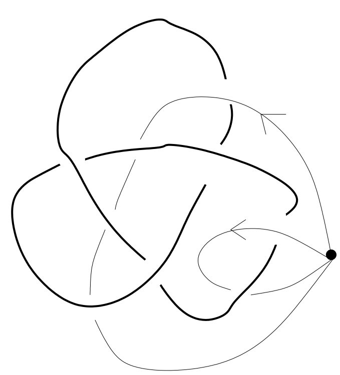
2 Basics II

2.1 Historical background

The mathematical study of knots was started in the late 19th century, but for a long time it remained very much a small backwater. The first tables of knots of up to ten crossings were produced by the Scottish physicist Peter Guthrie Tait and the American C. N. Little. Their work was really of an experimental nature. The method was to somehow construct an exhaustive list of all the possible knot diagrams with the required number of crossings, and then to reduce this list by finding equivalences between its members by *ad hoc* procedures. Needless to say, this work was slow and error prone. In Tait's case, much of the hard work of enumerating the diagrams to be considered had been carried out by an amateur mathematician, the clergyman Thomas P. Kirkman, who is now considered to be one of the founders of the branch of mathematics known as combinatorics.

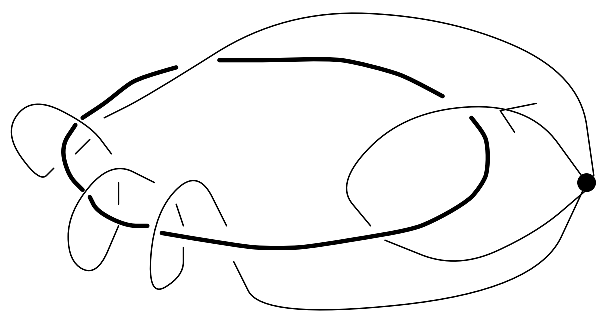
Up to the 1920's, there were few attempts to prove that the knots listed as distinct in the tables of Tait and Little really were inequivalent knots. The reason is that the available knot invariants were either difficult to calculate or not powerful enough. Some invariants, such as the crossing number, suffered from both defects.

Around 1900, Poincaré introduced an important topological invariant, which is now known as the *fundamental group* of a topological space.



The idea was to consider the equivalence classes of loops in the space under the equivalence relation known as homotopy. Here a loop simply means a continuous mapping from a circle into the space under consideration. In order to apply this to knot theory, the fundamental group of the *complement* of a knot K in $\mathbf{R}^3$ is considered. That is, we look at closed curves in $\mathbf{R}^3$ which do not intersect K, and we regard two such curves as equivalent if one of them can be continuously deformed into the other, of course without intersecting K in the process.

The set of all equivalence classes of such loops forms a *group* called the *group of the knot K*. This group is non-commutative in general, although in the simplest case, where K is the unknot, the group is isomorphic to the additive group of integers $\mathbf{Z}$. You can see this intuitively by taking K to be a circle. The integer associated with a loop which does not intersect the circle is the number of times that the loop is threaded through the circle K. The loop is taken as having an orientation, so that this number is taken with a positive or negative sign as appropriate. We shall study this idea in a slightly different form when we study *linking numbers* later in this chapter.



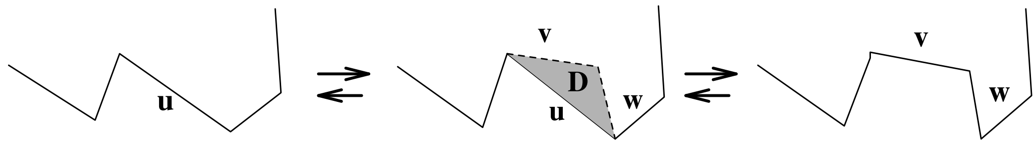
The group of the knot was recognised from the early days of the 20th century as a very powerful tool for distinguishing between different knots. However, as a practical means for distinguishing between the knots in the tables it had a serious shortcoming: the task of proving that two knot groups were different (that is, not isomorphic to one another) turned out to be just as difficult as the original task of proving that the knots were not equivalent.

Thus the knot group remained for some time a theoretical rather than a practical tool for distinguishing between different knots. All the same, Poincaré's work had an important influence on the way the problem came to be seen, because it showed that a difficult geometrical problem could be attacked by a purely algebraic method.

The next step, that of extracting a computable invariant from the knot group that still retained much of the full power of the knot group itself as an invariant, was taken by the American mathematician J. W. Alexander in the 1920's. He developed a polynomial invariant for knots, which turned out to be highly successful in confirming the accuracy of the tables by distinguishing between the knots listed in a mathematically rigorous way for the first time. This was really the beginning of knot theory as we know it today. In Chapter 3, we shall study the Alexander polynomial, from a direct combinatorial point of view that does not require the study of the knot group.

Alexander's work was complemented by that of the German mathematician Kurt Reidemeister, who developed a combinatorial approach to the subject. Taking knots simply as polygons in $\mathbf{R}^3$, Reidemeister introduced a way of doing knot equivalence by sequences of triangle moves in $\mathbf{R}^3$, which he called Δ-*moves*.

Let K be a polygonal knot or link in $\mathbf{R}^3$, and let D be a triangle in $\mathbf{R}^3$ with sides u, v, w such that u is one of the line segments which make up K and $D \cap K = u$. Then a new polygonal link K' is obtained by deleting u from K and inserting $v \cup w$ instead.



