4
Some topology

In the previous two chapters I argued for the existence of solitons by starting with a configuration in which ϕ approached different vacuum values at different parts of spatial infinity, with these values being chosen in such a way that the configuration could not be continuously deformed into a uniform vacuum solution over all of space. For both the kink and the U(1) vortex it was fairly easy to visualize the situation and convince oneself of the topological stability of the soliton. However, we want to be able to consider theories in which the space of vacuum solutions is more complicated, as well as theories in three (or more) spatial dimensions. In such cases the topological constraints can be harder to visualize. It is therefore useful to formulate matters more precisely. As I will explain in this chapter, the mathematical language of homotopy provides an ideal framework for this.

4.1 Vacuum manifolds

The general situation that we want to consider is a field theory with n scalar fields that can be assembled into an n-component column vector ϕ and a scalar field potential $V(\phi)$ that has a family of degenerate minima that form a manifold $\mathcal{M}$. Let us assume that this degeneracy is a consequence of a symmetry group G that is spontaneously broken to a subgroup H by the vacuum expectation value of ϕ. (One can certainly construct potentials with degenerate minima that are not related by a symmetry, but in such cases the degeneracy is usually broken by higher-order quantum corrections.) Given a value of ϕ, the action of an element g of G transforms ϕ to $D(g)\phi$, where $D(g)$ is the appropriate n-dimensional representation of G. In particular, if ϕ_0 minimizes V, then so does $D(g)\phi_0$ for any choice of g. Furthermore, our assumption that the degeneracy is entirely due to the symmetry implies that all minima of V are of this form.

One might then guess that there is a one-to-one correspondence between elements of G and minima of V. This is only true if G is completely broken. If G

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

is only partially broken, there is an unbroken subgroup H that can be defined by the requirement that it leaves ϕ_0 invariant; i.e., that $D(h)\phi_0 = \phi_0$ for any element h of H. Consequently, for any given $g \in G$ and $h \in H$, gh and g have the same effect; i.e., $D(gh)\phi_0 = D(g)\phi_0$. We can therefore define equivalence classes of elements of G by defining two elements g_1 and g_2 to be equivalent if $g_2 = g_1 h$ for some $h \in H$. The set of such equivalence classes is the coset space G/H. There is a one-to-one correspondence between these equivalence classes and the minima of V, so

$$\mathcal{M} = G/H. \tag{4.1}$$

A word of caution is appropriate here. In perturbative treatments of field theories the distinction between different Lie groups that share the same Lie algebra [e.g., SU(2) and SO(3)] is often ignored. It will be essential in our topological considerations to keep track of these distinctions and to be clear as to precisely which groups G and H are. This will be discussed in more detail in Sec. 4.3. However, it will helpful to first introduce the concepts of homotopy and the fundamental group.

4.2 Homotopy and the fundamental group $\pi_1(\mathcal{M})$

Let us start by considering closed paths, or loops, on a manifold $\mathcal{M}$. In particular, let us pick a point x_0 on $\mathcal{M}$ and restrict our attention to paths that begin and end at x_0. Any such path can be specified by a continuous function $f(t)$ taking values in $\mathcal{M}$, with $0 \le t \le 1$ and $f(0) = f(1) = x_0$. Figure 4.1 shows three such paths. In this figure, $\mathcal{M}$ is the two-dimensional Euclidean plane, but with the shaded region deleted. It is evident that path b can be smoothly deformed so that it coincides with path a. However, path c cannot be smoothly deformed into a, since that would require passing part of the path through the forbidden shaded region, which is not part of $\mathcal{M}$. The concept of "smoothly deformable" can be made more precise as follows. Let $f(t)$ and $g(t)$ be continuous paths beginning and ending at x_0. They can be smoothly deformed into one

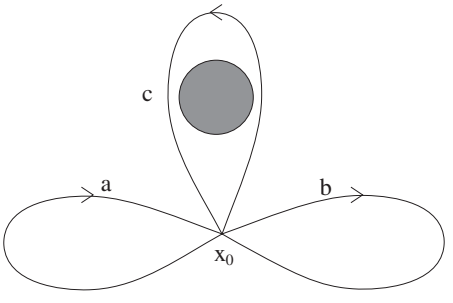


Fig. 4.1. Loops a and b are homotopic to each other, but not to loop c.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

another if and only if there is a continuous function $k(s,t)$ with $0 \leq s,t \leq 1$ such that

$$\begin{aligned} k(0,t) &= f(t), \\ k(1,t) &= g(t), \\ k(s,0) &= k(s,1) = x_0. \end{aligned} \tag{4.2}$$

Thus, $k(s,t)$ can viewed as a sequence of loops, labeled by s, that begin and end at x_0, with f being the first in the sequence and g the last. Paths f and g are said to be *homotopic at* x_0, and the family of paths that define the function k is a *homotopy*.

One can define a product on the space of paths. Given paths f and g, their product is defined as

$$(f \circ g)(t) = \begin{cases} f(2t), & 0 \leq t \leq 1/2, \\ \\ g(2t-1), & 1/2 \leq t \leq 1. \end{cases} \tag{4.3}$$

In other words, $f \circ g$ is the path obtained by going around f and then going around g. An inverse path f^{-1} can be defined as going around f in the reverse direction; i.e., $f^{-1}(t) = f(1-t)$.

The next step is to divide the paths on $\mathcal{M}$ into *homotopy classes*, with the homotopy class $[f]$ denoting the set of paths that are homotopic to f. The definition of the product of paths can be carried over to the product of homotopy classes. Because the product of any path homotopic to f with any path homotopic to g is homotopic to $f \circ g$, we have

$$[f] \circ [g] = [f \circ g]. \tag{4.4}$$

Note that not every path in $[f \circ g]$ is a product path. Equation (4.3) shows that the product of two paths must pass through x_0 three times (at $x = 0, 1/2$, and 1). However, there are certainly paths homotopic to this product that only go through x_0 at their beginning and end, and not at any intermediate points.

There is a trivial "path", $f(t) = x_0$, that just stays at x_0 for its entire length; let us denote the homotopy class of this path by $[I]$. It is easy to see that $[g] \circ [I] = [I] \circ [g] = [g]$ and $[g] \circ [g^{-1}] = [I]$ for any $[g]$. It is also clear that the multiplication of homotopy classes is associative.

Thus, multiplication of homotopy classes has a group structure, with $[I]$ being the identity. This group is denoted $\pi_1(\mathcal{M}, x_0)$, and called the *fundamental group of* $\mathcal{M}$ *at* x_0. Because there are generalizations $\pi_n(\mathcal{M}, x_0)$, which we will encounter later on, the fundamental group is also known as the *first homotopy group*.

If $\mathcal{M}$ is a connected manifold, the fundamental group does not depend on the choice of the base point x_0. To see this, note that any loop beginning and ending at x_0 can be mapped onto a loop beginning and ending at y_0 by adding a path from y_0 to x_0 at the beginning of the loop and then traversing this

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

path backward from x_0 to y_0 at the end of the loop. This mapping preserves all of the homotopy relations between loops, and thus defines an isomorphism between the fundamental groups at x_0 and y_0. If these groups are Abelian, this isomorphism does not depend on which path from y_0 to x_0 is chosen, but the detailed correspondence between homotopy classes of the two groups can be path-dependent if the groups are non-Abelian.

Because the fundamental group does not depend on the choice of the base point, one usual writes simply $\pi_1(\mathcal{M})$, and refers to the fundamental group of $\mathcal{M}$. If all loops on a manifold can be deformed to the trivial loop, then there is only a single homotopy class. We denote this by writing $\pi_1(\mathcal{M}) = 0$, and the manifold is said to be simply connected.

