CHAPTER 1: A CENTURY OF KNOT THEORY

In 1877 P. G. Tait published the first in a series of papers addressing the enumeration of knots. Lord Kelvin's theory of the atom stated that chemical properties of elements were related to knotting that occurs between atoms, implying that insights into chemistry would be gained with an understanding of knots. This motivated Tait to begin to assemble a list of all knots that could be drawn with a small number of crossings. Initially the project focused on knots of 5 or 6 crossings, but by 1900 his work, along with that of C. N. Little, had almost completed the enumeration of 10-crossing knots. The diagrams in Appendix 1 indicate the kind of enumeration he was seeking.

Tait viewed two knots as equivalent, or of the same type, if one could be deformed to appear as the other, and sought an enumeration that included each knot type only once. The difficulty of this task is illustrated by the four knots in Figure 1.1. For now a knot can be thought of simply as a loop of rope. With some effort it is possible to deform the second knot to appear untangled, like the first. On the other hand, no amount of effort seems sufficient to unknot the third or fourth. Is it possible that with some clever manipulation the third could be transformed to look like the fourth? If a list of knots is going to avoid knots of the same type appearing repeatedly, means of addressing such questions are needed.

When Tait began his work in the subject, the formal mathematics needed to address the study was unavailable. The arguments that his lists were complete are convincing, but the evidence that the listed knots are distinct was empirical. Developing means of proving that knots are distinct remains the most significant of the many problems introduced by Tait.

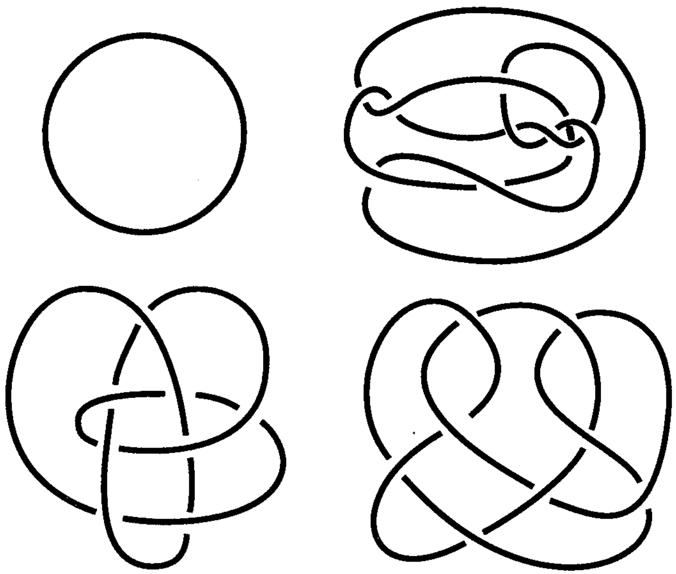


Figure 1.1

Work at the turn of the century placed the subject of topology on firm mathematical ground, and it became pos-

sible to define the objects of knot theory precisely, and to prove theorems about them. In particular, algebraic methods were introduced into the subject, and these provided the means to prove that knots were actually distinct. The greatest success in this early period was the proof by M. Dehn in 1914 that the two simplest looking knots, the right- and left-handed trefoils, illustrated in Figure 1.2, represent distinct knot types; that is, there is no way to deform one to look like the other.

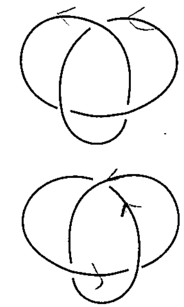


Figure 1.2

In 1928 J. Alexander described a method of associating to each knot a polynomial, now called the Alexander polynomial, such that if one knot can be deformed into another, both will have the same associated polynomial. This invariant immediately proved to be an especially powerful tool in the subject; a scan of Appendix 2 reveals that only 8 knots out of the 87 with 9 or fewer crossings share polynomials with others on the list.

Alexander's initial definitions and arguments were combinatorial, depending only on a study of the diagram of a knot, without reference to the algebra that had already proved so successful.

By 1932 the subject of knot theory was fairly well developed, and in that year K. Reidemeister published the first book about knots, *Knotentheorie*. The tools that he presented in the text are, in theory, sufficient to distinguish almost any pair of distinct knots, although as a practical matter for knots with complicated diagrams the calculations are often too lengthy to be of use.

One theme that was well established by this time was the study of families of knots. The most interesting family is formed by the torus knots, so called because they can be drawn to lie on the surface of a torus.

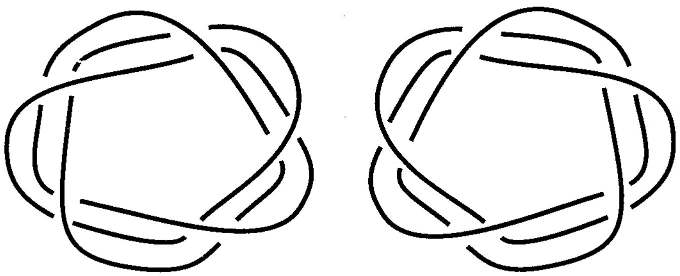


Figure 1.3

For any ordered pair of relatively prime integers, (p,q), with $p > 1$ and $|q| > 1$, there is a corresponding (p,q)-torus knot. Figure 1.3 illustrates the (3,5)-torus knot and the (3,−5)-torus knot. The right- and left-handed trefoils are easily seen to be the same as the (2,3) and (2,−3)-torus knots, respectively. These knots provide test cases for new techniques and building blocks for constructing more complicated examples. Dehn and O. Schreier used group theoretic methods to give the first proof that the (p,q) and (p',q')-torus knots are the same if and only if the (unordered) sets $\{p,q\}$ and $\{p',q'\}$ are the same. (The Alexander polynomial of the (p,q)-torus knot turns out to be $(t^{|pq|}-1)(t-1)/(t^{|p|}-1)(t^{|q|}-1)$, and except for an issue of sign, this too is sufficient to distinguish the torus knots.)

Soon after *Knotentheorie* appeared, H. Seifert made a significant discovery. He demonstrated that if a knot is the

boundary of a surface in 3-space, then that surface can be used to study the knot; he also presented an algorithm to construct a surface bounded by any given knot. Figure 1.4 illustrates a surface with knotted boundary. This approach was certainly of practical importance, as it gave efficient means for computing many of the known invariants. More important, it laid the foundation for the use of geometric methods into a subject that, until then, had been dominated by combinatorics and algebra.

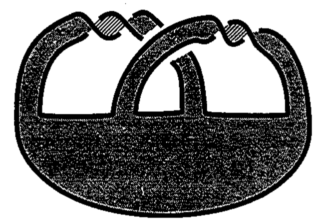


Figure 1.4

In 1947 H. Schubert used geometric methods to prove a key result concerning the decomposition of knots. Given any two knots, one can form their connected sum, denoted $K\#J$, as illustrated in Figure 1.5. (If knots are thought of as being tied in a piece of string, the connected sum of two knots is formed by tieing them in separate portions of the string so that they do not overlap.) A knot is called prime if it cannot be decomposed as a connected sum of nontrivial knots. (The appendix illustrates those prime knots with 9 crossings or less.) Schubert proved that any knot can be decomposed uniquely as the connected sum of prime knots. As an immediate corollary, if K is nontrivial, there is no knot J so that $J\#K$ is unknotted.