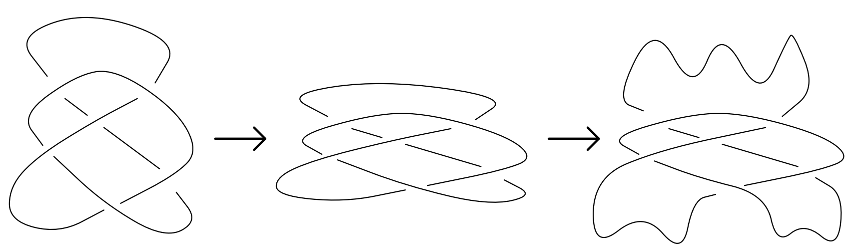
Clearly K' is equivalent to K; we say K' results from K by a Δ-move, and conversely K results from K' by a Δ^{-1}-move. The triangle D is allowed to be just a straight line segment with an intermediate point, so that inserting a new vertex counts as a Δ-move. Reidemeister proved that if two polygonal knots K_1 and K_2 are equivalent (in the sense that there is an orientation preserving homeomorphism of $\mathbf{R}^3$ which carries K_1 to K_2), then there is a finite sequence of Δ and Δ^{-1} moves leading from K_1 to K_2.

This result was the cornerstone of Reidemeister's combinatorial approach to knot theory. In this approach, not only is a knot determined by a finite amount of data, but also equivalences of knots are similarly determined by a finite amount of data. Using this approach, Reidemeister wrote the first ever book on knot theory in 1932. This classic work of only 74 pages was a summary of the state of knowledge of the subject.

2.2 Reidemeister Moves

Reidemeister's theorem implies that in order to decide whether two knot or link diagrams represent equivalent knots or links, it is sufficient to study the *projections* of the Δ and Δ^{-1} moves in $\mathbf{R}^2$.

In some cases the diagrams themselves are *planar isotopic*. This means that one diagram can be deformed continously into the other one in the plane itself. Such deformations do not alter the topology of the diagram. For example, we cannot change the number of crossings. All that can change is the size and shape of the diagram, as suggested by the example shown below.



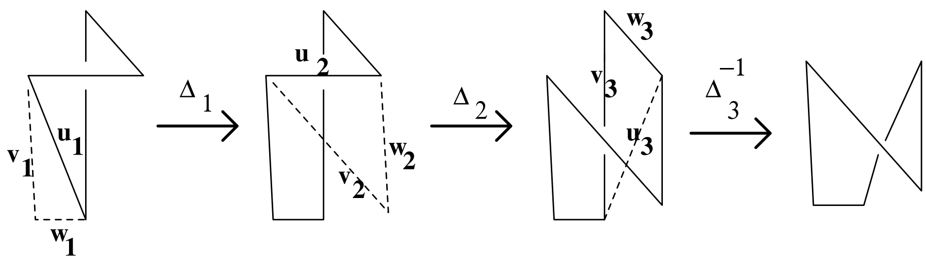
We look at this simple case first from the point of view of Δ-moves.

Projections of Δ-moves can give planar isotopies of knot diagrams. This happens in the following cases.

(1) The Δ move uses a triangle D which is itself in $\mathbf{R}^2$, *i.e.* the triangle D has one side u a segment of the knot diagram, not containing a crossing point, and the intersection of D with the knot diagram is u.

(2) Similar moves where u contains a crossing point, and one of the other two sides v of D intersects the crossing segment at a point which is not a vertex of the knot diagram or of D. In the new diagram, v is an overpass if u was an overpass.

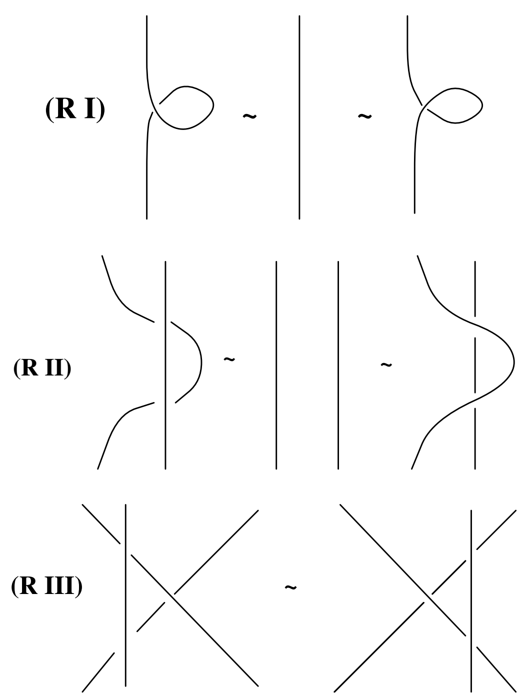
Examples of moves of these types are shown below.



You are encouraged to experiment a little to understand the effect of planar isotopies. Remember that they allow you to do "obvious" deformations of diagrams, but don't allow you to remove (or add) crossings.