Let us consider some examples:

(i) Any loop on the Euclidean plane can be continuously shrunk to a point, so R^2 is simply connected. The same is clearly true for all higher dimensional Euclidean spaces and, although a loop on a line may be a bit harder to visualize, for the real line R^1. Thus,

$$\pi_1(R^n) = 0. \tag{4.5}$$

(ii) Now consider the space shown in Fig. 4.1, the Euclidean plane with a disk removed. Loops can be characterized by the number of times they wind around the hole left by the disk, with counterclockwise (clockwise) windings counted positively (negatively). Thus, in the figure loops a and b have winding number 0, while loop c has winding number 1. Two loops with the same number of windings are homotopic, and under multiplication of loops the winding numbers add. Thus, the fundamental group of this manifold is just Z, the additive group of the integers. The same would be true if we put an outer boundary on the manifold and reduced it to a ring $r_1 < r < r_2$ enclosing the hole. Indeed, we could just shrink the ring to a circle, S^1, without changing the homotopy group. Thus,

$$\pi_1(S^1) = Z. \tag{4.6}$$

(iii) Consider next a two-sphere,[1] S^2. It may be obvious that any loop on this sphere can be shrunk to a point. If it isn't, imagine deleting from the sphere some point through which the loop does not pass. The sphere with a point deleted is topologically equivalent to the plane, so the result follows from Example (i). Similar arguments applied in higher dimensions show that

$$\pi_1(S^n) = 0, \qquad n \geq 2. \tag{4.7}$$

(iv) Figure 4.2 shows the "figure-eight space", a plane with two disks removed. Two loops, a encircling the left hole, and b encircling the right one, are shown.

[1] Recall that an n-sphere, S^n, is an n-dimensional manifold that can be viewed as a spherical hypersurface in $n+1$ Euclidean dimensions. In particular, a one-sphere is a circle and a two-sphere is the surface of a solid ball in ordinary three-dimensional space.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

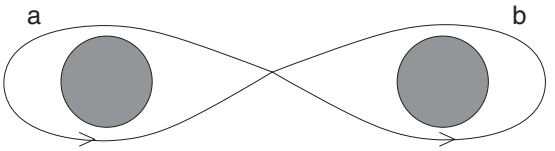


Fig. 4.2. The "figure-eight space". Loops a and b do not commute, and the fundamental group is non-Abelian.

Now consider the product loop $a \circ b \circ a^{-1} \circ b^{-1}$ that goes once clockwise and once counterclockwise around each hole. After some attempts at deforming the loops, you should be able to convince yourself that this product loop cannot be deformed to a point. On the other hand, the loop $a \circ a^{-1} \circ b \circ b^{-1}$ is clearly homotopic to the trivial loop. Thus, the fundamental group of this manifold contains elements $[a]$ and $[b]$ that do not commute, and so is non-Abelian.

4.3 Fundamental groups of Lie groups

Lie groups can be viewed as manifolds, and their fundamental groups are of particular interest. In examining these, the distinction between groups that share the same Lie algebra is crucial.[2] Let us start with the most familiar example of two groups with the same Lie algebra, SU(2) and SO(3).

SU(2) is the group of 2×2 unitary matrices with unit determinant. Any such matrix can be written in the form

$$U = b_0 + i\mathbf{b} \cdot \boldsymbol{\sigma}, \tag{4.8}$$

where the σ_j $(j = 1, 2, 3)$ are the Pauli matrices and

$$b_0^2 + b_1^2 + b_2^2 + b_3^2 = 1. \tag{4.9}$$

This last equation is just that for the unit three-sphere, so we see that as a manifold $\mathrm{SU}(2) = S^3$.

SO(3) is the group of rotations in three dimensions. Any element of the group can be identified by giving a unit vector $\hat{\mathbf{n}}$ that specifies the rotation axis, and a (counterclockwise) rotation angle about that axis that lies in the range $0 \leq \psi \leq \pi$. Note that rotations by π about $\hat{\mathbf{n}}$ and $-\hat{\mathbf{n}}$ have the same effect, and correspond to the same group element. As a manifold, SO(3) can be mapped onto a three-dimensional ball of radius π. The center corresponds to the identity element (rotation by $\psi = 0$ about any axis). All other elements lie on a radial line along $\hat{\mathbf{n}}$, with the distance from the origin being equal to ψ. Because $(\hat{\mathbf{n}}, \psi = \pi)$

[2] A more detailed discussion of Lie groups and Lie algebras is given in Appendix A.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

and $(-\hat{\mathbf{n}}, \psi = \pi)$ are the same group element, antipodal points on the surface of the ball must be identified.

The relation between the two groups is seen by writing Eq. (4.8) as

$$U = \cos(\psi/2) + i\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \sin(\psi/2). \tag{4.10}$$

While the range $0 \leq \psi \leq \pi$ covers all of SO(3), twice that range, $0 \leq \psi \leq 2\pi$, is needed to obtain all of SU(2). With this enlarged range, the two SU(2) matrices U, given by $(\hat{\mathbf{n}}, \psi)$, and $-U$, corresponding to $(-\hat{\mathbf{n}}, 2\pi - \psi)$, map to the same element of SO(3). Thus, SU(2) is a double cover of SO(3).

The relation between the topologies of the two groups can be understood in terms of this mapping. The two elements of SU(2) that are mapped to the same element of SO(3) lie on antipodal points of the three-sphere defined by Eq. (4.9). By taking one element from each such pair, we see that SO(3) corresponds to the upper half of the three-sphere, including the "equator" (which is actually a two-sphere), but with the caveat that antipodal points on the equator must be identified. This yields the previous construction of SO(3) in terms of a three-dimensional ball.[3]

The relation between the two groups can also be understood from a more algebraic point of view. The center of a group is defined as the set of group elements that commute with all elements of the group; this is in fact a subgroup. The center of SU(2) consists of two elements, the identity matrix I, and the matrix $z = -I$, with the latter corresponding to a rotation by 2π about any axis. These form the cyclic group with two elements, Z_2. Now suppose that we define an equivalence relation under which every SU(2) matrix U is equivalent to $zU = -U$. Because z commutes with every element of the group, this equivalence is compatible with the group multiplication, and the equivalence classes themselves form a group, $\mathrm{SU}(2)/Z_2$, which is just SO(3) itself.

Because SU(2) is topologically a three-sphere, we know from Example (iii) of the previous section that it is simply connected; every closed loop can be continuously contracted to a point. What does this tell us about SO(3)? By using the mapping of elements from SU(2) to SO(3), any path on SU(2) can be mapped to a path on SO(3). If the path is a closed loop on SU(2), it is obviously a closed loop on SO(3); since it is contractible on SU(2), it must also be contractible, and homotopic to the trivial loop, on SO(3). But consider a path on SU(2) that starts at some U_0 and ends at the antipodal point, zU_0. This is not a loop in SU(2), but because U_0 and zU_0 are mapped to the same element of SO(3), it is mapped to a closed loop in SO(3). However, this cannot be a contractible loop in SO(3), because that would imply that it could be smoothly deformed to a trivial SO(3) loop, which must correspond to a trivial SU(2) loop.

[3] It may be easier to visualize this by going to one fewer dimension. The upper half of a two-sphere (the upper half of the surface of a globe) is topologically the same as a disk enclosed by a circle, which is the two-dimensional "ball".

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

Hence, SO(3) must have at least two homotopy classes. In fact, that is all that it has. Going around this loop twice in SO(3) corresponds to a path in SU(2) that runs from U_0 to zU_0 and then back to U_0. This is a closed loop in SU(2), and hence must be contractible. Thus, we have

$$\pi_1(\mathrm{SU}(2)) = 0, \tag{4.11}$$

$$\pi_1(\mathrm{SO}(3)) = \pi_1(\mathrm{SU}(2)/Z_2) = Z_2. \tag{4.12}$$

These ideas can be extended to other Lie groups. For every Lie algebra, there is a unique simply connected group, known as the universal covering group. Let G be this group, and let K be either its center or a subgroup of its center. If G is semisimple, K is a finite group. In this case, by defining the elements g and kg to be equivalent, where g and k are arbitrary elements of G and K, we obtain a group G/K that is not simply connected. In fact, by an extension of the arguments given for SU(2) and SO(3),

$$\pi_1(G/K) = K. \tag{4.13}$$

If K is the full center of the covering group, then G/K is known as the adjoint group.

Turning now to the representations of these groups, recall that the irreducible representations of SU(2) can be labeled by a "spin" s that can be either an integer or a half-integer, but that the half-integer spin representations, for which a rotation by 2π is represented by the matrix $-I$, are not true (i.e., single-valued) representations of SO(3). In the general case, the true representations of G/K are those representations of G for which every element of K is represented by a unit matrix.

