown that it is, in fact, only non-zero for su($N > 2$).

This can be generalized to arbitrary rank by

$$d^{i_1 \dots i_n} = \frac{1}{n!} \sum_{\text{permutations } P} \mathrm{tr} T^{P(i_1)} \dots T^{P(i_n)}$$

where the sum is over all permutations. This can also be regarded as the case of a generalization of the Casimir concept, by denoting

$$\begin{aligned} C_j^k &= g^{a_1 \dots a_k} \sigma_{a_1} \dots \sigma_{a_k} \\ g^{a_1 \dots a_k} &= \frac{1}{k!} \sum_{\text{permutations } P} \mathrm{tr} T^{P(i_{a_1})} \dots T^{P(i_{a_k})} \end{aligned}$$

and thus the total symmetric tensor (10.7) leads to a third-order Casimir.

The invariant tensors play an important role in physics. If a quantity of a certain tensor rank should be constructed which is invariant under the action of the algebra or group, it must be necessarily a linear combination of the invariant tensors of the same rank. Thus, in su(3) any invariant tensor of rank 2 is proportional to the unit matrix, while for rank 3 it is a linear combination of ϵ^{abc} and d^{abc}. This is actually true also beyond compact Lie groups, and is the origins of the tensor decomposition. Together with the statement that the number of invariant tensors for a finite rank is finite, this is known as the primitiveness hypothesis. There is, to the knowledge of the lecturer, no full general proof of it, but it appears to apply to any case relevant in physics.

Moreover, this primitiveness hypothesis implies that when a tensor of higher rank should be constructed, it consists out of tensors constructed from lower rank, e. g. at rank

4 from $\delta^i_j\delta^k_l$ and $\delta^i_l\delta^j_l$ as well as of potentially new invariant tensors. E. g. for su(3) there is also a new invariant totally symmetric tensor at rank 4. It is essentially the statement that any representation of a given dimensionality can be decomposed into reducible and irreducible parts, and so can any vector in this dimensionality.

It is worthwhile to have a few examples of this statement. Take two tensors u and v of rank 1 in the 3-representation of su(3), i. e. their indices run from 1 to 3. Then it is possible to rewrite their tensor product $u \otimes v$ as

$$u^i v^j = \frac{1}{2}(u^i v^j + u^j v^i) + \frac{1}{2}\epsilon^{ijk}\epsilon_{klm}u^l v^m$$

This rewriting is at first not obvious. The logic behind is that the left-hand-side is a dimension 9 object. The first term on the right-hand side is a totally symmetric tensor, and has thus 6 different independent elements, and hence being a 6-dimensional representation. The second part is an anti-symmetric tensor, multiplying the proper antisymmetric tensor $\epsilon_{klm}v^l v^m$. It has a single lower component, and thus belongs to a conjugate representation. As there are only three elements, this must be a $\overline{3}$ representation. Thus,

$$(1,0) \otimes (1,0) = 3 \otimes 3 = 6 \oplus \overline{3} = (2,0) \oplus (0,1),$$

and a tensor product of two 3-representation can be decomposed into the two irreducible representations 6 and $\overline{3}$. This decomposition is actually unique. The important step was to note that the only possibilities of decomposing something into other objects was by invoking a decomposition using the invariant tensors. This is not obvious for the first term, but this could be written using the δ-tensor as well. Any attempt to use the symmetric tensor would only result again in the same construction, as this can only create a totally antisymmetric tensor.

Two more examples may be useful,

$$\begin{aligned}(1,0) \otimes (0,1) = 3 \otimes \overline{3} &= 8 \oplus 1 = (1,1) \oplus (0,0) \\ u^i v_j &= \left(u^i v_j - \frac{1}{3}\delta^i_j u^k v_k\right) + \frac{1}{3}\delta^i_j u^k v_k. \end{aligned} \tag{10.8}$$

Here, the first term has 8 independent possibilities, again constructed using two different contractions of δ-tensors. The second term has only a single independent object, $u^k v_k$, as the remainder is just an invariant tensor. Thus, this is a decomposition into the adjoint representation and into the trivial representation.

Finally,

$$\begin{aligned}(1,0) \otimes (1,1) = 3 \otimes 8 &= 15 \oplus \overline{6} \oplus 3 = (2,1) \oplus (0,2) \oplus (1,0) \\ u^i v^j_k &= \frac{1}{2}\left(u^i v^j_k + u^j v^i_k - \frac{1}{4}u^l v^j_l \delta^i - k - \frac{1}{4}\delta^j_k u^l v^i_l\right) \\ &\quad + \frac{1}{4}\epsilon^{ijl}\left(\epsilon_{lmn}u^m v^n_k + \epsilon_{kmn}u^m v^n_l\right) + \frac{1}{8}\left(3\delta^i_k u^l v^j_l - \delta^j_k u^l v^i_l\right)\end{aligned} \tag{10.9}$$

where again the counting of the symmetric and antisymmetric tensors shows the correct dimensionalities.