The more interesting equivalences of diagrams which can be done using the Δ moves are called *Reidemeister moves.*

Definition 2.1 *The three Reidemeister moves.*



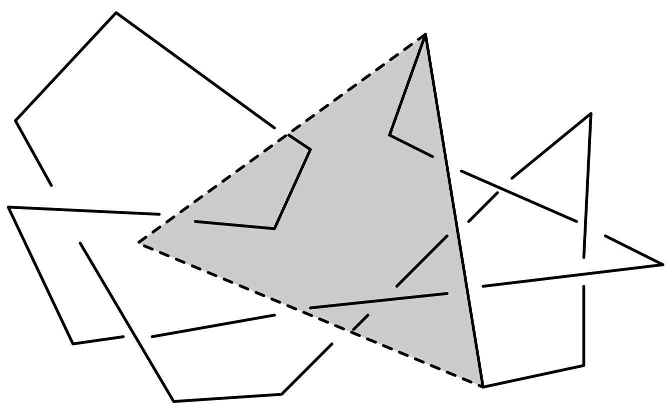
Note that it is part of the definition of the Reidemeister moves that the parts of the diagrams not shown are *identical* before and after the move. Moves I and II require two diagrams, but move III needs only one, as it makes no difference whether we think of the top string moving over or the bottom string under (or indeed the middle string through) the crossing of the other two.

It is intuitively obvious that if two knot diagrams are connected by a sequence of moves of type RI, RII and RIII, then they represent equivalent knots. It is also easy to see how each move can be obtained by projecting down Δ-moves (you can use planar isotopies as well to see this for case RIII).

Theorem 2.2 (Reidemeister) *Two knot diagrams K and K' represent equivalent knots if and only if K can be transformed into K' by a finite sequence of Reidemeister moves and planar isotopies.*

This theorem is a consequence of the theorem about Δ-moves in $\mathbf{R}^3$. We shall not prove it, but use it as the starting point of our work. *In other words, we shall simply regard Reidemeister's theorem as our working definition of equivalence of knots and links.*

The next diagram suggests how to break down a Δ-move into a number of Reidemeister moves.

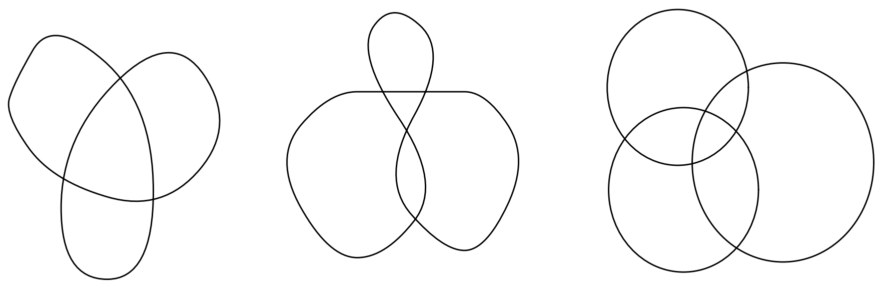


There is an important geometrical difference between the RI move and the other two. Moves RII and RIII can be carried out by a natural extension of planar isotopy — we can imagine the segments in the plane sliding over each other to give the equivalence. However, for RI it is necessary to pick up and untwist the loop using the third dimension.

Definition 2.3 *Knot diagrams K and K' are* regularly isotopic *if K can be transformed into K' by planar isotopies and Reidemeister moves RII and RIII only.*

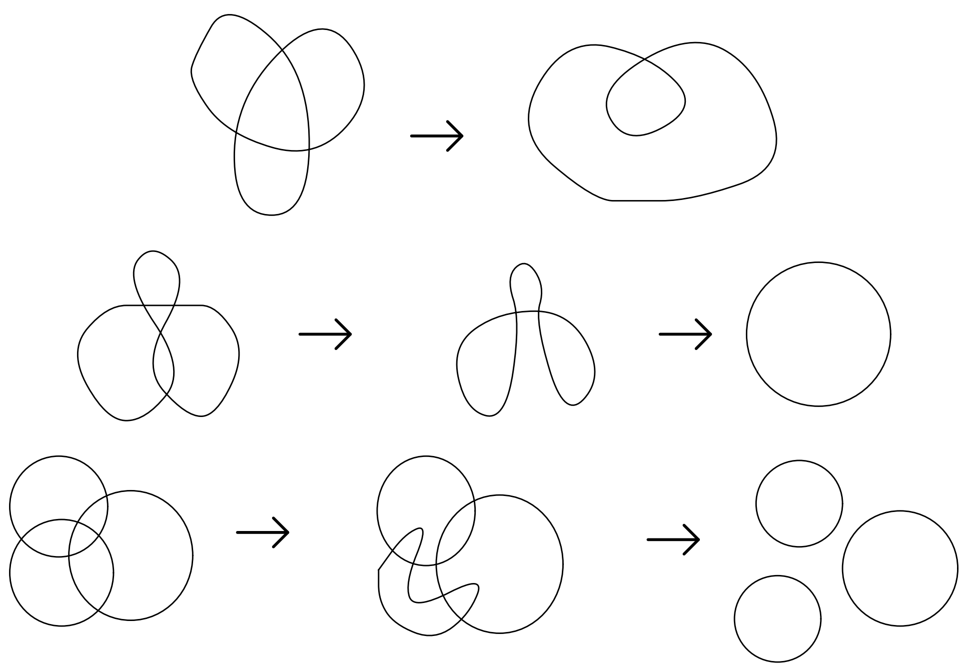
This definition will be useful much later in the course, when we come to study the Jones polynomial for knots and links.