The groups most often encountered in high energy physics applications are the unitary and orthogonal groups. The group SU(N) is simply connected for any $N \geq 2$. Its center consists of the matrices $e^{2\pi ik/N} I_N$ ($k = 0, 1, \ldots, N-1$), which form the cyclic group Z_N. We have already discussed the representations for $N = 2$. For $N = 3$, recall that any irreducible representation can be constructed from the direct product of p fundamental $\mathbf{3}$ and q antifundamental $\bar{\mathbf{3}}$ representations by suitable symmetrization or antisymmetrization and extraction of traces. The triality of such a representation is defined to be $p - q \pmod 3$. Only the representations with zero triality are representations of the adjoint group SU(3)/Z_3. These ideas can be extended to larger N in an obvious manner, although it should be noted that if N is not prime there are groups intermediate in size between the adjoint group and the covering group.

The group U(1), the multiplicative group with elements $e^{i\alpha}$, is topologically a circle. Hence, we see from Eq. (4.6) that

$$\pi_1(\mathrm{U}(1)) = Z\,, \tag{4.14}$$

so that U(1) is not simply connected. Its simply connected covering group is R, the additive group of real numbers, with $U(1) = R/Z$.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

None of the orthogonal groups are simply connected. SO(2) is identical to U(1), and so its fundamental group is Z. For $N \geq 3$, we have

$$\pi_1(\mathrm{SO}(N)) = Z_2, \qquad N \geq 3. \tag{4.15}$$

The covering group of SO(N) is known as Spin(N). The first few of these are more often recognized in other forms. We have already seen that Spin(3) = SU(2). The next, Spin(4), is identical to SU(2)×SU(2), while Spin(5) = Sp(4), and Spin(6) = SU(4). For further discussion of these and other compact Lie groups, see Appendix A.

As mentioned previously, in perturbative treatments of gauge field theories one is actually only concerned with the Lie algebra, and the distinction between the Lie groups that share that algebra is unimportant. This distinction often matters when using topology to study solitons. There is often some freedom in specifying the symmetry group. For the full symmetry group G of the theory, the only requirement is that all fields in the theory must transform under true representations of G. Hence, G can always be chosen to be the universal covering group, but it can also be taken to be a quotient group if some classes of representations are absent. The choice that is made for G will, however, determine the choice for the unbroken group H.

For future reference, it should be noted that the results in Eqs. (4.12) and (4.13) are instances of a more general result. Let G be a connected and simply connected Lie group, and let H be a subgroup of G. If H is not a connected group, it has a connected subgroup H_0 that contains the identity; if H is connected, $H_0 = H$. The cosets H/H_0 form a group that is often termed the *zeroth homotopy group*, $\pi_0(H)$. Then[4]

$$\pi_1(G/H) = \pi_0(H). \tag{4.16}$$

If H is connected, $\pi_0(H) = 0$. If H is a discrete group, $\pi_0(H) = H$.

4.4 Vortices and homotopy

In Chap. 3, I argued that a configuration with the asymptotic behavior shown in Fig. 3.1a could not be continuously deformed to a uniform vacuum solution, and hence that varying such a configuration until a local minimum of the energy was reached would produce a nontrivial solution. Let us now rephrase this argument in the language of homotopy.

In two dimensions, spatial infinity can be described as a circle at $r = \infty$. As θ varies from 0 to 2π, the values of the field $\phi(r = \infty, \theta)$ on this circle trace out a loop in the vacuum manifold $\mathcal{M}$. Roughly speaking, the argument is that if this loop is in a different homotopy class than the vacuum, then there must be a soliton solution. However, this is not quite right.

[4] For a proof of this theorem, see [49].

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

Homotopy, as defined by Eq. (4.2), is an equivalence relation between two loops that have a common end point. In the field theory context, this would relate two field configurations such that $\phi_1(\infty, 0) = \phi_2(\infty, 0) = \phi_0$ for some fixed ϕ_0. However, our arguments for the existence of a vortex involved continuous deformations of the initial configuration, but without any requirement that there be a point where the value of the field was held fixed. Thus, from a physical point of view our primary interest is in *free homotopy*, where the base point condition is omitted; two loops f and g are said to be freely homotopic if there is a continuous function that satisfies the first two, but not necessarily the third, lines of Eq. (4.2).

If two loops are homotopic with a fixed point, they are obviously also freely homotopic. The converse need not be true, even if the loops share a common base point. This is illustrated in Fig. 4.3. As long as it is attached to the base point x_0, loop g cannot be deformed into loop f. However, it certainly can be deformed into f if it is released from the base point. Hence, the two loops are freely homotopic, but not homotopic with a base point.

These two loops can be related with the aid of loop a; thus, g is homotopic to $a \circ f \circ a^{-1}$. More generally, two loops h and k with a common base point are said to be conjugate if there is a loop c such that h is homotopic to $c \circ k \circ c^{-1}$. This relationship defines a set of *conjugacy classes*. If the fundamental group is Abelian, each conjugacy class contains a single homotopy class, but a conjugacy class can contain several homotopy classes if $\pi_1(\mathcal{M})$ is non-Abelian.

I remarked previously that there is an isomorphism between fundamental groups with different base points, but that this isomorphism is not unique if the fundamental group is non-Abelian; i.e., the mapping of homotopy classes in $\pi_1(\mathcal{M}, x_0)$ to homotopy classes in $\pi_1(\mathcal{M}, y_0)$ may depend on the choice of the path connecting x_0 and y_0. However, this ambiguity is entirely within a conjugacy class; elements of a given conjugacy class at x_0 are always mapped to the same conjugacy class at y_0, regardless of the choice of path.

Another issue to be addressed is gauge invariance. We have seen that to obtain a finite energy vortex we need to work in a gauge theory. This means that many

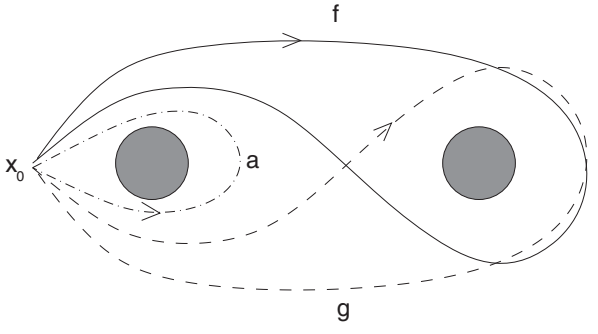


Fig. 4.3. Loops f and g are not homotopic. However, they are freely homotopic, because g is homotopic to afa^{-1}.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

different field configurations can represent the same physical vortex. How do we know that all of these correspond to the same conjugacy class?

Let $\phi_1(r,\theta)$ and $\phi_2(r,\theta)$ be two such configurations, so that

$$\phi_2(r,\theta) = g(r,\theta)\phi_1(r,\theta), \tag{4.17}$$

with the gauge transformation $g(r,\theta)$ being a smooth function. Now define a gauge transformation $k(s)$ that is a continuous function of s such that $k(0) = g^{-1}(r=0)$ and $k(1) = I$; the connectedness of the gauge group guarantees that such a function exists. Then [50]

$$F(s,\theta) = g\left(\frac{s}{1-s},\theta\right) k(s)\phi_1(\infty,\theta) \tag{4.18}$$

is a smooth function of s and θ with the property that

$$\begin{aligned} F(0,\theta) &= \phi_1(\infty,\theta), \\ F(1,\theta) &= \phi_2(\infty,\theta). \end{aligned} \tag{4.19}$$

Hence, $\phi_1(\infty,\theta)$ and $\phi_2(\infty,\theta)$ are freely homotopic, and our two gauge-equivalent configurations correspond to the same conjugacy class.

Let us briefly summarize where we are at this point. Every finite energy configuration can be assigned to a conjugacy class according to the behavior of the scalar field at spatial infinity, with this assignment being invariant under nonsingular gauge transformations. Configurations within the same conjugacy class can be smoothly deformed into each other, but those in different classes cannot be. If there is only a single conjugacy class, there are no topologically stable vortices. If there is more than one conjugacy class, then a configuration that minimizes the energy within a nontrivial class (i.e., a class that does not contain the trivial constant configuration) gives a topologically stable vortex solution.

This connects the existence of topological vortices to the existence of a fundamental group with more than one element. However, we have not yet connected the group structure of $\pi_1(\mathcal{M})$ to the physical properties of vortices. To do this, we must consider configurations that correspond to assemblies of several vortices.

To this end, consider a gauge theory with one or more types of vortex solutions. Let us assume that, as with the gauged U(1) vortex of Chap. 3, the energy density of each vortex is concentrated within a well-defined region of finite area, and that outside this region the Higgs field ϕ rapidly approaches a (spatially nonuniform) vacuum solution. The interactions between well-separated vortices will then be relatively weak, and one can envision assembling a number of them together to form a multivortex configuration.