The language of tensors now provides also a new possibility to generalize the Wigner-Eckart theorem. Consider

$$\langle v|W|u\rangle,$$

where the v and u are now arbitrary tensors. This corresponds to matrix elements

$$\langle v^{ij\dots}_{mn\dots}|W^{kl\dots}_{ab\dots}|u^{rs\dots}_{xy\dots}\rangle = \Gamma^{mn\dots kl\dots rs\dots}_{ij\dots ab\dots xy\dots}.$$

But any such object must be decomposable into invariant tensors. Thus, depending on the number of uncontracted indices, this is a sum

$$\langle v|W|u\rangle = \sum_i \lambda_i t_i, \tag{10.10}$$

where the t_i are the invariant tensors of the corresponding rank. Hence, such matrix elements can be decomposed into a finite sum with a finite number of numbers λ_i which are not determined by the group structure, and invariant tensors of the group. The undetermined numbers λ_i are those which in physics are determined from the dynamics of the system, e. g. the energy levels of hydrogen. This is just a generalization of the Wigner-Eckart theorem, expressed using the primitiveness assumption above. The evaluation of these matrix elements for $W = 1$ is yielding the Clebsch-Gordan coefficients.

10.6 Young tableaux

There exists a convenient graphical approach to constructing tensor products of representations, which also helps in finding the type of tensors in terms of invariant tensor deconstruction. These are called Young tableaux. It is based on the fact that the decomposition in terms of invariant tensors (10.10), e. g. (10.8) and (10.9), always occur in terms of either symmetric or antisymmetric invariant tensors. The dimensionalities then follow necessarily from the number of independent components of such tensors. The tensors themselves are then necessarily related to tensor products of the permutation group. And it is here where Young tableaux come into play.

A Young tableaux is build from (square) boxes. An elementary object is a single box. A tensor product is obtained from building out of the Young tableaux of the individual representations a new tableaux. There are two rules for this. For this, a box can be appended to the right of below an existing box. However, the resulting tableaux must be left-justified and the length of a row cannot exceed the length of a row above. The arrangement now makes statements about symmetrization and antisymmetrization. The indices are symmetrized along rows and antisymmetrized along columns.

Thus, tensorizing vector representations imply staking two single boxes. They can be formed in a row, yielding a symmetric rank two tensor or a column, yielding an antisymmetric rank two tensors. If the vector has three components, these are 9 in total. A symmetric combinations has six components, the antisymmetric one has three components. These representations do not interact, but can still be completely reducible. E. g., starting from for the spin 1 vector representation of so(3), 6 components can be given

in terms of a spin 2 and a spin 0 component, while 3 components form again a spin 1 representation, but now of antisymmetric matrices, rather than a vector.

Adding in a third, it is necessary to decide whether to add to the row or to the column. Adding to the row yields two possibilities, a row of three boxes, and a hook. The row is a symmetric rank 3 tensor, having 10 independent components, which can be written as a completely reducible combination of a spin 7 and a spin 3 representation, the latter needs to be symmetric. The hook is also rank three, but symmetric on two indices and antisymmetric on another pair, with one index shared. This gives eight independent components. There appear two possibilities, namely spin 7 and spin 0 or spin 5 and spin 3. But the former has just been found to be completely symmetric for rank three, but the other two are not yet fixed, so it is spin 5 and spin 3. In the same vain attaching it to the other two possibilites yields a completely antisymmetric rank three tensor, which has only one component, and thus spin zero, and another hook.

With the same logic, it is possible to create further tensor products and tensor decompositions. However, once complex representations are added, it is necessary to keep in mind that the indices in complex representations and complex conjugate representations are not equivalent, and thus need to be also to be treated independently. E. g. for the case of su(3), the row can only be a singlet, if the two boxes are from the different representations, while the column is then taking the other 8 elements, as a ll possibilities are different.

10.7 Bird tracks

It is furthermore possible to represent involved tensor products graphically. While it does not replace the necessity to calculate, e. g., Clebsch-Gordan coefficients explicitly, it is advantegous for keeping track as well as using graph theoretical tools to construct all possible tensor products. This procedure is called bird tracks.

It starts by using for a given tensor of some rank for each index an external line. Usually, it carries a multi-index, based on representation and magnetic quantum number. Open and closed boxes will symmetrized and antisymmetrized lines, which enter from one side, and leave from another side. Triangles (or 3-point vertices) fuse two lines into a single line, by replacing two representations by one of the constructed ones, in the sense of the Clebsch-Gordan coefficients. The possible representations can be constructed using the Young tableaux of section 10.6. Closed lines therefore correspond to traces over indices. A bird track with no external lines is therefore a number. E. g., a circle with a bisecting line will represent a Clebsch-Gordan coefficient. Since the external lines carry a magnetic quantum number, it is only one. Summing over magnetic lines will require an additional summation of diagrams, which is occasionally made implicit.

10.8 Semidirect product

While so far all tensor products have the property that they combine two groups (representations) G and H to a new group $G \otimes H$ such that the new group has elements (g, h) with the composition law $(g_1, h_1) \circ (g_2, h_2) = (g_1 g_2, h_1 h_2)$, this is by far not the only possibility, nor the only one relevant for physics.

An alternative are semidirect products, which rather satisfy

$$(g_1, h_1) \circ (g_2, h_2) = (g_1 \circ g_2, h_1 \circ f(g_1) h_2)$$

where $f(g)h$ is some automorphism of H, i. e. a structure-preserving mapping of H into H, but could be both an inner automorphism or an outer automorphism.