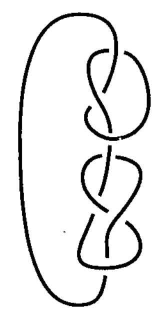


Figure 1.5

Unlike the problem of distinguishing knots, the problem of developing general means for proving that one knot *can* be deformed into another remained untouched. That changed in 1957. Early in the century Dehn gave an incorrect proof of what has become known as the Dehn Lemma. In rough terms, it stated that if a knot were indistinguishable from the trivial knot using algebraic methods, then the knot was in fact trivial. In 1957, C. Papakyriakopoulos succeeded in proving the Dehn Lemma, and it soon became the centerpiece of a series of major developments in the subject. One of special note occurred in 1968, when F. Waldhausen proved that two knots are equivalent if and only if certain algebraic data associated to the knots are the same. The interplay between algebra and geometry was essential to this work, and the connection was provided by Dehn's lemma.

The late 1950's through the 1970's were also marked by an extensive study of the classical knot invariants, and, in particular, how properties of the knot were reflected in the invariants. For instance, K. Murasugi proved that if a knot can be drawn so that the crossings alternate from over to under, then the coefficients of the Alexander polynomial alternate in sign. Figure 1.6 illustrates a nonalternating knot diagram—see how two successive overcrossings are marked. By the Murasugi theorem, it is impossible to find an alternating diagram for this knot, as it has Alexander polynomial $2t^6 - 3t^5 + t^4 + t^3 + t^2 - 3t + 2$. Murasugi's work also detailed relationships between

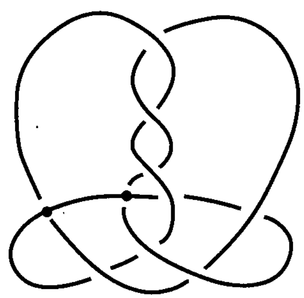


Figure 1.6

knot invariants and symmetries of knots, another major topic in the subject. Figure 1.7 illustrates three 9-crossing knots (9_4,9_{17}, and 9_{33} in the appendix.) Two of the diagrams appear quite symmetrical, while the last is striking in its asymmetry. Is it possible to deform the third knot so that it too displays a similar symmetry?

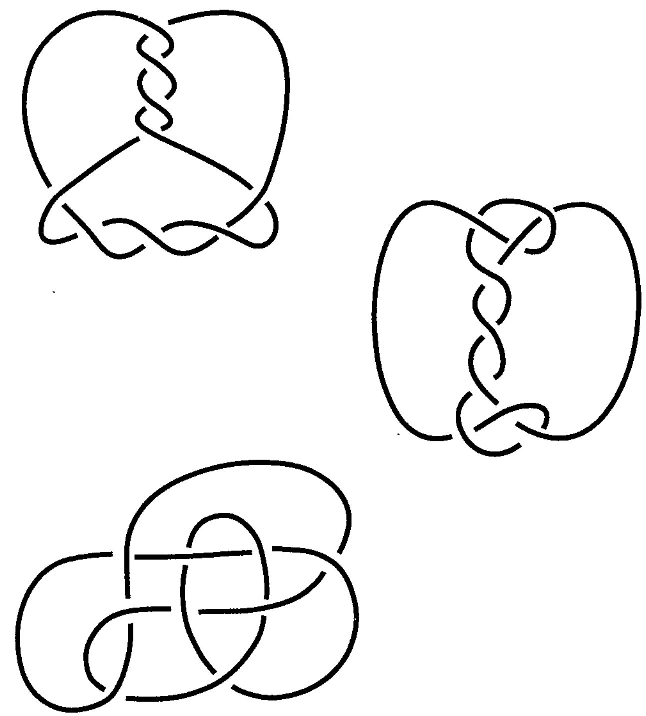


Figure 1.7

In a completely different direction, the investigation of higher dimensional knots, such as knotted 2-spheres in 4-space, became a significant topic. In 1960 the subject consisted of little more than a sparse collection of examples. By 1970 it had become a well-developed area of topology. It also had become a significant source of questions concerning classical knots.

Since 1970, knot theory has progressed at a tremendous rate. J. H. Conway introduced new combinatorial methods which, when combined with more recent work by V. Jones, have led to vast new families of invariants. New geometric methods have been introduced by W. Thurston (hyperbolic geometry) and by W. Meeks and S. T. Yau (minimal surfaces), and together these have provided significant new insights and results. Finally, in 1988 C. McA. Gordon and J. Luecke solved one of the fundamental problems in knot theory. Many of the methods of knot theory focus not on the knot itself, but on the complement of the knot in 3-space; Gordon and Luecke proved that knots with equivalent complements are themselves equivalent.

Knot theory remains a lively topic today. Many of the basic questions, some dating to Tait's first paper in the subject, remain open. At the other extreme, the results of recent years promise to provide many new insights.

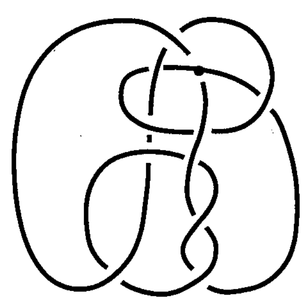


Figure 1.8

EXERCISES

1. If at a crossing point in a knot diagram the crossing is changed so that the section that appeared to go over the other instead passes under, an apparently new knot is created. Demonstrate that if the marked crossing in Figure 1.8 is changed, the resulting knot is trivial. What is the effect of changing some other crossing instead?

2. Figure 1.9 illustrates a knot in the family of 3-stranded *pretzel knots*; this particular example is the $(5,-3,7)$ pretzel knot. Can you show that the (p,q,r)-pretzel knot is equivalent to both the (q,r,p)-pretzel knot and the (p,r,q)-pretzel knot?

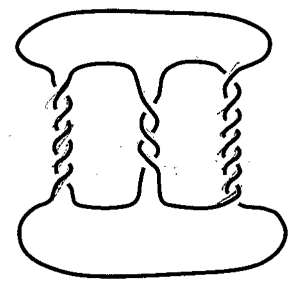


Figure 1.9

3. The subject of knot theory has grown to encompass the study of links, formed as the union of disjoint knots. Figure 1.10 illustrates what is called the *Whitehead link*. Find a deformation of the Whitehead link that interchanges the two components. (It will be proved later that no deformation can separate the two components.)

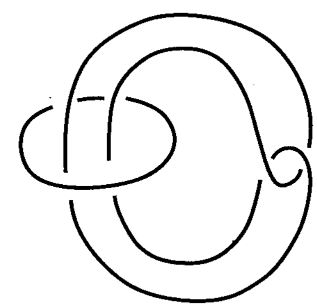


Figure 1.10

4. For what values of (p,q,r) will the corresponding pretzel knot actually be a knot, and when will it be a link? For instance, if $p = q = r = 2$, then the resulting diagram describes a simple link of three components, "chained" together.