Before this can be done, their asymptotic behaviors must be made compatible. For example, the U(1) vortex solutions of the previous chapter cannot be smoothly joined when each is written in a gauge where they take the form of Eq. (3.32). However, they can be combined if they are first gauge-transformed

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

into a form, such as that in Fig. 3.3b, where ϕ has a fixed phase θ_0 outside a wedge containing the vortex core.

More generally, consider a theory with vortices corresponding to nontrivial elements of $\pi_1(\mathcal{M})$. Let us assume that as one traverses a large circle enclosing vortex 1 the field ϕ traces out a loop in the vacuum manifold $\mathcal{M}$ that begins and ends at ϕ_1, that going around a circle enclosing vortex 2 the field traces out a loop beginning and ending at ϕ_2, and so on. In the U(1) theory, where $\mathcal{M}$ is a circle, one can choose the starting points of these loops so that the ϕ_j are all the same. This is not true in general, since for a larger $\mathcal{M}$ the various loops need not have any values of ϕ in common. However, as long as $\mathcal{M}$ is connected, the various ϕ_j can each be smoothly connected by a path f_j in $\mathcal{M}$ to a common value ϕ_0.

We can then use the following prescription to construct a multivortex configuration. To begin, deform each vortex so that ϕ lies on the vacuum manifold outside a circle of finite radius surrounding the vortex core. Next, gauge-transform the field of vortex j so that $\phi = \phi_j$ everywhere outside a wedge-shaped region that contains the vortex core. Next, surround this wedge with a larger wedge and gauge transform ϕ so that it smoothly varies along the path f_j from ϕ_j to ϕ_0 as one goes from the inner wedge to the outer wedge. Outside the outer wedge for each vortex $\phi = \phi_0$, so there is no problem in assembling the wedges into a smooth configuration

Figure 4.4 illustrates this for the case of three wedges. As one moves from the point A_j to B_j, the field traces out a loop in $\mathcal{M}$ corresponding to an element $[h_j]$ of $\pi_1(\mathcal{M}, \phi_j)$. Going along the path $C_j A_j B_j D_j$ then gives the corresponding element $[h_j]$ of $\pi_1(\mathcal{M}, \phi_0)$. Traversing the full loop indicated in the figure gives the product

$$[h_{\text{tot}}] = [h_1] \circ [h_2] \circ [h_3]. \tag{4.20}$$

For the U(1) theory of the previous chapter, with $\pi(\mathcal{M}) = Z$, this product is just the addition of the vorticities.

There were two points in this construction where somewhat arbitrary choices were made. One of these was the choice of the path f_j in $\mathcal{M}$ linking ϕ_j and ϕ_0, thus defining a map between the homotopy groups with different base points. As noted previously, such mappings may not be unique, and another choice for this path could have yielded a different element of $\pi_1(\mathcal{M}, \phi_0)$, although one that was in the same conjugacy class.

The second arbitrary choice was in the orientation of the wedges for the vortices. For example, if the wedge for vortex 2 in Fig. 4.4 had been oriented downward, Eq. (4.20) would have been replaced by

$$[h_{\text{tot}}] = [h_1] \circ [h_3] \circ [h_2]. \tag{4.21}$$

Neither of these ambiguities is of any consequence if the homotopy group is Abelian. The mapping between $\pi_1(\mathcal{M}, \phi_1)$ and $\pi_1(\mathcal{M}, \phi_0)$ is then path-independent and unique, and the products in Eqs. (4.20) and (4.21) are equal.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

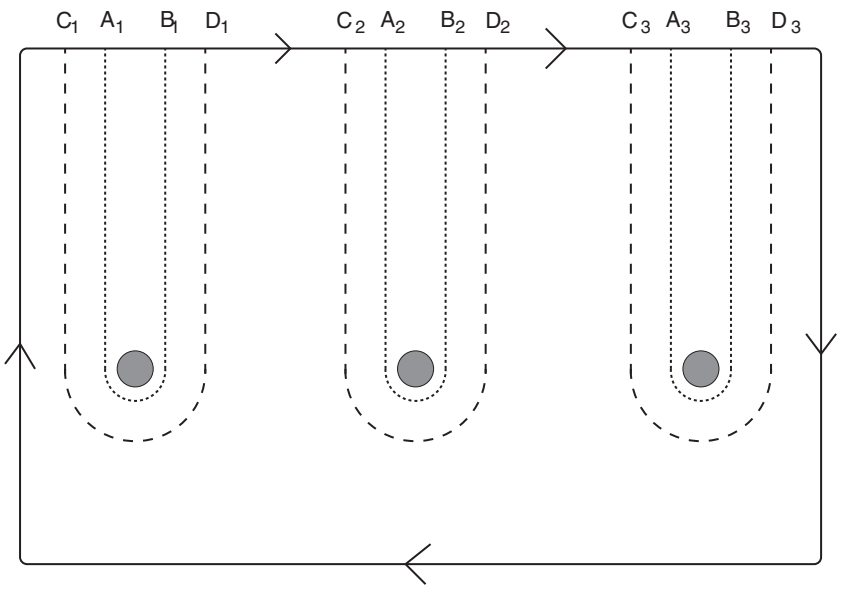


Fig. 4.4. The patching together of three vortex solutions, as described in the text. The shaded circles, running left to right, represent the cores of vortices 1, 2, and 3. Along the dotted lines the scalar field is equal to ϕ_1, ϕ_2, and ϕ_3, respectively. Along the dashed lines, and in the region outside these dashed lines, the field is equal to ϕ_0.

The elements of $\pi_1(\mathcal{M})$ define topological charges for the vortices that combine unambiguously in multivortex configurations. The possibilities when $\pi_1(\mathcal{M})$ is non-Abelian are more complex and will not be described in detail here. A detailed discussion of these can be found in [49].

4.5 Some illustrative vortex examples

It may be useful to illustrate the results of the previous sections with some examples.

(i) *U(1) = SO(2) symmetry broken by a complex scalar field*: This is the case studied in detail in Chap. 3. The vacuum expectation value of the Higgs field is of the form $\langle\phi\rangle = ve^{i\alpha}$, with $0 \leq \alpha < 2\pi$, so the vacuum manifold $\mathcal{M} = S^1$, and $\pi_1(\mathcal{M}) = \pi_1(S^1) = Z$. [Alternatively, we could make use of Eq. (4.16) by taking the gauge group G to be the covering group of U(1), which is the additive group of the real numbers, and the unbroken subgroup H to be the additive group of the integers.] The elements of $\pi_1(\mathcal{M})$ can be identified with the vorticity defined by Eq. (3.4). This was normalized so that it takes on integer values, and is additive when vortices are combined.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

(ii) *SO(N) symmetry broken to SO(N − 1)*: This can be achieved with a scalar field ϕ that transforms under the vector representation of SO(N) and has a vacuum expectation value that satisfies an equation of the form $\sum_{a=1}^{N} \phi_a^2 = v^2$. This defines an $(N-1)$-sphere, and so

$$\mathcal{M} = \mathrm{SO}(N)/\mathrm{SO}(N-1) = S^{N-1}. \tag{4.22}$$

For $N = 2$ we have the previous example, but for $N \geq 3$ we have $\pi_1(\mathcal{M}) = 0$, the trivial group, and there are no topologically stable vortices. This can also be seen by examining Fig. 3.1a. For $N = 2$, the arrows are confined to the plane of the paper and cannot be smoothly rotated to be all parallel. For $N = 3$, these arrows live in three dimensions and can be rotated to be perpendicular to the plane of the paper, giving a topologically trivial configuration that can be smoothly deformed into a pure vacuum.

(iii) *Z_2 vortices*: Consider an SO(3) gauge theory with two scalar fields, $\boldsymbol{\phi}$ and $\boldsymbol{\chi}$, that transform as SO(3) vectors governed by the scalar field potential

$$V(\boldsymbol{\phi}, \boldsymbol{\chi}) = \frac{\lambda_\phi}{4}(\boldsymbol{\phi}^2 - v_\phi^2)^2 + \frac{\lambda_\chi}{4}(\boldsymbol{\chi}^2 - v_\chi^2)^2 + g(\boldsymbol{\phi}\cdot\boldsymbol{\chi})^2. \tag{4.23}$$

The sign of g determines the relative orientation of $\boldsymbol{\phi}$ and $\boldsymbol{\chi}$ at the minima of V. If it is negative, these must be parallel, and the symmetry is broken to U(1). If instead $g > 0$, $\boldsymbol{\phi}$ and $\boldsymbol{\chi}$ must be orthogonal vectors of lengths v_ϕ and v_χ, respectively, and the SO(3) symmetry is completely broken. Hence, $\mathcal{M} = G/H = SO(3)$, and $\pi_1(\mathcal{M}) = Z_2$.[5] We thus have Z_2 vortices, with the property that a combination of two vortices is topologically trivial and can be smoothly deformed to the vacuum.