A physically relevant example is again the Galileo group including rotations and translations. It is a semidirect product of rotations and translations, since two rotations are just as usually composed, but the direction of the second translation has to be rotated by the first rotation, $(rr', t + rt')$, and thus rotations play the role of G and translations the role of H. The automorphism is then the rotation of the first translation.

Chapter 11

Continuous groups beyond compact Lie groups

Here, the topic of other groups, especially non-simple Lie groups, non-compact Lie groups as well as graded Lie groups will be discussed. It will also take up again the question how the groups can differ if the algebras coincide.

11.1 Pseudo groups

While in sections 9.2-9.4 the conventional special unitary and special orthogonal groups have been covered, it is useful to also introduce pseudo-versions of them. These groups are denoted by $\mathrm{SU}(p,q)$ and $\mathrm{SO}(p,q)$ and can be characterized to be the matrices M such that

$$\begin{aligned} M^a G M &= G \\ G &= \begin{pmatrix} 1_p & 0 \\ 0 & -1_q \end{pmatrix} \end{aligned}$$

where a is T for the special orthogonal case and $\dagger$ for the special unitary case. This can be generalized to the corresponding non-unimodular versions.

The relevance of these groups for physics is evident by noting that for the (special) orthogonal case and $p = 3$ and $q = 1$ (or $p = 1$ and $q = 3$) this defines the group of (proper) Lorentz transformations, as then G is the Minkowski metric.

In general, for any non-compact Lie group it is possible to choose the Cartan metric

$$g_{ab} = f^d_{ac} f^c_{bd}$$

such that it becomes $g = \mathrm{diag}(1_p, -1_q)$. As a consequence, all generators can be written as the sum of two generators, $T = K + M$ with $K^T = -K$ and $M^T = M$, such that the algebras created by the two sets of generators K and M obey

$$\begin{aligned} [K,K] &\sim K \\ [K,M] &\sim M \\ [M,M] &\sim -M \end{aligned}$$

and thus form two (noninvariant) subalgebras. This can be used to completely classify all non-compact simple Lie algebras, as the replacement $M \to iM$ turns both algebras into the same form as Lie algebras, and making the Cartan metric positive, and thus that of a compact Lie algebra. However, the process is still more complicated than for compact Lie algebras, and is therefore skipped here. However, this will be very useful in characterizing the Poincare algebra.

11.2 The Lorentz group

Taking up the issue of pseudogroups from section 11.1 again, a natural question is to ask what the representations are. Especially, given that the Lorentz group is a pseudo-group[1].

Considering the cases of SO(1,1) and SO(1,3) first[2], finite-dimensional representations are already known. In two dimensions the representation is given by

$$\Lambda = \begin{pmatrix} \cosh\alpha & \sinh\alpha \\ \sinh\alpha & \cosh\alpha \end{pmatrix}$$

while in four dimensions the group elements can be constructed from six basic elements

$$J_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad J_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\beta & 0 & \sin\beta \\ 0 & 0 & 1 & 0 \\ 0 & -\sin\beta & 0 & \cos\beta \end{pmatrix}; \quad J_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \cos\gamma & \sin\gamma \\ 0 & 0 & -\sin\gamma & \cos\gamma \end{pmatrix}$$

$$K_1 = \begin{pmatrix} \cosh\eta & \sinh\eta & 0 & 0 \\ \sinh\eta & \cosh\eta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad K_2 = \begin{pmatrix} \cosh\xi & 0 & \sinh\xi & 0 \\ 0 & 1 & 0 & 0 \\ \sinh\xi & 0 & \cosh\xi & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad K_3 = \begin{pmatrix} \cosh\zeta & 0 & 0 & \sin\zeta \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \cosh\zeta & 0 & 0 & \sinh\zeta \end{pmatrix}.$$

In the first three immediately the group elements of SO(3) are recognized, while the other three constitute the elements of a second SO(3), but with imaginary angels. In this sense, the group splits into the two subgroups as discussed in section 11.1.

However, already the two-dimensional case shows that this representation is not unitary, i. e. $\Lambda^\dagger \neq \Lambda^{-1}$. The reason can be traced back to the fact that the single generator of SO(1,1) is given by

$$\lambda = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$$

which is not Hermitian. Since furthermore $\mathrm{tr}\lambda^2 < 0$ this implies what was already anticipated: The group is not compact. As shown in section 8.3, this implies that it does not have any finite-dimensional unitary representations.

[1]The Poincare group is a tensor product of the Abelian translation group and the Lorentz group, and the latter plays no role in the following.

[2]These are the proper Lorentz group, while O(1,1) and O(1,3) would also include parity and time reversal.

However, there exists infinite-dimensional unitary representations. In this case, a function space is required. To introduce it, it is useful to have a different view on the Lorentz group.

The Lorentz group consists out of rotations J and boosts K. In general, commutators of J and K do not vanish. However, defining a skew version of these operators

$$\begin{aligned} A &= \frac{1}{2}(J + iK) \\ B &= \frac{1}{2}(J - iK) \end{aligned}$$

this is the case. The Lorentz algebra becomes then a direct product of two SU(2) algebras

$$\begin{aligned} [A_i, A_j] &= \epsilon_{ijk} A_k \\ [B_i, B_j] &= \epsilon_{ijk} B_k \\ [A_i, B_j] &= 0. \end{aligned} \tag{11.1}$$

Hence, any representation of the Lorentz group can be assigned two independent quantum numbers, which are either integer or half-integer. These are connected with physical objects, depending on how the transform under each of the two SU(2) subalgebras. E. g. scalars are then just twice the trivial case, $(0,0)$. Left-handed fermions and right-handed fermions, however, belong to the $(1/2,0)$ and $(0,1/2)$ representations, vectors like the momentum belong to the $(1/2,1/2)$ representation, and antisymmetric tensors like the generators of angular momentum to the $(1,0)+(0,1)$ representation.