It is worth mentioning that there is an analogous equivalence relation for diagrams made up of a finite number of closed differentiable curves in $\mathbf{R}^2$. These diagrams are the actual projections of differentiable links in $\mathbf{R}^3$, where simple intersections replace the crossings. Later we shall refer to such diagrams as *link universes.*

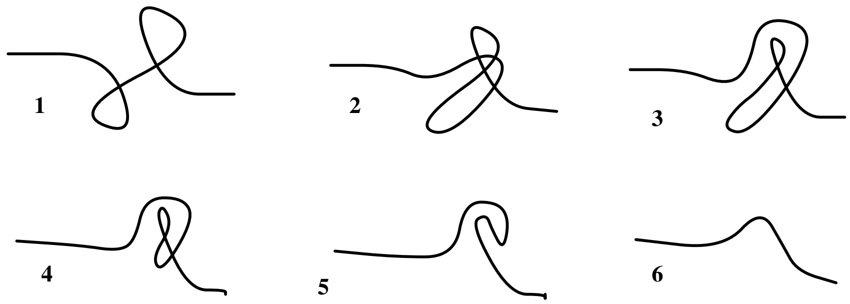


trefoil universe **figure eight knot universe** **Borromean rings universe**

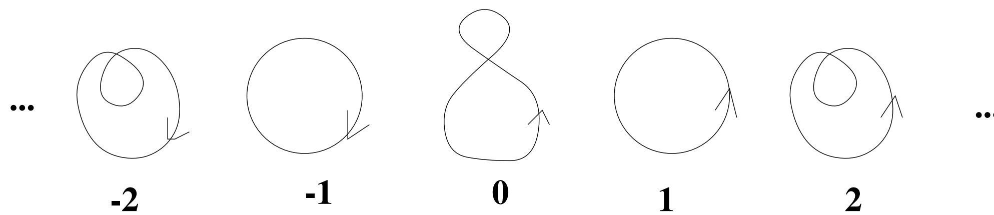
Two link universes are regularly isotopic if one can be deformed into the other through a family of diagrams of the same kind. The examples below show examples of regular isotopies. These include "moves" analogous to the RII and RIII moves. However, there is no analogue of the RI move, as an attempt to remove a loop by "pulling it tight" results in a curve with a cusp, *i.e.* a point where the curve is not differentiable.



The next example of a regular isotopy shows that loops with opposite orientations cancel each other out under this equivalence relation.



This is known as the "Whitney trick". Using the Whitney trick, it can be shown that every oriented knot universe is regularly isotopic to a circle with a number of positive or negative loops, *i.e.* to one of the curves in the doubly infinite sequence shown below.



2.3 The Colour Invariant

In this section we discuss a simple knot invariant, which can be used to prove that the trefoil knot is truly knotted, *i.e.* not equivalent to the unknot. The idea is to consider the knot diagram as it is drawn, as a set of curves (or polygons) in the plane with breaks at the underpasses. Each of these unbroken sections of the knot, running from one underpass to the next, will be called an *arc* of the diagram.

We now colour the arcs, if possible, using 3 colours, so that the following conditions are satisfied.

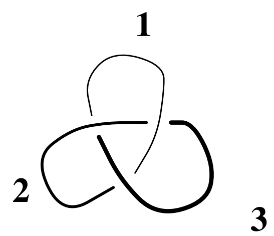
(i) At least two colours are used.

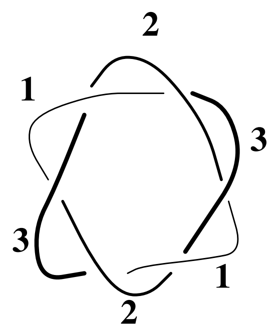
(ii) The three arcs which meet at a crossing either have all the same colour or have three different colours.

Of course condition (ii) is a local condition at each crossing. More accurately we should say "subarcs" or something like that, to allow for the possibility that the three parts of arcs which appear locally may actually be parts of fewer than 3 arcs in the whole knot diagram.

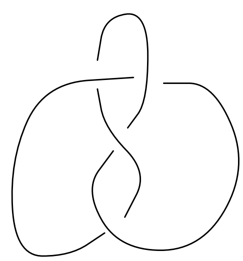
These conditions seem very arbitrary and artificial. However, they will provide us with a useful exercise in the use of the Reidemeister moves to establish that some feature of a knot or link diagram is actually an invariant of the underlying knot or link itself.

Here are some examples of 3-colourings.

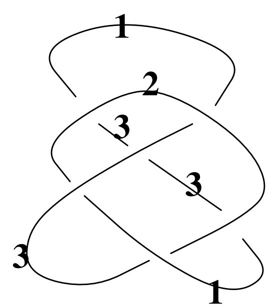




It is not hard to see by trial and error that the standard diagram for the figure eight knot



is not 3-colourable. Remember that we have to allow the possibility that the 3 arcs which meet at certain crossings all have the same colour. For example, the following is a valid 3-colouring of a diagram from Examples 1.



Of course, none of this has any interest unless we can show that the 3-colouring property is a property of the underlying knot itself, and not just an accidental feature of the choice of diagram. In other words, we have to prove that the property is a *knot invariant*. (Unlike numerical invariants, this invariant has only two possible "values". Either the diagram can be 3-coloured or it can not. So we are really aiming to divide all knots and links into just two big equivalence classes by means of this invariant.)

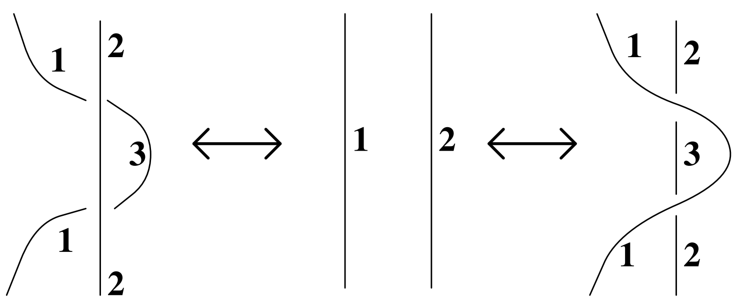
To prove the invariance, we shall show for each Reidemeister move in turn that if the diagram can be 3-coloured before the move takes place, then it can also be 3-coloured after the move. Of course, the colours assigned to the arcs can be changed. Since every knot equivalence can be carried out by a finite sequence of Reidemeister moves and planar isotopies (which obviously don't affect the colouring property, as we just follow each arc through the isotopy), this will prove invariance.