In this latter case we can write down a rotationally invariant ansatz,

$$\begin{aligned} \boldsymbol{\phi} &= v_\phi\,(0,0,1), \\ \boldsymbol{\chi} &= v_\chi f(r)\,(\cos\theta, \sin\theta, 0), \\ \mathbf{A}_j &= \epsilon_{jk}\hat{x}^k \frac{a(r)}{r}\,(0,0,1), \end{aligned} \tag{4.24}$$

in which $\boldsymbol{\phi}$ is constant in space while $\boldsymbol{\chi}$ and $\mathbf{A}_j$ are essentially embeddings of the U(1) Higgs and gauge fields into the larger group. We could just as easily use an ansatz in which $\boldsymbol{\chi}$ was constant and $\boldsymbol{\phi}$ was the field whose orientation was twisted from point to point. It is easy to see that there are solutions of both types (and possibly others, as well). Stability is a more difficult issue, and a detailed study would be needed to see whether the heavier of these two solutions is stable under small perturbations against decay to a lighter vortex.

[5] Alternatively, we could take G to be the covering group SU(2), in which case H would be the Z_2 subgroup composed of the two elements I and –I. The result for $\pi_1(\mathcal{M})$ would then follow from Eq. (4.16).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

Now consider the ansatz

$$\begin{aligned} \boldsymbol{\phi} &= v_\phi\,(0,0,1), \\ \boldsymbol{\chi} &= v_\chi f(r)\,(\cos n\theta, \sin n\theta, 0), \\ \mathbf{A}_j &= \epsilon_{jk}\hat{x}^k \frac{a(r)}{r}\,(0,0,1), \end{aligned} \tag{4.25}$$

that corresponds to an embedding of a vorticity n solution of the U(1) theory. Because we have a Z_2 topological charge, this solution has vanishing topological charge if n is even, and unit charge if n is odd. In particular, the $n = -1$ antivortex solution is topologically equivalent to the vortex solution of Eq. (4.24). But let us suppose that the two Higgs fields have very different energy scales, with $v_\phi \gg v_\chi$, while $\lambda_\phi \sim \lambda_\chi$. Because $\langle\boldsymbol{\phi}\rangle$ breaks SO(3) to U(1), at energies well below v_ϕ one would appear to have a U(1) gauge theory with a single complex Higgs field formed from the components of $\boldsymbol{\chi}$ orthogonal to $\langle\boldsymbol{\phi}\rangle$. That theory would have vortices with ordinary integer topological charge, and solutions with different values of n could not be deformed into one another. The vortex and the antivortex would not be equivalent.

These two points of view are reconciled by considering the homotopy that takes an $n = 2$ SO(3) configuration to the vacuum. In the course of this homotopy, the untwisting of the $\boldsymbol{\chi}$ field must be accompanied, at intermediate stages, by a twisting of the $\boldsymbol{\phi}$ field. At large distances, where the covariant derivatives vanish, this can be accomplished by a gauge transformation and is unproblematic. However, within the vortex core there is gradient energy associated with the twisting of $\boldsymbol{\phi}$. Because the scale of this energy is set by v_ϕ, there is a large potential energy barrier that must be traversed. As a result, the $n = 2$ configuration, although not topologically stable, is dynamically protected from decay into the vacuum sector.[6]

This illustrates that while topology is an important guide, it is not the whole story.

(iv) *Weinberg–Salam theory and semilocal vortices*: The electroweak interactions of the Standard Model have a gauged SU(2) symmetry, with coupling constant g, and a gauged U(1) symmetry, with coupling constant g'. The Higgs field is a complex doublet ϕ. The minima of $V(\phi)$ occur when $\phi^\dagger\phi = v^2/2$, where $v \neq 0$ is determined by the Higgs field mass and self-coupling. Writing $\phi = (\phi_1, \phi_2)^t$, we have

$$(\mathrm{Re}\,\phi_1)^2 + (\mathrm{Im}\,\phi_1)^2 + (\mathrm{Re}\,\phi_2)^2 + (\mathrm{Im}\,\phi_2)^2 = \frac{v^2}{2}. \tag{4.26}$$

This is the equation for a three-sphere. Since $\pi_1(S^3) = 0$, there are no topological vortices in this theory.

[6] The protection against decay is only absolute at the classical level. The decay could proceed quantum mechanically via tunneling through the potential energy barrier, although the rate for this would be exponentially small.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

However, there are some interesting related possibilities. First, consider a similar theory, but with vanishing SU(2) coupling g. There would then be a local U(1) symmetry, but only a global SU(2) symmetry. As we have seen, finiteness of the energy requires that the covariant derivative of the scalar field falls sufficiently rapidly at large distance. This could be achieved if the long-distance twisting of ϕ were entirely within the gauged U(1) group, but not if ϕ were rotated by elements of the SU(2), because the effects of the gradients could not be compensated by a coupling to an SU(2) gauge field. Consequently, a vortex configuration with

$$\phi = \begin{pmatrix} 0 \\ f(r)e^{i\theta} \end{pmatrix} \tag{4.27}$$

and an appropriate U(1) gauge field might actually be stable, because the homotopy connecting the asymptotic field to a uniform vacuum configuration would pass through configurations with SU(2) twisting of ϕ, and thus infinite energy. Configurations such as these have been termed *semilocal vortices* [51].

However, depending on the parameters of the theory, a configuration such as that in Eq. (4.27) could be unstable even without unwinding. The usual topological argument for the stability of such a configuration uses the fact that while avoiding a singularity at the origin forces $f(0)$ to vanish, the need to minimize the energy from the potential forces f to rapidly approach v as r increases. Now suppose that instead of setting $\phi_1 = 0$ we consider configurations of the form

$$\phi = \begin{pmatrix} g(r) \\ f(r)e^{i\theta} \end{pmatrix} \tag{4.28}$$

with g not necessarily vanishing. The potential is minimized if $f^2 + g^2 = v^2$, so there is not necessarily any energetic penalty in letting f be small, or even vanishing, as long as there is a compensating increase in g. Hence, a configuration with nonzero vorticity in f might be unstable against the transformation of the point zero at the origin into an ever increasing region of vanishing f. At spatial infinity the measured vorticity would remain nonzero, but it would eventually become unobservable at any finite r. A detailed numerical analysis shows that such an instability is in fact present if the scalar self-coupling is large compared to the U(1) gauge coupling. On the other hand, for small scalar self-coupling this instability is absent and semilocal vortices can exist [52–54].

Still focusing on this latter case, let us restore the SU(2) gauge coupling, but with $g \ll g'$. Although there is no longer an infinite energy barrier preventing the unwinding of the field of Eq. (4.27), one might well expect there to be a finite barrier that would preserve the semilocal vortex. This has been confirmed by numerical analysis. However, this analysis also shows that this barrier disappears before g reaches the observed electroweak value. Hence, the SU(2)×U(1) electroweak theory does not have stable vortex solutions [55–57].

(v) *Alice strings*: Consider an SO(3) gauge theory with a Higgs field transforming according to the five-dimensional irreducible representation [58]. This field can be represented as a traceless symmetric 3×3 matrix that transforms as

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

$$\phi \to R\phi R^t, \tag{4.29}$$

where R is an SO(3) rotation matrix. Now let $V(\phi)$ be such that it is minimized when two of the eigenvalues of ϕ are equal, so that ϕ is of the form

$$\phi = R \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -2a \end{pmatrix} R^t, \tag{4.30}$$

or, equivalently,

$$\phi = a(I - 3\mathbf{e}\mathbf{e}^t), \tag{4.31}$$

with $\mathbf{e}$ a real unit three-vector. Thus each possible vacuum is specified by a unit vector $\mathbf{e}$, but with the caveat that $-\mathbf{e}$ and $\mathbf{e}$ specify the same vacuum. To identify the vacuum manifold $\mathcal{M}$, we first note that the space of unit three-vectors is the unit two-sphere. Making $\mathbf{e}$ and $-\mathbf{e}$ equivalent corresponds to identifying antipodal points on the two-sphere, giving a space that is the same as a disk (a circle and the area it encloses) with antipodal points on its boundary identified. A closed loop on this space can include a jump from one point on the boundary to the antipodal point, as shown in Fig. 4.5. By continuous deformations of the path (e.g., by bringing $B \sim B'$ and $C \sim C'$ together and then moving the combined point away from the disk boundary) one can subtract or add additional jumps, but only in pairs. Thus, there are two homotopy classes of loops: those which include an odd number of such jumps, and those with either an even number or no jump at all. The homotopy group $\pi_1(\mathcal{M})$ is therefore Z_2. [A similar argument gives an alternative demonstration that the fundamental group of SO(3) is Z_2.]