Each of the subgroups are themselves compact. But their tensor product giving the Lorentz group is not, as their product is only a semi-direct product. The direct product delivers the conventional $SO(4) \sim SU(2) \otimes SU(2)$, the Euclidean rotation group.

Now, finally the unitary representation is given by a functional space, i. e. a space of functions $\phi^i_{\dot{j}m}(p)$ such that $p^2 = m^2$ and p is an n-dimensional vector with $p^2 = p_0^2 - \vec{p}^2$ living in the n-dimensional non-unitary representation of the Lorentz group. The indices i and $\dot{j}$ are multiindices, which describe the representation the functions form of the two SU(2) subgroups. The action of the Lorentz group is then

$$\Lambda \phi^i_{\dot{j}m}(p) = \lambda^i_k \lambda^{\dot{l}}_{\dot{j}} \phi^k_{\dot{l}m}(\lambda p), \tag{11.2}$$

where the λ are the corresponding non-unitary representations. In this case, $\Lambda^\dagger = \Lambda^{-1}$.

The parameter m classifies the little group of the orbits, where $m > 0$, $m = 0$, and $m < 0$ are the strata for the little groups SO(3), SO(2), and SO(1,2), where only the former two appear in physics. Correspondingly, m is the eigenvalue of the (lowest-order) Casimir of this representation with respect to the orbits, while the spin is the Casimir for the non-unitary finite-dimensional representations.

It is possible to upgrade the Poincare group further to the so-called conformal group. This is done by adding two more space-time transformations, corresponding to 5 more generators in four dimensions, to the group,

$$\begin{aligned} x^\mu &\to \lambda x^\mu \\ x^\mu &\to \frac{x^\mu + a^\mu x^2}{1 + 2x^\nu a_\nu + a^2 x^2}, \end{aligned}$$

with arbitrary vector a_μ. The first is a scale transformation, which is also called dilatation. The second is the special conformal transformation. The scale symmetry is in so far remarkable in physics, as only scaleless theories can be conformal. Hence, a conformal theory never has any kind of intrinsic mass scale. As a consequence, all infinite-dimensional, unitary representations of the conformal group, which still need to satisfy all the requirements of the representations of the Poincare group as well, can only be those states in (11.2) which satisfy $m^2 = 0$. Note that the finite-dimensional spin representations of the rotation group are not affected.

11.3 Covering groups

The probably most important consequence of considering topological groups is the insight that the seemingly surprising identification of different Lie algebras in chapter 8 is due to the fact that the corresponding Lie groups are identical in the patch containing the unit element, and are thus locally isomorphic in this patch. However, the difference is that not all of the respective groups are simply connected. But it can be shown that for every set of locally isomorphic Lie groups there is one unique Lie group which is simply connected. This group is called the covering group.

E. g. for SO(3) and SU(2), the covering group is SU(2), since SO(3) is not simply connected. The explicit mapping for group elements is given by

$$r_{ij} = \mathrm{tr} u^\dagger \sigma_i u \sigma_j \tag{11.3}$$

where u is an arbitrary element of SU(2) and r the corresponding element of SO(3), and the σ_i are the corresponding generators of SU(2). This also shows how the simple connectedness of SU(2) is lost, as the double appearance of the u elements eliminates signs. Especially, elements proportional to the negative unit element of SU(2) are mapped into an element proportional to the positive unit of SO(3). This is generic. Since for any covering the number of continuous parameters needs to be the same, as it would otherwise not be a homomorphism, the different patches can only differ by a discrete group. Furthermore, the difference is exactly the center of the group SU(2), which is the discrete group Z_2, and thus in a sense SO(3)$\sim$SU(2)/Z_2. If the difference would not be the center of the covering group, then the different elements would not commute in the vicinity of the identity, and therefore would correspond to some different group element expanded close to the identity, but this would be continuous in contradiction to the discreteness. Thus, in general for any group H, it is homomorphic to its covering group G, up to its center Z, and thus $H \sim G/Z$. In reverse, $H/G \sim Z$.

However, the covering group does not need to be a simple Lie group. E. g. for SO(4) the covering group is SU(2)$\times$SU(2), nor does it need to be a simple product group.

Studying (11.3), it is seen that the left-hand-side is a three-dimensional representation, and thus the adjoint representation. This immediately shows that all center elements of a group are mapped to the identity in the adjoint representation, which is therefore not a faithful representation. This is generally true. The reason is that the adjoint representation forms automatically a basis in its representation space, and then the only

element commuting with every other element can be the unit matrix, and therefore the whole center is mapped to the unit matrix.

Hence, coinciding Lie algebras can yield different Lie groups, which then differ by connectedness and the center. Conversely, the adjoint representation of the algebra is faithful, if the center is trivial.