5. Describe the general procedure for drawing the (p,q)-torus knot. What happens if p and q are not relatively prime?

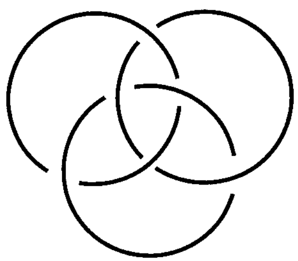


Figure 1.11

6. The link in Figure 1.11 is called the *Borromean link*. It can be proved that no deformation will separate the components. Note, however, that if one of the two components is removed, the remaining two can be split apart. Such a link is called *Brunnian*. Can you find an example of a Brunnian link with more than 3 components? (H. Brunn described families of such examples in 1892.)

7. The knots illustrated in Figure 1.12 were, until recently, assumed to be distinct, and both appeared in many knot tables. However, Perko discovered a deformation that turns one into the other. As a challenging exercise, try to find it.

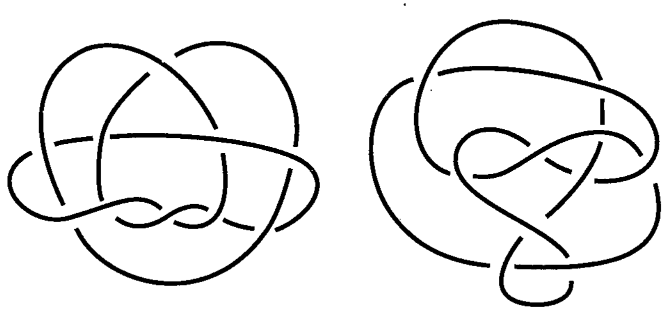


Figure 1.12

CHAPTER 2: WHAT IS A KNOT?

There are many definitions of knot, all of which capture the intuitive notion of a knotted loop of rope. For each definition there is a corresponding definition of deformation, or equivalence. This chapter will concentrate on one pair of such definitions, and mention another. (Results at the foundations of geometric topology relate the various definitions. Such matters will not be presented here, and do not affect the work that follows.) The goal for now is to demonstrate how the notion of knotting can be given a rigorous mathematical formulation, and to give the reader a flavor of the problems and techniques that occur at this basic level of the subject.

1 Wild Knots and Unknottings

Considering a pair of definitions that are not appropriate, and seeing how they fail, demonstrates some unexpected subtleties and the need for precision and care in finding the right approach. One might define a knot as a continuous simple closed curve in Euclidean 3-space, R^3. To be precise, such a curve consists of a continuous function f from the closed interval $[0,1]$ to

R^3 with $f(0) = f(1)$, and with $f(x) = f(y)$ implying one of the three possibilities:
(1) $x = y$,
(2) $x = 0$ and $y = 1$, or
(3) $x = 1$ and $y = 0$.
This is illustrated schematically in Figure 2.1.

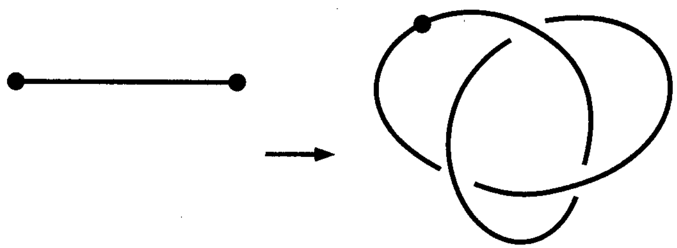


Figure 2.1

Unfortunately, with this definition the infinitely knotted loop illustrated in Figure 2.2 would be admitted into

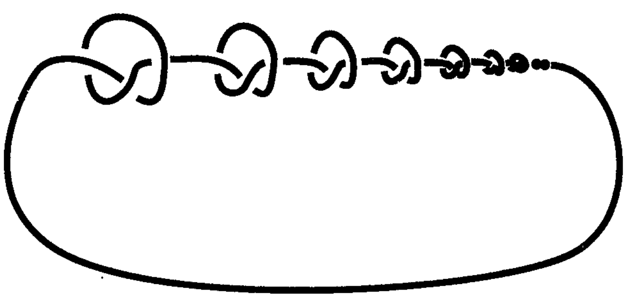


Figure 2.2

our studies. Such pathological examples are distant from the intuitive notion of a knot and the physical knotting that the theory is modelling, and so must be avoided.

Suppose for the moment that a definition similar to that indicated above were suitable. How would the idea of a deformation be captured? A natural choice would be to say that a knot J is a deformation of K if there exists a family of knots, K_t, $0 \leq t \leq 1$, with $K_0 = K$, $K_1 = J$, and with K_t close to K_s, for t close to s. Of course the idea of knots being close would have to be defined as well.

Once again, an example indicates the difficulty of using a definition based on continuity. In Figure 2.3 several steps of a deformation of a knot into an unknotted loop are illustrated. Note that at every step of the deformation the loop is a continuous simple closed curve. Somehow the definition must rule out such deformations.

One remedy is to introduce differentiability into the discussion. For instance, if the function f is required to be differentiable, with unit velocity, the possibility of a wild knot is eliminated; for the knot in Figure 2.3, the tangent is varying rapidly near the wild point where the small knots bunch up, and there is no continuous way to define a tangent direction at that wild point. Introducing differentiability into the definition of deformation is also possible, but more difficult.

An alternative solution is to use polygonal curves instead of differentiable ones. This approach avoids many technical difficulties and at the same time eliminates wild knotting, as polygonal curves are finite by nature. A theorem relating the two approaches is proved in the appendix of the text by Crowell and Fox, a good starting point for readers interested in this aspect of the theory.

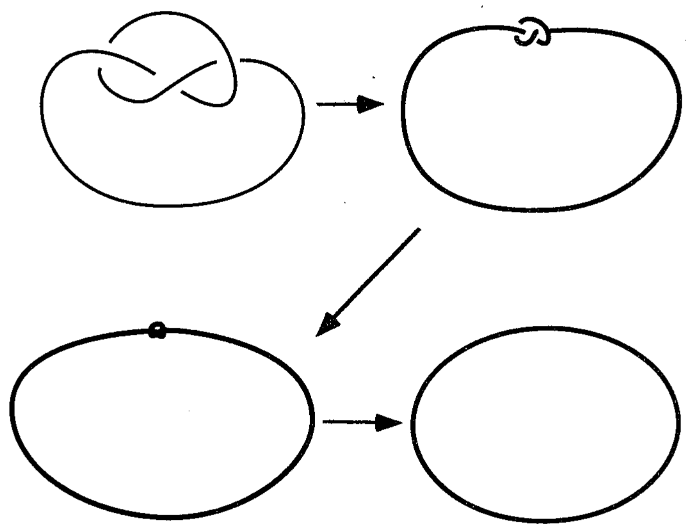


Figure 2.3

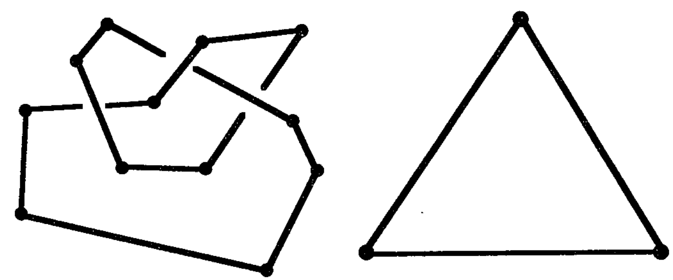
2 The Definition of a Knot The simplest definitions in knot theory are based on polygonal curves in 3-space. Essentially a knot is defined to be a simple closed curve formed by "joining the dots."