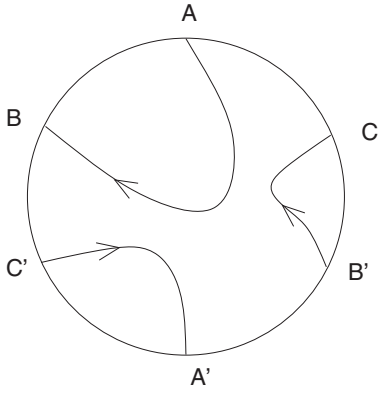


Fig. 4.5. A closed loop on a disk with antipodal points identified. As drawn, the loop has three "jumps". If B′ and C (and thus B and C′) are brought together, two of these can be eliminated. However, it is never possible to subtract or add an odd number of jumps.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

We can identify the unbroken gauge group by considering the vacuum with

$$\phi = \phi_0 \equiv \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & -2a \end{pmatrix} \equiv a(I - \mathbf{e}_0\mathbf{e}_0^t). \tag{4.32}$$

The unbroken subgroup H clearly includes the U(1) subgroup consisting of rotations about $\mathbf{e}_0$; i.e., about the 3-axis. However, it also includes another component, disconnected from the first, consisting of rotations by π about any axis perpendicular to the 3-axis. Because these rotations by π do not commute with the elements of the U(1), H is not the direct product $\mathrm{U}(1)\times Z_2$, but rather the semidirect product group $\mathrm{U}(1)\rtimes Z_2 = \mathrm{Pin}(2)$.

Now consider a vortex solution in which ϕ at large distances is of the form

$$\begin{aligned} \phi(r,\theta) &\approx R(\theta)\phi_0 R(\theta)^{-1} \\ &= a[I - 3\mathbf{e}(\theta)\mathbf{e}(\theta)^t], \end{aligned} \tag{4.33}$$

where $\mathbf{e}(\theta) = R(\theta)\mathbf{e}_0$. This will be topologically stable if $R(\theta)$ traces out an uncontractible loop as θ ranges from 0 to 2π; i.e., if $\mathbf{e}(2\pi) = -\mathbf{e}(0)$. At each value of θ the unbroken U(1) subgroup is the set of rotations about $\mathbf{e}(\theta)$, generated by a charge operator $Q(\theta)$. If one makes a full circuit of the vortex, the result is a reversal of $\mathbf{e}$, and thus a change in the sign of Q. Hence, the U(1) charges are double-valued and the sign of a particle's charge is ambiguous. This can be represented by imagining a branch cut starting at the vortex and either ending at another vortex or running out to infinity, and saying that a particle's charge changes sign when the particle crosses the branch cut. The exact location of the branch cut is not physically meaningful and not gauge invariant, but the existence of a cut is. The vortices (or rather, the strings that are their extension into three dimensions) that give rise to this phenomenon have been termed "Alice strings", in an allusion to Lewis Carroll's novel [58].

Even though the sign of a charge is not gauge-invariant, it is physically meaningful to ask whether a pair of particles have the same or opposite charges (e.g., the two particles cannot annihilate if they have the same charge, but might be able to if they have opposite charges). However, suppose that we have an Alice string in three spatial dimensions that closes on itself to make a circular loop. The branch cut in two spatial dimensions becomes a surface in three dimensions that is bounded by the string loop, but whose location is otherwise arbitrary. Now start with two particles with equal charges and take one of them along a path that goes around the string loop and then returns to its original position. This reverses the sign of the particle's charge, so it would seem that the two particles have opposite sign, and so zero net charge. On the other hand, the electric flux through a Gaussian surface enclosing both the string loop and the two particles would be unchanged, as should be expected from charge conservation. It is tempting to try to reconcile these facts by saying that the U(1) charge of the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

particle was transferred to the string loop when the particle crossed the surface of branch cuts. However, the time of this transfer would depend on the arbitrary choice for locating the surface. Instead, we are led to the conclusion that we can assign a physically meaningful gauge-invariant charge to the entire system comprising the two particles and the loop, but that this charge cannot be localized within the system in a gauge-invariant manner. In another allusion to Carroll, such charges have been termed "Cheshire charges" [59].

For further discussion of Alice strings and Cheshire charges, see [60–62].

(vi) *A non-Abelian fundamental group*: Consider the same SO(3) gauge theory as in the previous example, but with a potential that is minimized when the eigenvalues of ϕ are all distinct, so that

$$\phi = R \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & a_3 \end{pmatrix} R^t, \tag{4.34}$$

with $a_1 \neq a_2 \neq a_3$. This is invariant only under the identity or rotations by π about either the x-, the y-, or the z-axis, so the unbroken subgroup of SO(3) is now a discrete group with four elements.

For determining $\pi_1(\mathcal{M})$ it is actually more convenient to take the original gauge group G to be the covering group, SU(2), so that we can make use of Eq. (4.16). This doubles the size of the unbroken group H, which now has eight elements, corresponding to the SU(2) matrices $\pm I$, $\pm\sigma_x$, $\pm\sigma_y$, and $\pm\sigma_z$. These form the quaternion group, Q, which is clearly non-Abelian. Because $H = Q$ is discrete, Eq. (4.16) gives

$$\pi_1(G/H) = \pi_0(H) = Q \tag{4.35}$$

and we have a non-Abelian fundamental group, as promised. Vortices in this theory are discussed in [63].

4.6 Higher homotopy groups

There is a natural generalization of the fundamental group. The latter classifies closed loops, which are maps from a circle, S^1, to a given manifold. The higher homotopy groups, $\pi_n(\mathcal{M})$, classify maps from an n-sphere, S^n, to the manifold.

Let us first consider the second homotopy group, $\pi_2(\mathcal{M}, x_0)$. Although the base point x_0 is explicitly indicated, we will see below that, as with π_1, the group is independent of the choice of base point. This is the natural object for classifying particle-like solitons in three spatial dimensions, where spatial infinity is a two-sphere. (The line-like strings that are the extensions of vortices to three dimensions are still classified by π_1.)

We start by mapping the two-sphere to a finite region of the plane. One way to visualize this is to imagine puncturing a hole in the surface of a balloon,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

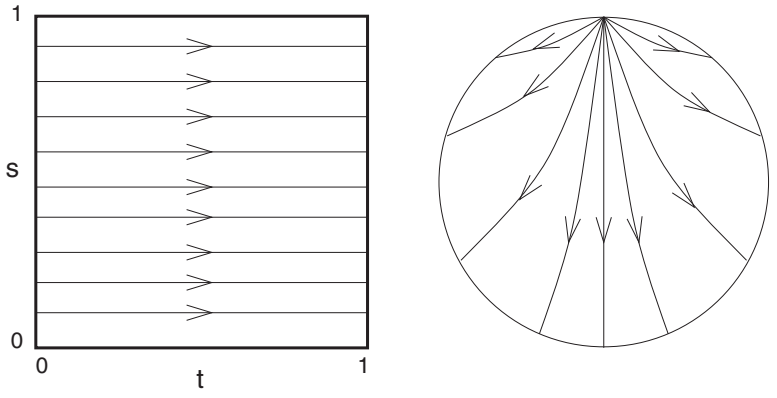


Fig. 4.6. The mapping from a square to a sphere. The entire perimeter of the square is mapped onto the north pole of the sphere, and the horizontal lines of constant s are mapped onto the curves shown on the sphere. As one goes along each of these curves, t runs from 0, at the north pole, to 1/2, at the edge of the diagram, and then to 1 as the curve returns to the north pole on the (hidden) opposite side of the sphere.

and then stretching and flattening out this surface until the infinitesimal circle surrounding the puncture becomes the perimeter of a square. This square can be covered by coordinates s and t, with $0 \le s, t \le 1$, as illustrated in Fig. 4.6. A smooth function on the sphere becomes a smooth function of s and t with the additional constraint that it has the same value everywhere on the perimeter; i.e., $f(0,t) = f(1,t) = f(s,0) = f(s,1)$. A convenient mapping of these coordinates back to the sphere is shown in Fig. 4.6. Here s labels the loops, with $s = 0$ and $s = 1$ being degenerate loops that reduce to a point and $s = 1/2$ being the largest loop. The distance along each loop is parameterized by t.