It is for this reason interesting to enumerate for the simple Lie groups centers and covering groups. In fact, only the SO(n) groups are not simply connected, but have a covering group which is different only by a Z_2 or a Z_4 or a $Z_2 \times Z_2$ factor and called the spin group Spin(n). Necessarily the spin groups have the same Lie algebra as the SO(n) group, and thus fall otherwise into the same classification. Furthermore, the centers of SU(n) are Z_n, of E_7, SO($2n$), and Sp($2n$) Z_2, of E_6 Z_3, and in all other cases trivial.

The concept of covering groups now also gives insight into the somewhat mysterious spinor representations of the SO(n) groups in section 9.4.3. Since the SO(n) are not simply connected, representations exist, which are not faithful, but double-valued, i. e. they map one group element into two different representation elements. The SO(n) are the only such groups, and the result are the double-valued spinor representation. The corresponding representation of the covering group is, however, unique. This also explains the example of so(3) in section 9.4.3: In su(2), the elements multiplied by -1 are ordinary group elements. These are factored out when going to the non-connected so(3). Since the spinor representation is equivalent to su(2), these elements nonetheless exist, but as a double-valued 'ghost' image of the su(2) elements. Similar considerations apply to SO(5), where the covering group is Sp(4), as in the rank 2 case the lacedness does not play a role.

As can be deduced from this discussion, there are groups for which there exists continuous, unitary irreducible representations, which are either not faithful (adjoint representations for groups with a non-trivial center) or double-valued (spinor representations of the disconnected groups SO(n)). However, these groups all have the same algebra, i. e. they are isomorphic at the algebra level. As a consequence, the group which for a given representation is both faithful and single-valued is called the corresponding true group. E. g. for the isomorphic algebras so(3)$\sim$su(2), for the adjoint representation the group SO(3) is the true group, but for the spinor representation it is the SU(2) group.

11.4 True groups

11.4.1 True groups and the center

Th full classification of representations according to the true groups is rather complicated, especially in the non-simple case. Therefore, only the cases will be considered which are of direct relevance in physics. Here, it becomes especially important that not actually simple groups are the most common case in (particle) physics, but rather just compact Lie groups, with possible Abelian factor groups. The best known example is the standard model of particle physics.

The actual calculation is somewhat involved, and it is best to proceed in several steps.

Note that due to the discussion in chapter 10, it is possible to construct arbitrary

representations by tensor products of the fundamental representations.

Start out with the groups without center, i. e. $SO(2n+1)$, G_2, F_4, and E_8. It is here important that the groups are considered, and not the algebra. Since these groups do not have a center, their adjoint representation is faithful. Furthermore, all tensor representations of them are single-valued.

For the groups with center Z_2, i. e. $SU(2)$, $SO(2n)$, $Sp(2n)$, and E_7, there are two possibilities for their tensor representations. Either they are build from an even or odd number (rank) of fundamental representations. For odd tensors, permutations of the indices permits to explicitly include the center, while this is not possible for an even rank. Thus, in the latter cases, the true group is G/Z_2 rather than the groups themselves.

The same is also true for the spinor representations of $SO(2n+1)$: If they are build from an even number of fundamental spinor representations, the true group is $SO(2n+1)\times Z_2$, but otherwise the original group.

This shows already how the structure emerges: Since center elements are roots of unity, it is necessary to be able to associate to every element such a factor under permutations of the fundamental representations building a given representation. E. g. for E_6 with center Z_3, the true group is again E_6 if the rank modulo 3 is non-zero, but E_6/Z_3 otherwise.

This effect can be seen as follows: A group element acting on a tensor will act on every fundamental representation simultaneously. In case of Z_2, there are two elements, and thus the values can be at most 1 or -1, and thus

$$\begin{aligned} t_1 &\rightarrow (-1)t_1 \\ t_1\otimes t_2 &\rightarrow (-1)(-1)t_1\otimes t_2 = t_1\otimes t_2 \\ t_1\otimes t_2\otimes t_3 &\rightarrow (-1)(-1)(-1)t_1\otimes t_2\otimes t_3 = (-1)t_1\otimes t_2\otimes t_3 \end{aligned}$$

and therefore a non-trivial action is only possible for an odd number of representations.

If the center is Z_3, the effect is, e. g. for the element $\exp(i2\pi/3)$

$$\begin{aligned} t_1 &\rightarrow (\exp(i2\pi/3))t_1 \\ t_1\otimes t_2 &\rightarrow (\exp(i2\pi/3))(\exp(i2\pi/3))t_1\otimes t_2 = (\exp(i4\pi/3))t_1\otimes t_2 \\ t_1\otimes t_2\otimes t_3 &\rightarrow (\exp(i2\pi/3))(\exp(i2\pi/3))(\exp(i2\pi/3))t_1\otimes t_2\otimes t_3 = (\exp(i2\pi))t_1\otimes t_2\otimes t_3, \end{aligned}$$

and likewise for the other element. Therefore the effect is trivial for the case of modulo 3 representations.

This still leaves the interesting cases of $SO(2n)$ with center Z_2 and $SU(n)$ with center Z_n.