For any two distinct points in 3-space, p and q, let $[p,q]$ denote the line segment joining them. For an ordered set of *distinct points*, $(p_1,p_2,\dots,p_n)$, the union of the segments $[p_1,p_2]$, $[p_2,p_3],\dots,[p_{n-1},p_n]$, and $[p_n,p_1]$ is called a closed

polygonal curve. If each segment intersects exactly two other segments, intersecting each only at an endpoint, then the curve is said to be *simple*.

□ **DEFINITION.** *A knot is a simple closed polygonal curve in* R^3.

Figure 2.4a illustrates the simplest nontrivial knot, which is called the *trefoil*, drawn as a polygonal curve. The *unknot*, or trivial knot, is defined to be the knot determined by three noncollinear points, as illustrated in Figure 2.4b. (Note that picking a different set of three points yields a different "unknot." This ambiguity will be resolved in discussing deformations and equivalence, and in the exercises.)



(*a*) (*b*)

Figure 2.4

Knots are usually thought of, and drawn, as smooth curves and not jagged ones. An informal way of dealing with this is to view smooth knots as polygonal knots constructed from a very large number of segments. That a smooth knot can be closely approximated by a polygonal curve is intuitively clear. The formal way of dealing with this problem is to study the relationship between polygonal and differentiable knots. Knots will often be drawn smoothly in this book, but this is for aesthetic reasons, and all the figures could have been drawn polygonally instead.

There is one important observation to be made about the definition. A knot is defined to be a subset of 3-space, the union of a collection of segments. Various choices of ordered sets of points can define the same knot. For instance, cyclicly permuting the order of the points does not alter the underlying knot. Also, if three consecutive points are collinear, then eliminating the middle one does not change the underlying knot. This last observation about eliminating points along segments leads to a useful definition.

□ **DEFINITION.** *If the ordered set $(p_1,p_2,\ldots,p_n)$ defines a knot, and no proper ordered subset defines the same knot, the elements of the set $\{p_i\}$ are called vertices of the knot.*

Finally, even if one's goal is to study only knots, links of many components will arise.

□ **DEFINITION.** *A link is the finite union of disjoint knots. (In particular, a knot is a link with one component.) The unlink is the union of unknots all lying in a plane.*

Notice that the condition that the components of the unlink lie in a single plane is essential; examples of nontrivial links with each component unknotted have already been described. As with the definition of the unknot, ambiguities appear here; for instance, in the definition of the unlink does it matter what plane is used? Following the definition of equivalence presented in Section 3, these issues can be addressed.

EXERCISES

2.1. The ordering of the points $\{p_i\}$ used to define a knot is essential. Show that by correctly changing the ordering of the points, one might not get a knot at all. (Hint: with the vertices reordered a closed curve will still result, but is it necessarily simple?) Also, show that by changing the ordering of the points $\{p_i\}$ defining the trefoil, the resulting knot can be deformed into the unknot.

2.2. It is not clear from the definition that a knot has only one set of vertices. Prove that in fact the vertices of a knot form a well-defined set.

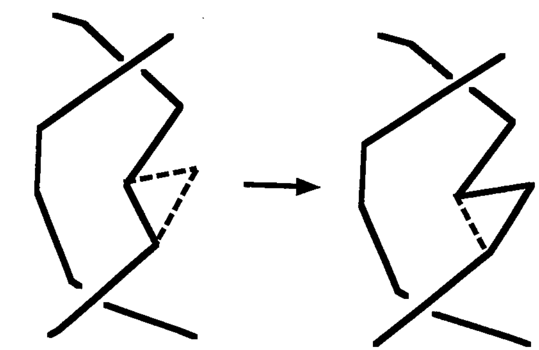
3 Equivalence of Knots, Deformations

The next step is to give a mathematical formulation of the idea of deforming knots. This is done with the notion of equivalence, which is in turn defined via elementary deformations.

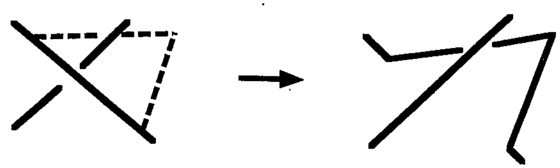
□ **DEFINITION.** *A knot J is called an elementary deformation of the knot K if one of the two knots is determined by a sequence of points $(p_1,p_2,\ldots,p_n)$ and the other is determined by the sequence $(p_0,p_1,p_2,\ldots,p_n)$, where (1) p_0 is a point which is not collinear with p_1 and p_n, and (2) the triangle spanned by (p_0,p_1,p_n) intersects the knot determined by $(p_1,p_2,\ldots,p_n)$ only in the segment $[p_1,p_n]$.*

Here a triangle is the flat surface bounded by the edges $[p_0,p_1]$, $[p_1,p_n]$, and $[p_n,p_0]$. It is defined formally as $T=\{xp_0+yp_1+zp_n \mid 0\le x,\ y,\ z,\text{ and } x+y+z=1\}$.

The second condition in the definition assures that in performing an elementary deformation the knot does not cross itself. Figure 2.5a illustrates an elementary deformation, and 2.5b illustrates a deformation which is not permitted. As examples have indicated, such crossings can change a knot into a different type of knot. Of course, the point of the definition is to make these ideas precise.



(*a*)



(*b*)

Figure 2.5

Knots K and J are called equivalent if K can be changed into J by performing a series of elementary deformations. More precisely:

□ **DEFINITION.** *Knots K and J are called equivalent if there is a sequence of knots* $K = K_0, K_1, \ldots, K_n = J$, *with each* K_{i+1} *an elementary deformation of* K_i, *for i greater than* 0.

This notion of equivalence satisfies the definition of an equivalence relation; it is symmetric, transitive, and reflexive, three facts that the reader can verify.

Knot theory consists of the study of equivalence classes of knots. For instance, proving that it is impossible to deform one knot into another is the same as proving that the two knots lie in different equivalence classes. Proving that a knot is nontrivial consists of showing that it is not contained in the equivalence class of the unknot.

TERMINOLOGY

It is usual in the subject to blur the distinction between a knot and its equivalence class. For instance, rather than say that a knot is equivalent to the unknot, one just states that the knot is unknotted. Similarly, when it is said that two knots are distinct, it is meant that the knots are inequivalent. This convention seldom can cause confusion, but will be avoided in ambiguous situations.

EXERCISES

3.1. Suppose a knot lies in a plane, and bounds a convex region in that plane. (Convex means that any segment with endpoints in the region is entirely contained in the region.) Prove that the knot is equivalent to a knot with 3 vertices. That is, describe how to construct a sequence of knots, each an elementary deformation of the previous one, starting with the convex planar knot and ending with a knot having exactly 3 vertices. Hint: Apply induction on the number of vertices.

3.2. Suppose that K and J are unknots lying in the same plane. (Recall that this means that K and J are each determined by three noncollinear points.) Show that K and J are equivalent by describing a method for finding the appropriate sequence of elementary deformations.