Homotopy is defined by generalizing Eq. (4.2). Let $f(s,t)$ and $g(s,t)$ be two maps, both of which are equal to x_0 everywhere on the perimeter of the square. These are homotopic if there is a function $k(s,t,u)$ with $0 \le s, t, u \le 1$ such that

$$\begin{aligned} k(s,t,0) &= f(s,t), \\ k(s,t,1) &= g(s,t), \\ k(s,0,u) = k(s,1,u) &= k(0,t,u) = k(1,t,u) = x_0. \end{aligned} \tag{4.36}$$

Just as before, this relation defines a set of homotopy classes.

The product of two maps can be defined by

$$(f \circ g)(s,t) = \begin{cases} f(s,2t), & 0 \le t \le 1/2\ , \\ \\ g(s,2t-1), & 1/2 \le t \le 1\ . \end{cases} \tag{4.37}$$

This is illustrated graphically in Fig. 4.7. As was true with loops, this product can be carried over to homotopy classes. We want to show that these

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

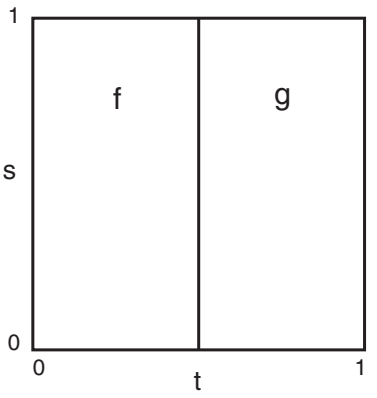


Fig. 4.7. Diagrammatic illustration of the product of two maps $f(s,t)$ and $g(s,t)$. The functions are equal to x_0 everywhere on the heavy solid lines.

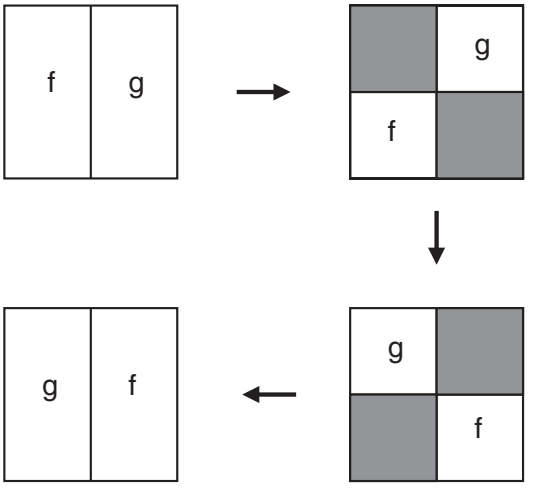


Fig. 4.8. A demonstration that π_2 is always Abelian. The functions are equal to x_0 everywhere on the heavy solid lines and in the dark shaded regions. The transformations indicated by the arrows are all smooth deformations of one map into the next.

homotopy classes with this product form a group. Associativity is straightforward to demonstrate. The identity element is the homotopy class containing the constant map $f_0(s,t) = x_0$. The existence of an inverse follows from the observation that $f(s,t) \circ f(s,-t)$ is homotopic to the identity map. Thus, we have a based homotopy group $\pi_2(\mathcal{M}, x_0)$.

However, note the series of maps shown in Fig. 4.8. Each is clearly homotopic to the previous one, with the net result being that $f \circ g$ and $g \circ f$ are homotopic. Thus, in contrast with the first homotopy group, π_2 is always Abelian.

As with the first homotopy group, $\pi_2(\mathcal{M}, x_0)$ does not depend on the choice of x_0, provided that $\mathcal{M}$ is a connected manifold. In the case of π_1, this could

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

be seen by considering a path connecting two possible base points. For π_2, this path must be inflated to become a thin tube, but the demonstration is otherwise analogous. We can therefore drop the reference to the base point, and simply write $\pi_2(\mathcal{M})$.

Nevertheless, just as with π_1, the isomorphism between $\pi_2(\mathcal{M}, x_0)$ and $\pi_2(\mathcal{M}, y_0)$ can depend on the path of the tube from x_0 to y_0. If two such paths can be combined to form a noncontractible loop from x_0 to y_0 and back to x_0, then there will be maps that can be deformed into each other if the base point condition is removed, but not otherwise. Thus, these maps will be freely homotopic, but not homotopic. This situation can arise even if the first homotopy group of $\mathcal{M}$ is Abelian, although it is necessary that this group be nontrivial [48].

When ambiguities of this sort are absent, three-dimensional particle-like solitons can be associated with a unique homotopy class, with an associated additive topological charge. On the other hand, if there is an ambiguity, the soliton is associated with a set of homotopy classes, and there are subtleties in defining the topological charges of multisoliton configurations. This is seen, for example, in configurations containing both magnetic monopoles and Alice strings [61].

This discussion of π_2 can be carried over, with the obvious generalizations, to the π_n with $n \geq 3$. Like π_2, these higher homotopy groups are always Abelian. In particular, we will encounter π_3 in our study of Yang–Mills instantons.

4.7 Some results for higher homotopy groups

In our later discussions we will need a number of results concerning the homotopy groups of spheres. Let us begin with $\pi_n(S^n)$, which classifies maps from one n-sphere to another. For $n = 1$, we saw previously that $\pi_1(S^1) = Z$, with the integer associated with a map being the winding number N that counts the number of times that the first S^1 winds around the second. If the mapping at a point is counted with a positive or negative sign according to whether the two circles are being traversed in the same or opposite directions, respectively, then the total count at every point on the target circle is N. Explicitly, if θ denotes the angle on the first circle, and $\alpha(\theta)$ the angle to which this is mapped on the second circle, the winding number that was defined in Eq. (3.4) can be written as

$$N = \frac{1}{2\pi} \int_0^{2\pi} d\theta \, \frac{d\alpha}{d\theta}. \tag{4.38}$$

A winding number can be defined in an exactly analogous fashion for maps from S^2 to S^2. Using the standard spherical coordinates θ and ϕ for the first sphere and $\alpha(\theta, \phi)$ and $\beta(\theta, \phi)$ for the second, we can define the winding number

$$\begin{aligned} N &= \frac{1}{4\pi} \int d^2\Omega \, \frac{\sin\alpha}{\sin\theta} \left(\frac{d\alpha}{d\theta} \frac{d\beta}{d\phi} - \frac{d\beta}{d\theta} \frac{d\alpha}{d\phi} \right) \\ &= \frac{1}{4\pi} \int d\theta \, d\phi \, \sin\alpha \left(\frac{d\alpha}{d\theta} \frac{d\beta}{d\phi} - \frac{d\beta}{d\theta} \frac{d\alpha}{d\phi} \right). \end{aligned} \tag{4.39}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

It is useful to rewrite this in terms of Cartesian components. Let us define the unit vector

$$\hat{\mathbf{e}}(\mathbf{r}) = (\sin\alpha\cos\beta, \sin\alpha\sin\beta, \cos\alpha), \tag{4.40}$$

with $\alpha(\mathbf{r})$ and $\beta(\mathbf{r})$ functions of three-dimensional Euclidean space. Converting Eq. (4.39) to Cartesian coordinates, we find that the winding number of $\hat{\mathbf{e}}(\mathbf{r})$ on a sphere of fixed radius is

$$\begin{aligned} N &= \frac{1}{8\pi}\epsilon^{ijk}\epsilon^{abc}\int dS^i\, \hat{e}^a\, (\partial_j \hat{e})^b\, (\partial_k \hat{e})^c \\ &= \frac{1}{8\pi}\epsilon^{ijk}\int dS^i\, \hat{\mathbf{e}}\cdot\partial_j\hat{\mathbf{e}}\times\partial_k\hat{\mathbf{e}}, \end{aligned} \tag{4.41}$$

where dS^i is the surface element on the sphere.