Proposition 2.4 *The three Reidemeister moves preserve the 3-colouring property.*

Proof (**RI move**) Note that only two arcs are involved at the crossing created or removed by a RI move. If we introduce a new crossing by an RI move, we simply leave both arcs the same colour as the original one. Likewise, removing a crossing by an RI move is no problem, since the two arcs must already be the same colour.

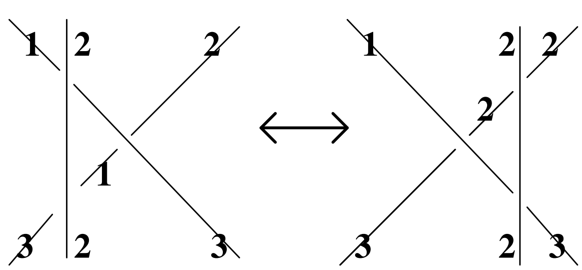
(**RII move**) There is a trivial case, similar to the case of an RI move, where all the arcs involved in the RII move have the same colour before and after the move. (Notice that the requirement that all 3 colours are used refers to the diagram *as a whole*, not to the part of the diagram involved in a local move. However, we must be careful not to "lose" colours from a part of the diagram in which all the arcs connecting it to the rest of the diagram have the same colour.)

The more interesting case is illustrated below.



The idea is to colour the new arc on the underpass with the third colour.

(**RIII move**) Aside from the trivial case (all one colour) there are two basic possibilities to be considered. Either all three colours occur at all three crossings in the RIII diagram, or there is exactly one crossing where only one colour occurs. These cases are interchanged by the RIII move.



Notice that only one arc changes colour. None of the others can change, because they lead to parts of the diagram that we have not drawn. You should check carefully that all possibilities have been accounted for.

With this result, we have achieved a major goal. Since the trefoil knots (left or right) are 3-colourable, and the unknot is *not* 3-colourable (why?), the trefoil knot(s) are not equivalent to the unknot. Or, as anyone except a knot theorist would say, they are really knotted! However, as we have seen, we can't distinguish between the figure eight knot and the unknot by this method.

We can also use this invariant for links with more than one component. However, notice that a split unlink with more than one component *can* be 3-coloured. However, the Hopf link obviously can't be 3-coloured (why not?). So this gives us our first proof that the Hopf link really is linked! Later in the course, we shall find more proofs of this fact.

Remark 2.5 You need to be really careful in using the 3-colouring property. For example, the "star of David" link above can *not* be distinguished from the unlink by this invariant, precisely because it *can* be 3-coloured.

We mention briefly a generalisation of this invariant.

Definition 2.6 *Let n be an integer ≥ 3. A knot or link K can be* coloured *mod n if we can assign an integer to each arc in some diagram of K so that the following conditions are satisfied.*

(i) *At each crossing the integers a and b assigned to the (subarcs of the) underpass and the integer c assigned to the (subarc of the) overpass satisfy the linear congruence*

$$a + b \equiv 2c \bmod n.$$

(ii) *The integers assigned to the arcs are not all congruent mod n.*

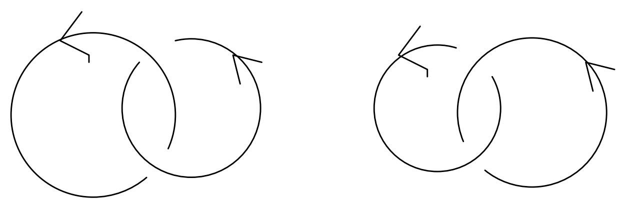
Of course, it is equivalent to use just one representative from each congruence class mod n as the "colours" to be assigned to the arcs. The case $n = 3$ is the one we have discussed. The proof that colourability mod n is a link invariant is left to you as an exercise.

Using this general family of invariants, we can distinguish many more knots and links. For example, the figure eight knot can be coloured mod 5 (see Examples 2). Since the unknot cannot be coloured mod n for any value of n (why?), this gives a method to distinguish between the figure eight knot and the unknot.

2.4 Linking Numbers

In Chapter 1 we defined *oriented* knots and links. Recall that this simply means that each component has an arrow on it to indicate the positive sense. Obviously planar isotopies and the three Reidemeister moves can be considered as transformations of oriented knot diagrams. Reidemeister's theorem tells us that two oriented knots or links are equivalent if and only if one can be changed into the other by a finite sequence of Reidemeister moves and planar isotopies, where all such moves and isotopies now have to preserve the orientations.

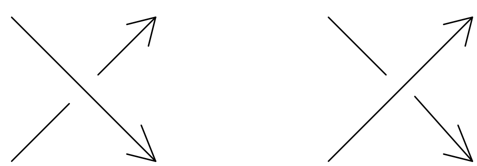
To appreciate the difference that orientation makes, let's go back to our rope models. There are two different "Hopf links" if we take account of orientation, but with the arrows removed we can no longer distinguish between these. We observed this already in Section 1.2.