3.3. Exercises 3.1 and 3.2 show that two convex knots in a plane determine equivalent knots. This result is true for nonconvex knots, and is called the Schonflies Theorem. Prove the Schonflies theorem for planar knots with 4 and 5 vertices.

3.4. Is every knot with exactly 4 vertices unknotted?

3.5. Let K be a knot determined by points $(p_1,p_2,\ldots,p_n)$. Show that there is a number z such that if the distance from p_1 to p_1' is less than z, then K is equivalent to the knot determined by $(p_1',p_2,\ldots,p_n)$. Similarly, show there is a z such that every vertex can be moved a distance z without changing the equivalence class of the knot. (These are both detailed arguments in epsilons and deltas.)

3.6. Prove, using 3.5, that a knot can be arbitrarily translated or rotated by a sequence of elementary deformations.

3.7. Generalize the definition of elementary deformation, and equivalence, to apply to links. (Your definition should not permit one component to pass through another.)

4 Diagrams and Projections

Although a knot is a subset of space, all our work takes place in a plane. The pictures in this book all lie on a flat piece of paper and your practice

is done on a flat blackboard or piece of paper as well. How is it that a diagram on a piece of paper gives a well-defined knot? This is answered by formalizing the notion of knot diagram.

The function from 3-space to the plane which takes a triple (x,y,z) to the pair (x,y) is called the projection map. If K is a knot, the image of K under this projection is called the *projection* of K. A projection of the *figure-8* knot (knot 4_1 in the appendix) is illustrated in Figure 2.6.

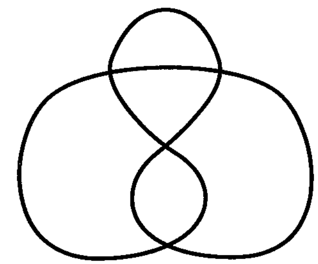


Figure 2.6

It is possible that different knots can have the same projection. Once the curve is projected into the plane, it is no longer clear which portions of the knot passed over other parts. To remedy this loss of information, gaps are left in the drawings of projections to indicate which parts of the knot pass under other parts. Such a drawing is called a knot *diagram*. In this book all the drawings of knots are really knot diagrams.

At this point the distinction between knots and equivalence classes of knots appears. Many different knots can have the same diagram, as the diagram indicates that certain portions of the knot pass over other portions, but not how high above they pass. It turns out that this does not matter! *If two knots have the same diagram they are equivalent*. To state this formally as a theorem requires a more careful study of projections.

Suppose that a knot has a projection as illustrated in Figure 2.7a. If that knot is rotated slightly in space, the resulting knot will have a projection as illustrated in

Figure 2.7b! Such knot projections have to be avoided as too much information has been lost in the projection.

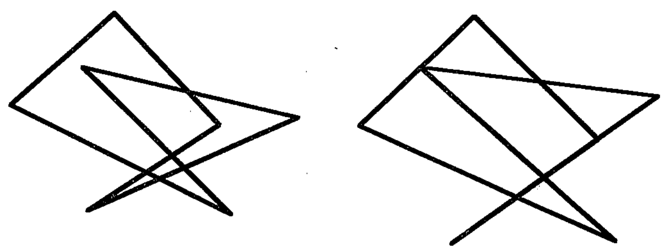


Figure 2.7

□ **DEFINITION.** *A knot projection is called a regular projection if no three points on the knot project to the same point, and no vertex projects to the same point as any other point on the knot.*

There are two theorems that make regular projections especially useful. The first states that if a knot does not have a regular projection then there is an equivalent knot nearby that does have a regular projection. The second states that if a knot does have a regular projection then all nearby knots are equivalent and also have regular projections. The notion of nearby is made precise by measuring the distance between vertices.

□ **THEOREM 1.** *Let K be a knot determined by the ordered set of points $(p_1,\dots,p_n)$. For every number $t > 0$ there is a knot K' determined by an ordered set $(q_1,\dots,q_n)$ such that the distance from q_i to p_i is less than t for all i, K' is equivalent to K, and the projection of K' is regular.*

□ **THEOREM 2.** *Suppose that K is determined by the sequence $(p_1,\dots,p_n)$ and has a regular projection. There is a number $t > 0$ such that if a knot K' is determined by $(q_1,\dots,q_n)$ with each q_i within a distance of t of p_i, then K' is equivalent to K and has a regular projection.*

Knot diagrams are only defined for knots with regular projections. The theorem relating knots to diagrams is the following:

□ **THEOREM 3.** *If knots K and J have regular projections and identical diagrams, then they are equivalent.*

PROOF

One approach is the following. First arrange that K is determined by an ordered sequence $(p_1,\dots p_n)$ and J is determined by the sequence $(q_1,\dots,q_n)$ with the projection of p_i and q_i the same for all i. This may require introducing extra points in the defining sequences for both knots.

Next perform a sequence of elementary deformations that replace each p_i with a q_i in the defining sequence for K. These moves are first applied to all vertices which do not bound intervals whose projections contain crossing points. Finally each crossing point can be handled. □

TERMINOLOGY

A knot diagram consists of a collection of arcs in the plane. These arcs are called either *edges* or *arcs* of the diagram. The points in the diagram which correspond to double points in the projection are called *crossing points*, or just *crossings*. Above the crossing point are two segments on the knot; one is called an *overpass* or *overcrossing*, the other the *underpass* or *undercrossing*. Notice that the number of arcs is the same as the number of crossings.

With Theorem 3 it is now possible to blur the distinction between a knot and its diagram. There is usually no confusion created by not distinguishing a knot diagram from an equivalence class of a knot. To be clear, though: a knot is a subset of 3-space, knots determine equivalence classes of knots, and knots with regular projections have diagrams, which are drawings in the plane.

EXERCISES

4.1. Fill in the details of the proof of Theorem 3.

4.2. Sketch a proof of Theorem 1. (A proof can make use of Exercise 6, Section 3. A projection is regular as long as 1) no line joining two vertices is parallel to the vertical axis, 2) no vertices span a plane containing a line parallel to the vertical axis, and 3) there are no triple points in the projection. Argue that the knot can be rotated slightly to achieve conditions 1 and 2, and then deal with triple points.)

4.3. Prove Theorem 2. The previous hint should help here.

4.4. Show that the trefoil knot can be deformed so that its (nonregular) projection has exactly one multiple point.

5 Orientations

Knots can be oriented, or, informally, given a sense of direction. Recall that a knot is determined by its (ordered) set of vertices. If the ordered set of vertices is $(p_1, \ldots, p_n)$, then, as noted earlier, any cyclic permutation of the vertices gives the same knot. It is also true that reversing the order of the vertices will yield the same knot.

□ **DEFINITION.** *An oriented knot consists of a knot and an ordering of its vertices. The ordering must be chosen so that it determines the original knot. Two orderings are considered equivalent if they differ by a cyclic permutation.*

The orientation of a knot is usually represented by placing an arrow on its diagram. The connection with the definition of orientation should be clear.

The notion of equivalence is easily generalized to the oriented setting. If a knot is oriented, an elementary deformation results in a knot which is naturally oriented. Hence, an elementary deformation of an oriented knot is again an oriented knot.