To discuss the $SU(n)$ case, introduce the rank index of a representation

$$t = \sum^{l} s\mu_s \mod n+1$$

where μ_s are the corresponding Dynkin weights of the representation, and the sum is over the involved representations. The weight μ_s counts the number a given representation appears, and s is the number the fundamental representation appears. E. g., for $SU(2)$

$$t = 2j \mod 2+1,$$

where j is the spin of the representation and for SU(3) it is the triality of the representation: The rank index characterizes thus the rank of the representation modulo the size of the center.

This result implies that in a representation identified by rank index t for a group with center Z_n, the application of a center element ω (with $\omega^n = 1$) will yield the total effect ω^t. The complexity arises from the fact that there may still be a subgroup of the center, which is non-trivially realized on this representation. To proceed, note that there is some number f such that

$$\begin{aligned} n &= f n_0 \\ t &= f t_0 \end{aligned}$$

such that n_0 and t_0 are relatively prime. It follows

$$(\omega^t)^{n_0} = \omega^{f t_0 n_0} = (\omega^n)^t = 1 = (\omega^n)^{t_0},$$

and n_0 is the smallest integer for which this is true. Thus, the representation with rank index t carries a representation of the center group $\mathrm{Z}_{n_0} = \mathrm{Z}_n/\mathrm{Z}_f$. The true group is therefore G/Z_f.

Take as an example SU(3). The possible rank indices are 1, 2, and 3. The possible values for f are therefore 1, 1, and 3. Thus, the true group is SU(3) for $t = 1$ and $t = 2$, but $\mathrm{SU}(3)/\mathrm{Z}_3$ for $t = 3$. The latter includes the adjoint representation, which is a totally antisymmetric tensor product of three fundamental representations. The fundamental representations and the 6 representation, being a symmetric tensor product of two fundamental representations, are of the former type. Note that the other in physics particular important representations of SU(3), 10, $\overline{10}$, and 27 also all have rank index 3.

This sequence can be more involved. Consider SU(4), with rank indices 1-4. The values of f are 1, 2, 1, and 4. Thus, the true group can be either SU(4), $\mathrm{SU}(4)/\mathrm{Z}_2$, or $\mathrm{SU}(4)/\mathrm{Z}_4$. For SU(6), the sequence of f is 1, 2, 3, 2, 1 and 6. In general, if n is prime, then f can be either only 1 or n. For the adjoint representation f is always n, and thus the true group for the adjoint representation is always $\mathrm{SU}(n)/\mathrm{Z}_n$.

For the SO($2n$) groups, the situation is a little bit more subtle, but it is along similar lines possible to define a rank index, and then determine the corresponding true groups. It will not be detailed here, but yields that the true groups for the spinor representations are always the covering groups, while in all other cases it is the original group, but sometimes with part, but not all, of the center divided out. Thus, all tensor representations of SO($2n$) contain at least the center Z_2.

11.4.2 True groups and Abelian subgroups

While so far only the true groups of simple groups have been considered, physics often requires to consider the situation with non-simple groups. Generically, this becomes quite involved. Therefore, the general strategy will be discussed here for the case most relevant to particle physics of groups $\mathrm{SU}(p)\times\mathrm{U}(1)$ and $\mathrm{SU}(p)\times\mathrm{SU}(q)\times\mathrm{U}(1)$, which cover e. g. the gauge group of the standard model. In the context of physics the quest for the true

groups becomes very important, as quantum mechanics requires wave functions to be single-valued. Thus, a meaningful quantum theory can only be built with objects in single-valued representations, and therefore in the true groups.

To start, it is helpful to recall that the continuous unitary irreducible representations of U(1) are $\exp(im\phi)$, where the integer m is both the rank and the Dynkin index of the representation. Note, however, that only $m = 1$ is faithful, as in all other cases multiple group elements are mapped to the same value.

The simplest case is SU(p)×U(1) with p prime. Then there exist two possible groups, SU(p)×U(1) and SU(p)×U(1)/Z_p=U(p). The last equality is obtained by identifying the elements of Z_p in SU(p) with the corresponding ones in U(1). E. g., for SU(2)×U(1) this is achieved by identifying the element diag$(-1,-1)$ from SU(2) with the element (-1) from U(1). This is generalized by identifying the diagonal matrix element with entries ω (with $\omega^p = 1$) with the number ω'.

Another view is illustrated by the fact that the p-dimensional representation of U(p) is the direct product of SU(p) and U(1), and then the identification is manifest: This is just the set of p-dimensional special unitary matrices multiplied by an arbitrary phase, and then there is no distinction between the factor ω from either parts, as both yield a unit matrix with the ω on the diagonal. However, SU(p)×U(1) yields more directly a direct sum of dimension $p+1$, where then the difference of both center elements is manifest: The one is a p-dimensional unit matrix times ω appended with a one-dimensional unit matrix, while the other is a positive p-dimensional unit matrix appended with a one-dimensional diagonal matrix times ω.

Such an identification of elements is only possible if ω^t and ω'^m become one at the same time, and thus $t = m \mod p$. If this is the case, the two representations of the center can be identified, and thus the true group is U(p), and otherwise SU(p)×1. The p and $p+1$ dimensional representations are particular examples of this with $m = t = 1$, and representations $(1,1)$ and $(1,0)+(0,1)$ for (t,m).