In this section we shall introduce an invariant of oriented 2-component links, called the *linking number*. This invariant will be of fundamental importance in later parts of the course.

We begin by assigning a *sign* to each of the crossings in an oriented knot or link diagram,

Definition 2.7 *At a crossing of two oriented line segments in* $\mathbf{R}^2$*, suppose that the overpass is rotated through an angle* $< \pi$ *to align it with the underpass. Then the crossing is* positive *if the rotation is anticlockwise and* negative *if the rotation is clockwise.*



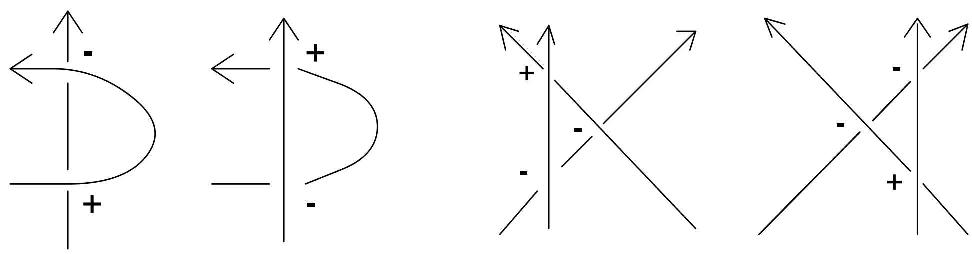
ε = +1 **ε = -1**

Using this idea we can define the linking number, as follows. Denote the two components by α and β. For each crossing point x of α with β, let $\epsilon(x) = +1$ if x is a positive crossing and let $\epsilon(x) = -1$ if x is a negative crossing.

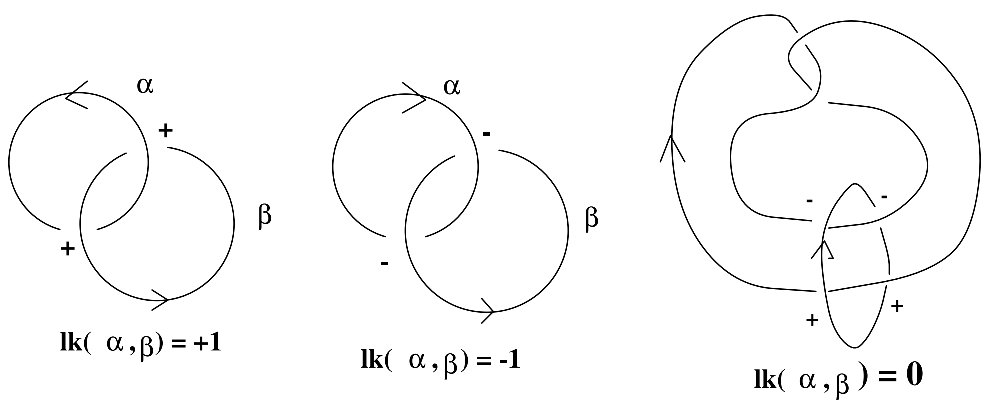
Definition 2.8 *The* linking number *of* α *and* β *is* $\mathrm{Lk}(\alpha, \beta) = \frac{1}{2}\sum_x \epsilon(x)$.

Proposition 2.9 *The linking number is an invariant of oriented link type.*

Proof Clearly planar isotopies do not change anything. We consider the effects of the three Reidemeister moves in turn. RI moves also do not contribute to the linking number, since they affect only self-crossings of α or β, and these are not counted in the sum. RII moves add (or subtract) both a $+1$ and a -1. And RIII moves do not alter the summation, because the relative position of any two of the three lines remains the same.



Finally note that since there must be an even number of crossing points altogether, the linking number is an integer. (A proof of this will be given in Proposition 2.10.)



The third example here is known as the *Whitehead link*. It is an example of a non-trivial link with linking number 0. This example shows that it is not possible to tell whether a link diagram with two components represents an unlink or not, simply by calculating the linking number. However, if the linking number is different from 0, we can certainly say that the link cannot be split. Of course, the linking number is of no help to us in studying the possible knottedness of the two individual components.

The Borromean rings (Section 1.1) provide another example where linking numbers fail to capture the reason why the three rings cannot be separated. Here we can calculate three linking numbers, one for each pair of components, but of course these linking numbers will be 0, since we know that if one link is removed, then the other two form a split unlink.

If two different components of a link cross at x, then we can change the sign $\epsilon(x)$ by reversing the orientation on *one* of the components. If the orientation on both components is reversed, then $\epsilon(x)$ is unchanged. In the case where x is a self-crossing of a single component, it follows that $\epsilon(x)$ is independent of the choice of orientation of this component. Such self-crossings are omitted in calculating linking numbers.