□ **DEFINITION.** *Oriented knots are called oriented equivalent if there is a sequence of elementary deformations carrying one oriented knot to the other.*

One of the hardest problems that arises in knot theory is in distinguishing equivalence and oriented equivalence. The first examples of knots which are equivalent but not oriented equivalent were described by H. Trotter in 1963; for example, the (3,5,7)-pretzel knot can be oriented in two ways, and Trotter showed the resulting oriented knots are not oriented equivalent, even though they are the same when orientations are ignored.

Another related definition will be useful later.

□ **DEFINITION.** *The reverse of the oriented knot determined by the ordered set of vertices $(p_1, \ldots, p_n)$, is the oriented knot K^r with the same vertices but with their order reversed. An oriented knot K is called reversible if K and K^r are oriented equivalent. If K is not oriented, it is called reversible if for some choice of orientation it is reversible.*

EXERCISES

5.1. Formulate a definition of oriented link.

5.2. Any oriented knot, or link, determines an unoriented link. Simply ignore the orientation. Given a knot, there are at most two equivalence classes of oriented knot that determine its equivalence class, ignoring orientations. (Why?)

(a) What is the largest possible number of distinct oriented n component links which can determine the same unoriented link, up to equivalence? Try to construct an example in which this maximum is achieved. (Do not attempt to prove that the oriented links are actually inequivalent. This will have to wait until more techniques are available.)

(b) Show that any two oriented links which determine the unlink as an unoriented link are oriented equivalent.

5.3. Explain why if an unoriented knot is reversible, then for any choice of orientation it is reversible.

5.4. Show that the (p,p,q)-pretzel knot is reversible.

5.5. The knot 8_{17} is the first knot in the appendix that is not reversible, a difficult fact to prove. Find inversions for some of the knots that precede it. Several are not obvious.

5.6. Classically, what has been defined here as the reverse of a knot was called the inverse. The change in notation arose from high-dimensional considerations that will be discussed in Chapter 9. The inverse is now defined as follows. Given an oriented knot, multiplying the z-coordinates of its vertices by -1 yields a new knot, K^m, called the *mirror image*, or *obverse* of the first. The *inverse* of K is defined to be K^{mr}.

(a) How are the diagrams of a knot and its obverse and inverse related?

(b) Given a knot diagram it is possible to form a new knot diagram by reflecting the diagram through a vertical line in the plane, as illustrated in Figure 2.8. What operation on knots in 3-space does this correspond to?

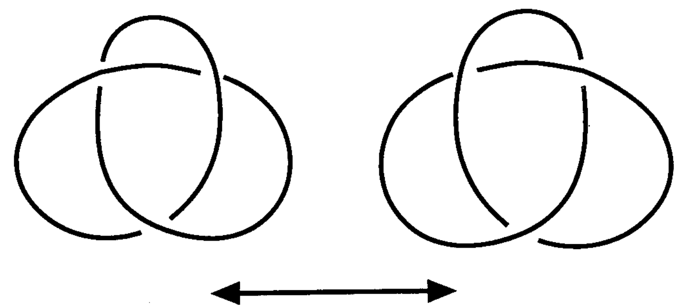


Figure 2.8

(c) Show that the operation described in part b) yields a knot equivalent to the obverse of the original knot.

CHAPTER 3:
COMBINATORIAL TECHNIQUES

The techniques of knot theory which are based on the study of knot diagrams are called combinatorial methods. These techniques are usually easy to describe and yet provide deep results. For instance, in this chapter such methods will be used to prove that nontrivial knots exist and then to demonstrate that there is in fact an infinite number of distinct knots.

Combinatorial tools often appear as unnatural or ad hoc. In many cases alternative perspectives, though more abstract, can provide insights. One of the successes of algebraic topology is to provide such perspectives, but in some cases, the efficacy of combinatorial techniques remains mysterious. Recent progress in combinatorial knot theory will be described in Chapter 10.

1 Reidemeister Moves

In what ways are diagrams of equivalent knots related? Clearly, even a single elementary deformation can have a dramatic effect on the diagram. Some of the simplest changes in a diagram that can occur when a knot is deformed are illustrated in Figure 3.1. In the figure only that portion of the diagram where a change occurs is illustrated.

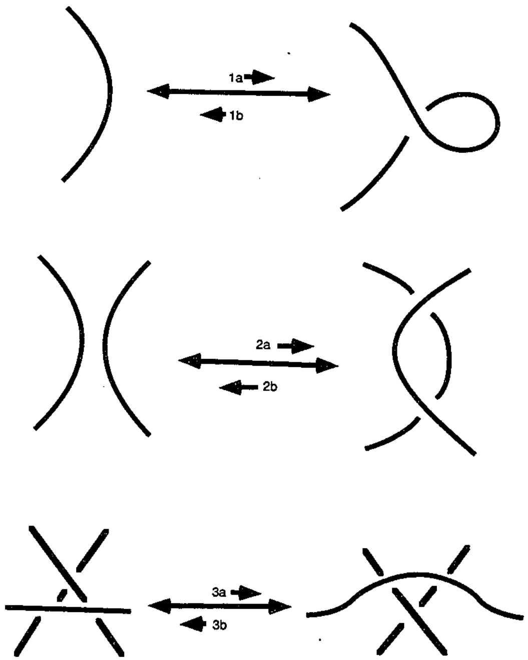


Figure 3.1

Each of the three figures represents a pair of possible changes in a diagram; each operation is paired with its inverse. These six simple operations which can be performed on a knot diagram without altering the corresponding knot are called *Reidemeister moves* . The key observation in combinatorial knot theory was made by Alexander and Briggs:

□ **THEOREM 1.** *If two knots (or links) are equivalent, their diagrams are related by a sequence of Reidemeister moves.*

PROOF

If you have already done some of the exercises showing that different diagrams can represent the same knot then this result should seem intuitively clear. In turning one diagram into the other the only changes that you ever need to make are these Reidemeister moves. The full proof is a detailed argument keeping track of a number of cases, but the main ideas are fairly simple.

Suppose that K and J represent equivalent knots, and that both have regular projections. Then K and J are related by a sequence of knots, each obtained from the next by an elementary deformation. A small rotation will assure that each knot in the sequence has a regular projection, and thus the proof is reduced to the case of knots related by a single elementary deformation.

Again after performing a slight rotation, it can be assured that the triangle along which the elementary deformation was performed projects to a triangle in the plane. That planar triangle might contain many crossings of the knot diagram. However, it can be divided up into many small triangles, each of which contains at most one crossing. This division can be used to describe the single elementary deformation in a sequence of many small elementary deformations; the effect of each on the diagram is quite simple. The proof is completed by checking that only Reidemeister moves have been applied. □

EXERCISES

1.1. Show that the change illustrated in Figure 3.2 can be achieved by a sequence of two Reidemeister moves.

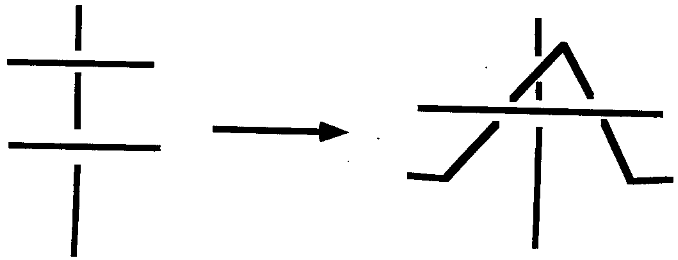


Figure 3.2

1.2. Find a sequence of Reidemeister moves that transforms the diagram of the unknot drawn in Figure 3.3 into a diagram without crossings. Here is a harder exercise: What is the least number of Reidemeister moves needed for such a sequence? Can you prove that this is the least number that suffices?