In the next case is SU(p)×SU(q)×U(1) with $p \neq q$ and both relative prime. The center elements satisfy $\omega^p = 1$ and $\sigma^q = 1$. There are, of course, again elements ω' and σ' in the U(1) which also exponentiate in the same way to one. With the same argumentation, the possible true groups are then

- SU(p)×SU(q)×U(1)
- U(p)×SU(q)=SU(p)×SU(q)×U(1)/Z_p, identifying $\omega = \omega'$
- SU(p)×U(q)=SU(p)×SU(q)×U(1)/Z_q, identifying $\sigma = \sigma'$
- S(U(p)×U(q))=SU(p)×SU(q)×U(1)/Z_{p+q}, identifying both $\omega = \omega'$ and $\sigma = \sigma'$

The last case is actually the one relevant for the standard model of particle physics, as all particles appear to be in the fundamental representation of the full standard model gauge group.

11.4.3 True groups and little groups

It is possible to show that for any continuous unitary and irreducible representation of a group there are at most a finite number of little groups. Furthermore, there is a smallest, non-trivial little group, in the sense that this little group is a (conjugated) subgroup of all little groups. However, there may be more than one maximal little group, i. e. little groups which are not related as (conjugated) subgroups to other little groups, but still are little groups of the representation in question.

A number of general statements can be made about the strata of these little groups. The first is that the stratum of the minimal little group is open and dense in the representation. It is thus, in a sense, maximal, and therefore also called the generic stratum. In the context of Morse theory, only the second-order invariant has a extremum in the generic stratum, while all other invariants have extrema in other strata, which is stated without proof.

The opposite are, modulo a norm, discrete strata, i. e. strata with a finite number of elements. The orbits in such strata are called critical strata. These strata are often, but not always, associated with maximal subgroups. In fact, it can be shown that the little group is maximal if there is only a single, modulo norm, orbit in its stratum, which is therefore a singlet under the action of this little group. An example were the fixed vectors for SO(n). It is important that, for a fixed representation, not necessarily a subgroup of the group is also a little group. In fact, it is often not. However, it is usually possible to construct some representation of which a given subgroup of a group is a little group, but this may not necessarily be an irreducible representation, nor easy. In the context of Morse theory, it is possible to show that all invariants, except for the second-order one, have extrema on critical orbits, which is thus an alternative characterization.

Another feature is that the little group may have the same or lesser rank than the original group, or, if it is discrete, even have rank zero. Since the rank is associated with the number of charges, this plays an important role in physics. In this context it is sometimes useful to define the little space as the space which is left invariant by a little group. E. g., for SO(3) this is the two-dimensional subspace left invariant by the SO(2) little group in the vector representation of the group.

11.5 Clifford algebra

A generalization of the Clifford algebra in section 9.4.3 is one where the right-hand side is a more general metric tensor $g_{\mu\nu}$ instead of the Euclidean $\delta_{\mu\nu}$

$$\{\gamma_\mu, \gamma_\nu\} = \gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2g_{\mu\nu}1,$$

The metric $g_{\mu\nu}$ needs to be a non-degenerate, symmetric matrix. The (identical) index range N of μ and ν defines the algebra, and therefore this is actually a family of algebras.

The probably best known case is the one where there are four-different values for μ and ν and $g_{\mu\nu}$ is the Minkowski metric. Then, the lowest-dimensional representation is four-dimensional, and given by the so-called Dirac matrices. If the range is three, the

lowest-dimensional representation is two, and the γ_μ coincide with the Pauli matrices. Interestingly, there is no three (and actually no odd) dimensional representation of the Clifford algebra.

This plays the same role for the pseudo group SO(1,$N-1$) as the Euclidean Clifford algebra did for the compact Lie group SO(N) is section 9.4.3, and is closely related to fermions in physics. Again, also the Clifford representations are homomorphic to the representations of SO(1,$N-1$). This becomes immediately obvious by decomposing the representations into their SU(2) content using the Pauli matrices. In fact, the Clifford group generated from the Clifford algebra is a covering group of these groups, and identical to the aforementioned spin groups Spin(n).

Note that the concept of Clifford algebras can be further generalized, but this leads substantially beyond the scope of this lecture, and it is actually rarely encountered in physics.

11.6 Grassmann algebra

A, at first sight, deceptively similar algebra to the Clifford algebra is the Grassmann algebra

$$\{\xi_a, \xi_b\} = 0.$$

However, its properties are very different. The most remarkable fact is that $\xi_a \xi_a = 0$, i. e., the elements are nilpotent. This implies that there is no one-dimensional representation. Higher-dimensional representations are given by nilpotent matrices. However, since nilpotent matrices cannot be inverted, this algebra cannot generate a group under matrix multiplication, in contrast to the Clifford algebra, or any Lie algebra. However, it still forms a semigroup, as all other properties of a group, except the existence of an inverse, and thus neutral element, are satisfied.

However, it is possible to form a group under addition. This can be extended by including also ordinary numbers to form a vector space. This vector space will be useful to introduce graded algebras in section 11.7.