Proposition 2.10 *Let L be a 2-component link with components α and β. Suppose that L can be split by switching the overpass and underpass at some subset of the crossings of α with β, say $c_1, c_2, \ldots, c_n$. Then the linking number of L is the sum of the signs of these crossings, that is*

$$\mathrm{Lk}(\alpha, \beta) = \sum_{i=1}^{n} \epsilon(c_i).$$

Proof For each k with $1 \leq k \leq n$, let L_k be the link obtained from L by switching the crossings c_i, $1 \leq i \leq k$. Also let $L = L_0$. Thus L_i is obtained from L_{i-1} by switching the ith crossing c_i. If this was a positive crossing, then c_i contributes $1/2$ to $\mathrm{Lk}(L_{i-1})$ and contributes $-1/2$ to $\mathrm{Lk}(L_i)$. Similarly, if c_i was a negative crossing, then c_i contributes $-1/2$ to $\mathrm{Lk}(L_{i-1})$ and contributes $1/2$ to $\mathrm{Lk}(L_i)$. In either case we have

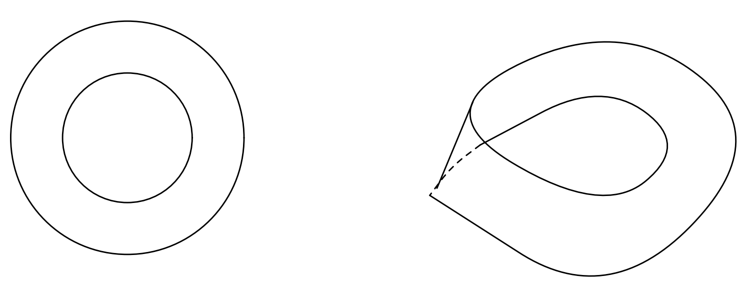
$$\mathrm{Lk}(L_{i-1}) - \mathrm{Lk}(L_i) = \epsilon(c_i).$$

Summing these equations for $1 \leq i \leq n$ gives $\mathrm{Lk}(L_0) - \mathrm{Lk}(L_n) = \sum_i \epsilon(c_i)$. But $L_0 = L$ and $\mathrm{Lk}(L_n) = 0$, since L_n is a split link. This completes the proof. $\square$

In particular, we can take $c_1, c_2, \ldots, c_n$ to be the set of crossings where α goes over β, or vice versa.

2.5 Linking, twisting and writhing

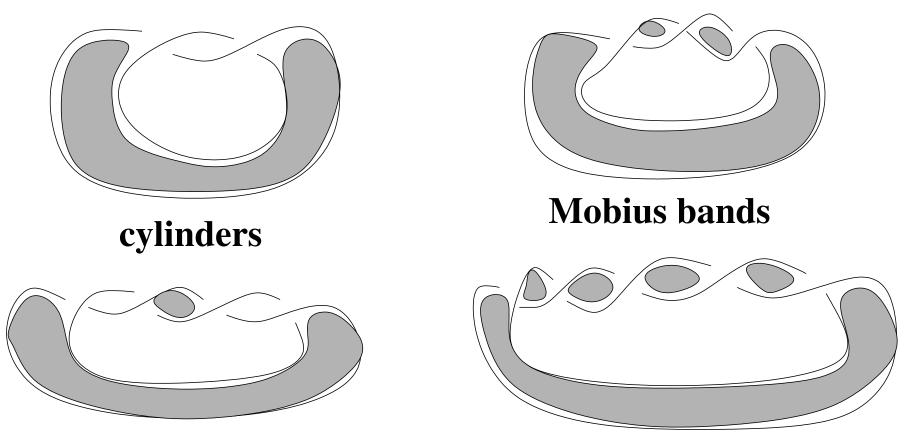
So far we have been thinking of knots and links physically as closed loops of rope, and mathematically as closed curves in $\mathbf{R}^3$. There is another physical model which has an interesting relation with linking numbers. Suppose that we have a ribbon in $\mathbf{R}^3$, again in the form of a closed loop. The ribbon can have two different topological forms, a *cylinder* and a *Möbius band*.



The topological distinction between these surfaces can be seen in a number of ways. For example the cylinder has two "sides", in the sense that an imaginary insect crawling on the surface has to cross an edge in order to get from one side to the other. The same insect crawling on the Möbius band can reach the "other" side by walking once round the band. For this reason, the Möbius band is often called a "one-sided" surface, though topologists generally prefer to call it a "non-orientable" surface. Here we do not attempt a proper mathematical treatment of these matters; this is done in the course MT3121 (Topology of Surfaces). For our purposes, it will be enough to observe that we could colour one side of the cylinder red and the other green, so that the two colours meet only at the edges of the cylinder, but that a similar attempt to paint a Möbius band will result in one colour only being used.

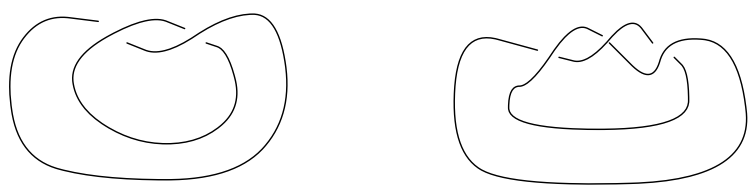
Another way to distinguish between the cylinder and the Möbius band is to look at their boundaries. The cylinder has two separate edges, while the Möbius band has only one edge: topologically speaking, the Möbius band has a circle as its boundary, just as a disc has. Of course, the "one-sidedness" property also distinguishes the Möbius band from the disc.

What about a ribbon with two or more twists? It turns out that every even number of twists gives a surface topologically equivalent to the cylinder, and every odd number of twists gives a surface topologically equivalent to the Möbius band.