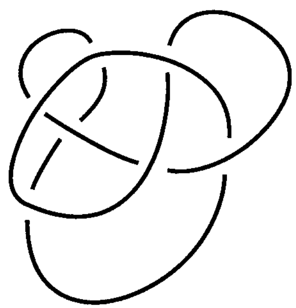


Figure 3.3

2 Colorings The method of distinguishing knots using the "colorability" of their diagrams was invented by Ralph Fox. The procedure is simple: A knot diagram is called *colorable* if each arc can be drawn using one of three colors, say red (R), yellow (Y), and blue (B), in such a way that 1) at least

two of the colors are used, and 2) at any crossing at which two colors appear, all three appear. Figure 3.4 illustrates a coloring of a knot diagram. Exercise 2.1 is a quick problem, asking you to check which of the diagrams for knots with 7 or fewer crossings, as illustrated in Appendix 1, are colorable.

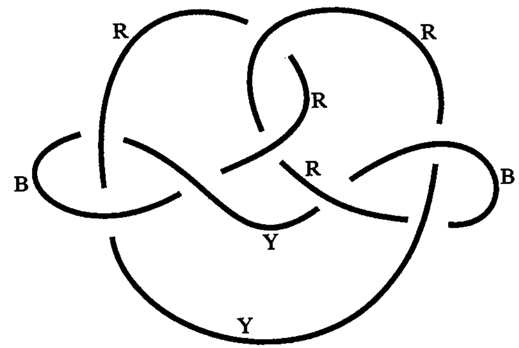


Figure 3.4

Is it possible that some diagrams for a knot are colorable while others are not? Our first result in combinatorial knot theory is that the answer is no.

□ **THEOREM 2.** *If a diagram of a knot, K, is colorable, then every diagram of K is colorable.*

Hence the following definition makes sense:

□ **DEFINITION.** *A knot is called colorable if its diagrams are colorable.*

The proof of Theorem 2 is the model for most of the proofs of later combinatorial results. But before giving it, one immediate consequence should be noted; nontrivial knots exist! Clearly the unknot is not colorable because its standard projection cannot be colored. It follows that any colorable knot is nontrivial. Further consequences appear in the exercises.

PROOF

(Theorem 2) It is sufficient to show that if a Reidemeister move is performed on the colorable diagram of a knot, then the resulting diagram is again colorable. Hence, the proof breaks into six steps, one for each Reidemeister move. Each step consists of checking various cases and none is difficult, although some are a bit tedious. One step is presented here; the others are left to the exercises.

Suppose that Reidemeister move 2b is performed on a colored knot diagram. It must be shown that the new diagram is again colorable. There are two cases. In the first, the arcs are colored with two (and hence three) colors, as illustrated in Figure 3.5a. (Only the affected portions of the knots are included in these illustrations.) The new diagram can be colored as before, with the altered section colored as in Figure 3.5b. As two colors still appear, the resulting diagram is still colorable.

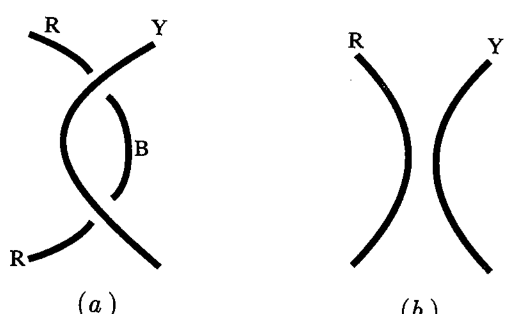


Figure 3.5

The second possibility is that both of the affected arcs start out colored with the same color. In this case, after performing the Reidemeister move the arcs can still be colored with that same color and the rest of the diagram can be colored as it was originally. All the requirements of colorability are still satisfied.

Checking Reidemeister moves 1a, 1b, and 2a, are all as simple as this. Moves 3a and 3b present a few more cases to check. □

EXERCISES

2.1. Which of the knot diagrams with seven or fewer crossings, as illustrated in Appendix 1, are colorable?

2.2. For which integers n is the $(2,n)$-torus knot in Figure 3.6a colorable? The knot illustrated in Figure 3.6b is called the *n-twisted double of the unknot*, where $2n$ is the number of crossings in the vertical band. The trefoil results when $n = -1$. What if $n = 0$ or 1? For which values of n is the n-twisted double of the unknot colorable?

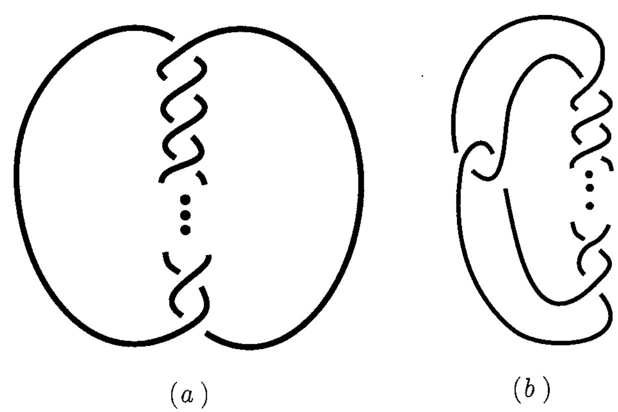


Figure 3.6

2.3. Discuss the colorability of the (p,q,r)-pretzel knots.

2.4. (a) Prove the coloring theorem for Reidemeister move 1a.

(b) How many cases need to be considered in proving Theorem 2 for Reidemeister move 3a?

(c) Check each of these cases.

(d) Complete the proof of Theorem 2.

2.5. Given an oriented link of two components, J and K, it is possible to define the *linking number* of the components as follows. Each crossing point in the diagram is assigned a sign, +1 if the crossing is right-handed and −1 if it is left-handed. (A *right-handed* crossing is a crossing at which an observer on the overcrossing, facing in the direction of the overcrossing, would view the undercrossing as passing from right to left. Right and left crossings are illustrated in Figure 3.7.) The linking number of K and J, $\ell k(K,J)$, is defined to be the sum of the signs of the crossing points where J and K meet, divided by 2.