Due to the nilpotency, the set $\mathcal{S}$ of independent Grassmann numbers with $a = 1, ..., N$ base numbers are

$$\mathcal{S} = \{1, \alpha^a, \alpha^{a_1}\alpha^{a_2}, ..., \alpha^{a_1} \times ... \times \alpha^{a_N}\},$$

where all a_i are different. This set contains therefore only 2^N elements. There are no more elements, as the square of every Grassmann number vanishes, and by anti-commuting thus any product containing twice the same Grassmann number vanishes. Of course, each element of $\mathcal{S}$ can be multiplied by ordinary complex numbers c, and can be added. This is very much like the case of ordinary complex numbers. Such combinations z are called supernumbers, and take the general form

$$z = c_0 + c_a\alpha^a + \frac{1}{2!}c_{ab}\alpha^a\alpha^b + ... + \frac{1}{N!}c_{a_1...a_N}\alpha^{a_1} \times ... \times \alpha^{a_N} = c_0 + c_S. \tag{11.4}$$

Here, the factorials have been included for later simplicity, and the coefficient matrices can be taken to be antisymmetric in all indices, as the product of α^as are antisymmetric.

For $N = 2$ the most general super-number is therefore

$$z = c_0 + c_1\alpha^1 + c_2\alpha^2 + c_{12}\alpha^1\alpha^2,$$

where the antisymmetry has already been used. Sometimes the term c_0 is also called body and the remaining part soul. It is also common to split the super-number in its odd and even (fermionic and bosonic) part. Since any product of an even number of Grassmann numbers commutes with other Grassmann numbers, this association is adequate. For $N = 2$, e. g., the odd or fermionic contribution is

$$c_1\alpha^1 + c_2\alpha^2,$$

while the even or bosonic contribution is

$$c_0 + c_{12}\alpha^1\alpha^2.$$

Since the prefactors can be complex, it is possible to complex conjugate a supernumber. The conjugate of a product of Grassmann-numbers is defined as

$$(\alpha^a ... \alpha^b)^* = \alpha^b ... \alpha^a \tag{11.5}$$

Note that this implies that a product of an even number of Grassmann numbers is imaginary while an odd number is real,

$$\begin{aligned} \alpha^* &= \alpha \\ (\alpha\beta)^* &= \beta\alpha = -\alpha\beta, \end{aligned}$$

due to the anti-commutation when bringing the product back to its original order.

An important property of a super number z is its Grassmann parity $\pi(z)$. It differentiates between numbers which commute or anti-commute, and thus takes the values 0 or 1. Hence, for two super numbers

$$z_1 z_2 = (-1)^{\pi(z_1)\pi(z_2)} z_2 z_1,$$

the Grassmann parity can be used to determine the sign of permutations. Note that only supernumbers with only even or odd numbers of Grassmann numbers have a definite Grassmann parity. Hence, super numbers with definite Grassmann parity 0 or 1 are therefore called even or odd.

Finally, a norm can be defined as

$$|z|^2 = |c_0|^2 + \sum_{k=1}^{\infty} \sum_{\text{Permutations}} \frac{1}{k!} |c_{a_1 ... a_k}|^2,$$

such that it is possible to give meaning to the statement that a super number is small.

To construct groups or representations of Grassmann algebras it is often useful to refer to Grassmannian, in the physics context often also called fermionic, dimensions. Ordinary real coordinates are then referred to as bosonic. This is also called a superspace formalism, especially in the context of supersymmetry.

11.7 Graded algebras

A graded algebra is an algebra which satisfies

$$\left[t^a, t^b\right\} = t^a t^b - (-1)^{\eta_a \eta_b} t^a t^b = i f^{abc} t_c,$$

where the η_i are known as gradings of the elements t_a, and are 0 for bosonic and 1 for fermionic generators. Because of the symmetry properties of the left-hand-side, this implies that f^{abc} is zero except when $\eta_c = \eta_a + \eta_b$. Still, if the t_a are Hermitian it follows that $f^*_{abc} = -f_{bac}$. Also a super-Jacobi identity follows

$$(-1)^{\eta_c \eta_a} \left[\left[t_a, t_b\right\}, t_c\right\} + (-1)^{\eta_a \eta_b} \left[\left[t_b, t_c\right\}, t_a\right\} + (-1)^{\eta_b \eta_c} \left[\left[t_c, t_a\right\}, t_b\right\} = 0.$$

It has the usual, but graded, implication for the relation of the structure constants. Note that the grading of a composite operator is in general given by $(\sum \eta_i) \mod 2$ with the gradings of the constituent operators η_i. Furthermore, any transformation based on this algebra involves necessarily a mixture of ordinary complex numbers and Grassmann numbers, and therefore the parameters receive also a grading.

Thus, a graded algebra is just a mixture of conventional algebras and Grassmann algebras. Thus, using superspaces, it is possible to construct a group, and then representations in the usual way.

While this form is the most common in physics, the concept can be generalized. Generators with different gradings belong to two different vector spaces. The gradings form a representation of Z_2, and the grading is therefore called a Z_2 grading. A Z_2-graded algebra is also known as a superalgebra. Especially, generators v_i in the first vector space V and generators w_i from the second vector space W can be chained $\circ$, and obey then

$$\begin{aligned} v_1 \circ v_2 &\in V \\ v_1 \circ w_1 &\in W \\ w_1 \circ v_1 &\in W \\ w_1 \circ w_2 &\in V. \end{aligned}$$

It is likewise possible to generalize this to a Z_N grading. In this case, the generators v_i^j belong to N different vector spaces V_j, and obey

$$v_i^j \circ v_k^l \in V_{j+k \mod N}.$$

Correspondingly a Z_N-graded algebra involves a $\mod N$-type change to the gradings, and so does the Jacobi identity.