This integral is invariant under smooth variation of $\hat{\mathbf{e}}$. To see this, consider the variation $\hat{\mathbf{e}} \to \hat{\mathbf{e}}' = \hat{\mathbf{e}} + \mathbf{v}$ where $\mathbf{v}$ is infinitesimal. Because $\hat{\mathbf{e}}$ and $\hat{\mathbf{e}}'$ are both unit vectors, we must require that $\mathbf{v}$, like $\partial_k\hat{\mathbf{e}}$, be orthogonal to $\hat{\mathbf{e}}$. To first order in $\mathbf{v}$, the change in N is

$$\begin{aligned} \delta N &= \frac{1}{8\pi}\epsilon^{ijk}\int dS^i\, (2\hat{\mathbf{e}}\cdot\partial_j\hat{\mathbf{e}}\times\partial_k\mathbf{v} + \mathbf{v}\cdot\partial_j\hat{\mathbf{e}}\times\partial_k\hat{\mathbf{e}}) \\ &= \frac{1}{4\pi}\epsilon^{ijk}\int dS^i\, \hat{\mathbf{e}}\cdot\partial_j\hat{\mathbf{e}}\times\partial_k\mathbf{v} \\ &= \frac{1}{4\pi}\epsilon^{ijk}\int dS^i\, \partial_j\,(\hat{\mathbf{e}}\cdot\mathbf{v}\times\partial_k\hat{\mathbf{e}}) \\ &= 0. \end{aligned} \tag{4.42}$$

The last term on the first line vanishes because the three factors are all orthogonal to $\hat{\mathbf{e}}$, and the last line follows from the fact that the integration is over a surface without boundary.

Furthermore, N is invariant under perturbations of the integration surface, as long as $\hat{\mathbf{e}}$ remains well defined. Hence, if we have a field $\boldsymbol{\phi}(\mathbf{r})$ that transforms as an SO(3) vector and define $\hat{\mathbf{e}} = \boldsymbol{\phi}/|\boldsymbol{\phi}|$, the winding number of $\hat{\mathbf{e}}$ over a surface is invariant under deformations of the surface that do not take it through a zero of $\boldsymbol{\phi}$. Arguments analogous to those for the vortex case then show that N is equal to the total number of zeros of $\boldsymbol{\phi}$ in the region enclosed by the surface of integration, with each zero being counted with a plus or minus sign according to the sign of the winding on an infinitesimal sphere enclosing the zero.

These results are readily generalized to arbitrary n, so we have

$$\pi_n(S^n) = Z. \tag{4.43}$$

There is a second result that generalizes from the $n = 1$ case. By removing a point from S^n to obtain R^n, and noting that R^n is simply connected, we showed that all maps from S^1 into S^n with $n > 1$ are contractible. Essentially the same argument shows that maps from any sphere to a higher-dimensional sphere are contractible; i.e., that

$$\pi_k(S^n) = 0, \qquad n > k. \tag{4.44}$$

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

The situation for maps from a larger sphere to a smaller one is less simple. For all $k > 1$, $\pi_k(S^1) = 0$, but for $1 < n < k$ the group $\pi_k(S^n)$ is nontrivial, with a more complicated dependence on k and n.

We will also make use of some results concerning higher homotopy groups of Lie groups. First, a theorem due to Cartan [64] implies that

$$\pi_2(G) = 0, \quad G \text{ compact, connected,} \tag{4.45}$$

for any compact connected Lie group G. Next, recall that the group SU(2) is topologically equivalent to a three-sphere. Setting $n = 3$ in Eq. (4.43), we have

$$\pi_3(\mathrm{SU}(2)) = Z. \tag{4.46}$$

SU(2) is a subgroup of every compact simple Lie group G, so every topologically nontrivial map from the three-sphere to SU(2) can also be viewed as a map from the three-sphere to G. Because G is a larger manifold, one might wonder if this map could be untwisted in G, even though it could not be untwisted in SU(2). The answer turns out to be no and we have

$$\pi_3(G) = Z, \quad G \text{ compact, connected, and simple.} \tag{4.47}$$

Finally, there is an important theorem concerning $\pi_2(G/H)$ that will prove to be very useful in our study of monopoles in the next chapter. It is analogous to the theorem for the fundamental group that was given in Eq. (4.16). One route [50] to obtaining this result is as follows. Let ϕ be a scalar field transforming under a gauge group G, and let H be the subgroup that leaves a particular vacuum value, ϕ_0, invariant. We are interested in maps from a two-sphere to the manifold of vacua, G/H. Any such map is given by a function $\phi(s,t)$, with $\phi = \phi_0$ when either s or t is equal to 0 or 1. Since all vacuum values of ϕ are related by gauge transformations, we can write

$$\phi(s,t) = g(s,t)\phi_0, \tag{4.48}$$

where $g(s,t)$ is an element (or rather a representation of an element) of G. However, $g(s,t)$ is not uniquely determined, since $g(s,t)h(s,t)\phi_0 = g(s,t)\phi_0$ for any element $h(s,t)$ of H.

Now consider the parameterization of the sphere given by Fig. 4.6. We can arbitrarily set $g = I$ at the beginning of each loop; i.e., $g(s,0) = I$. Also, because the $s = 0$ and $s = 1$ loops are simply points, $g(0,t) = g(1,t) = I$. Thus, $g = I$ on three sides of the square. On the fourth side, all that we know is that $g(s,1)\phi_0 = \phi_0$, which implies that

$$g(s,1) \equiv h(s) \tag{4.49}$$

is an element of H. The previous conditions imply that $h(0) = h(1) = I$, so $h(s)$ defines a loop in H. We thus have a correspondence between a map $\phi(s,t)$ from S^2 to G/H and a map $h(s)$ from S^1 to H.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005

Now suppose that there are two such maps, $\phi_1(s,t)$ and $\phi_2(s,t)$, that yield the same loop $h(s)$. If we write

$$\begin{aligned}\phi_1(s,t) &= g_1(s,t)\phi_0,\\ \phi_2(s,t) &= g_2(s,t)\phi_0,\end{aligned} \tag{4.50}$$

then $g_1 = g_2$ everywhere on the boundary of the square. Hence, $g_3(s,t) \equiv g_2^{-1}(s,t)g_1(s,t)$ is equal to the identity everywhere on the boundary, and thus gives a map from the two-sphere to G. If this map is homotopic to the identity map, $g(s,t) = 1$, then ϕ_1 and ϕ_2 will be homotopic to each other.

However, we saw above that $\pi_2(G) = 0$ for any compact connected Lie group G. Therefore any map from the two-sphere to G, including $g_3(s,t)$, is homotopic to the identity map, and hence any two maps $\phi_1(s,t)$ and $\phi_2(s,t)$ that yield the same $h(s)$ are homotopic to each other. This clearly extends to the case where ϕ_1 and ϕ_2 yield different, but homotopic, loops $h_1(s)$ and $h_2(s)$. Thus, we have shown that there is a one-to-one mapping of elements of $\pi_2(G/H)$ to elements of $\pi_1(H)$.

Does this mapping yield all elements of $\pi_1(H)$? To answer this question, consider the series of loops $g_t(s) \equiv g(s,t)$. We have $g_0(s) = 1$ and $g_1(s) = h(s)$. Because $g(s,t)$ is a smooth function, this gives a homotopy connecting the loops $g_1(s)$ and the trivial loop $g_0(s)$. Hence, $h(s)$, *viewed as a loop in* G, is homotopic to the identity. Conversely, if we are given any $h(s)$ that is homotopic, as a loop in G, to the identity map, that homotopy defines a $g(s,t)$ and a $\phi(s,t)$. Thus, the mapping from $\pi_2(G/H)$ is onto the subgroup of $\pi_1(H)$ that is mapped onto the identity element of $\pi_1(G)$. If $\pi_1(G) = 0$, i.e., if G is simply connected, then all of $\pi_1(H)$ is mapped onto the single element of $\pi_1(G)$. The mapping from $\pi_2(G/H)$ to $\pi_1(H)$ is then both one-to-one and onto, so

$$\pi_2(G/H) = \pi_1(H), \quad G \text{ compact, connected, and simply connected.} \tag{4.51}$$

In the application of this result to the search for solitons in gauge theories G is always compact and connected, so it is only the last condition that is nontrivial. However, this too is always satisfied if G is taken to be the covering group of the Lie algebra, with H then being the appropriate subgroup of this covering group.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 23:09:45, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9781139017787.005