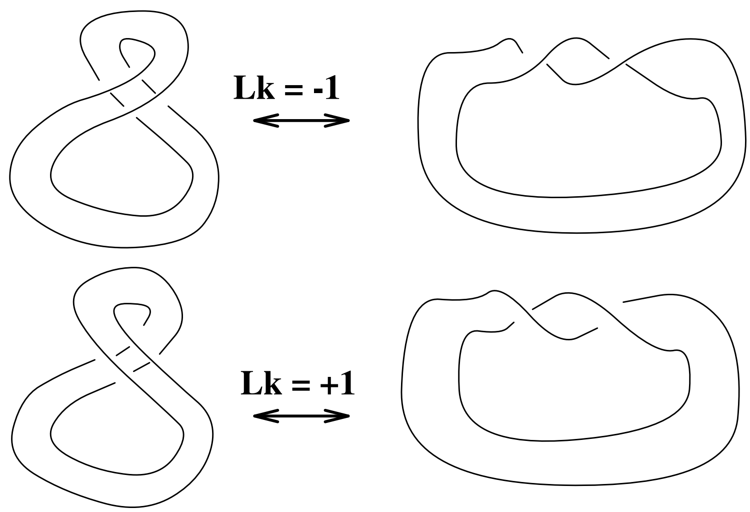


To see this, cut the ribbon, undo an even number of twists, and join it up again as it originally was. The surface itself is not changed by this drastic procedure. What has changed is the "embedding" of the surface in $\mathbf{R}^3$. That is to say, we cannot undo the even number of twists by manipulating the ribbon in space, in the same way that we manipulate a knot.

We can prove this last statement by using our knowledge of knot theory. The two curves which form the boundary of a ribbon with two twists form a Hopf link, and so have a linking number which is $+1$ or -1, depending on the choice of orientations. If we could untwist the ribbon in $\mathbf{R}^3$, then by so doing we should have unlinked this link. And by the invariance of linking numbers, we know that this is impossible. Similarly, the one curve which forms the boundary of a ribbon with 3 twists is easily seen to be a trefoil knot. Since the boundary of a standard Mobius band is an unknot, the 3 twist ribbon can not be manipulated into a standard Möbius band in $\mathbf{R}^3$. We say that it is a different embedding of the Möbius band in $\mathbf{R}^3$ from the usual one, just as the trefoil knot is a different embedding of a topological circle in $\mathbf{R}^3$ from the usual one.



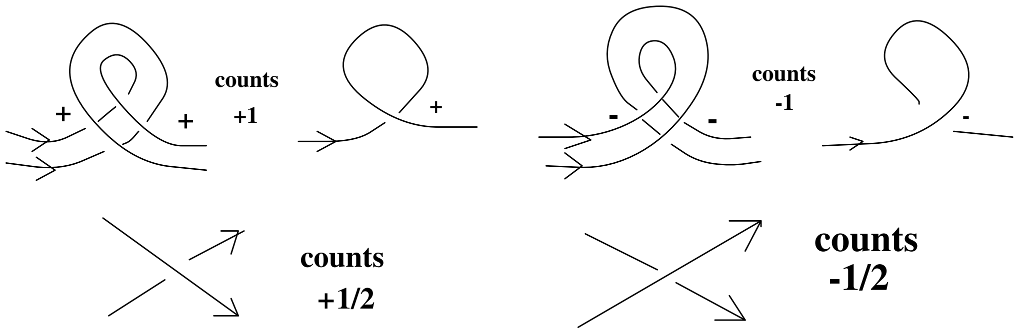
The next diagrams show more embeddings of ribbons in $\mathbf{R}^3$. It is easy to see that these are cylinders, not Möbius bands, by painting them red on one side and green on the other. Hence the ribbons each have two boundary curves, which form a 2-component link L. Physical experiment should convince you that the ribbon with the curl and the ribbon with two twists are equivalent.



We note that the linking numbers agree, if we orient the boundary curves the same way.

The linking numbers also tell us that the direction of the twists is physically significant. Let us agree to give the same orientation to both boundary curves of a ribbon. Then the linking number of the two boundary curves is an invariant of the embedding of the ribbon. This is because an equivalent embedding of the ribbon would have an equivalent link as its boundary.

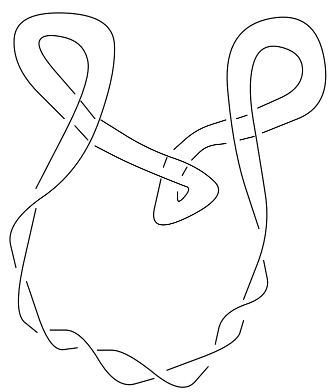
These examples reveal an interesting geometric phenomenon: we can exchange *twists* in an embedded ribbon for *coils*. The linking number $\mathrm{Lk}(L)$ of the boundary link L can be calculated from the total number of twists and coils.



If we denote by $\mathrm{Tw}(L)$ the sum of twist contributions, and by $\mathrm{Wr}(L)$ the sum of the curl (or *writhe*) contributions, we get the equation

$$\mathrm{Lk}(L) = \mathrm{Tw}(L) + \mathrm{Wr}(L).$$

Try this out on the following example.



Of course, taken separately, neither $\mathrm{Tw}(L)$ nor $\mathrm{Wr}(L)$ is an invariant of the link L. However, they can be defined in terms of the differential geometry of the two component curves of L. (You may know something about the differential geometry of curves in $\mathbf{R}^3$ from MT3151.)

This formula for the linking number has been used by biologists to study certain forms of DNA. These molecules are in closed, linked, double-stranded form, somewhat like the links drawn above. The coiling and twisting in the DNA can be observed in a direct way using the electron microscope. Different energy levels in the molecules are associated with different amounts of coiling and twisting, higher energy levels being associated with more and more coiling until a "supercoiled" state is reached in which the small coils are themselves grouped in larger coils. (You can see this effect by twisting a length of telephone cable.) We shall return to applications of knot theory to DNA in Chapter 5.