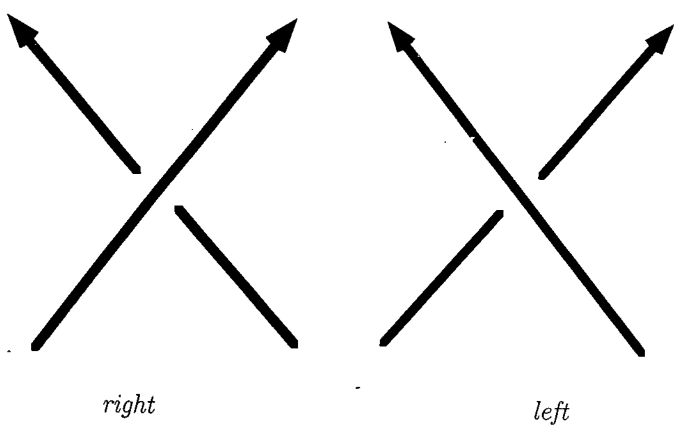


Figure 3.7

(a) Use the Reidemeister moves to prove that the linking number depends only on the oriented link, and not on the diagram used to compute it.

(b) Figure 3.8 illustrates an oriented *Whitehead link*. Check that it has linking number 0.

(c) Construct examples of links with different linking numbers.

2.6. This exercise demonstrates that the linking number is always an integer. First note that the sum used to compute linking numbers can be split into the sum of the signs of the crossings where K passes over J, and the sum of the crossings where J passes over K.

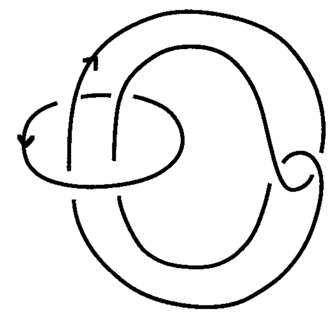


Figure 3.8

(a) Use Reidemeister moves to prove that each sum is unchanged by a deformation.

(b) Show that the difference of the two sums is unchanged if a crossing is changed in the diagram.

(c) Show that if the crossings are changed so that K always passes over J, the difference of the sums is 0. (This link can be deformed so that K and J have disjoint projections.)

(d) Argue that the linking number is always an integer, given by either of the two sums. (This is the usual definition of linking number. The definition in Exercise 2.5 makes it clear that $\ell k(K,J) = \ell k(J,K)$.)

2.7. The definition of colorability is often stated slightly differently. The requirement that at least two colors are used is replaced with the condition that all three colors appear.

(a) Show that the unlink of two components has a diagram which is colorable using all three colors and another diagram which is colorable with exactly two colors.

(b) Why is it true that for a knot, once two colors appear all three must be used, whereas the same statement fails for links?

(c) Explain why the proof of Theorem 2 applies to links as well as to knots.

2.8. Prove that the Whitehead link illustrated in Figure 3.8 is nontrivial, by arguing that it is not colorable.

2.9. In this exercise you will prove the existence of an infinite number of distinct knots by counting the number of colorings a knot has.

If a knot is colorable there are many different ways to color it. For instance, arcs that were colored red can be changed to yellow, yellow arcs changed to blue, and blue arcs to red. The requirements of the definition of colorability will still hold. There are six permutations of the set of three colors, so any coloring yields a total of six colorings. For some knots there are more possibilities.

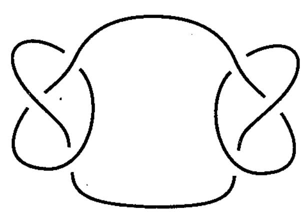


Figure 3.9

(a) Show that the standard diagram for the trefoil knot has exactly six colorings.

(b) How many colorings does the *square knot* shown in Figure 3.9 have?

(c) The number of colorings of a knot projection depends only on the knot; that is, all diagrams of a knot will have the same number of colorings. Outline a proof of this.

(d) Use the connected sum of n trefoils, illustrated in Figure 3.10, to show that there are an infinite number of distinct knots.

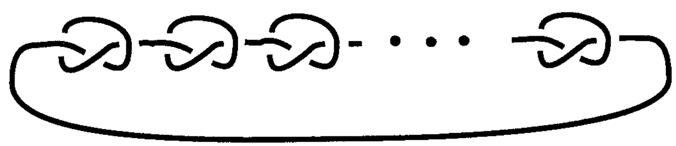


Figure 3.10

3 A Generalization of Colorability, mod p Labelings

How can colorability be generalized? Is it possible to use more than three colors to describe new methods of distinguishing knots? There are actually several ways to generalize colorability, the first of which is presented in this section.

In describing the method of colorings in the previous section, instead of labeling the arcs of the knot diagram with colors, three integers, 0,1, and 2, could have been used. The condition on colorings at crossings translates into the simple statement that if the overcrossing is labeled x and the two other arcs y and z, then the difference $2x - y - z$ is divisible by 3, or, more succinctly, $2x - y - z = 0 \pmod 3$. (Check that this condition is equivalent to the coloring condition.) A possible generalization immediately appears:

□ **DEFINITION.** *A knot diagram can be labeled mod p if each edge can be labeled with an integer from 0 to $p-1$ such that 1) at each crossing the relation $2x - y - z = 0 \pmod p$ holds, where x is the label on the overcrossing and y and z the other two labels, and 2) at least two labels are distinct.*

Figure 3.11 illustrates a mod 7 labeling of a knot.

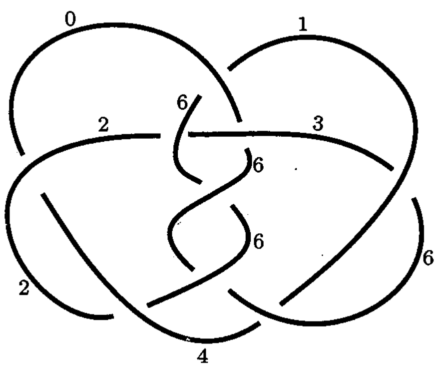


Figure 3.11

For reasons that will be made clear in the exercises, p will be restricted to the odd primes. In Exercise 3.3 the reader is invited to check that whether or not a knot

diagram can be labeled mod p depends only on the equivalence class of the knot, so that Theorem 2 generalizes to this new situation. Figure 3.12 illustrates one step; if Reidemeister move 2b is performed on a labeled diagram, the resulting diagram can again be labeled.

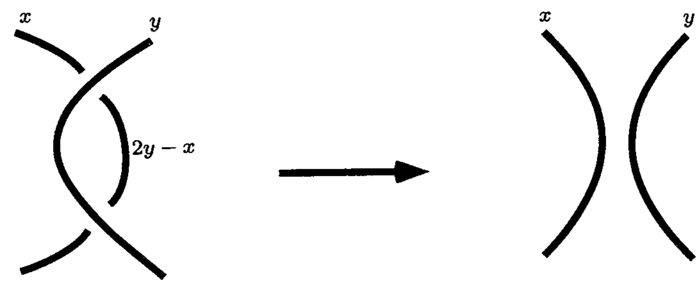


Figure 3.12

□ **THEOREM 3.** (*Labeling theorem*) *If some diagram for a knot can be labeled* mod p *then every diagram for that knot can be labeled* mod p.

EXERCISES

3.1. Determine which knots with 6 or fewer crossings can be labeled mod 5.

3.2. For what primes p can the trefoil knot diagram be labeled mod p?

3.3. Prove Theorem 3 by showing that if any Reidemeister move is performed on a labeled diagram, the resulting diagram can again be labeled.

3.4. Show that if all the labels of a knot that is labeled mod 3 are multiplied by 5, the resulting labeling is a labeling mod 15. This gives some indication as to why p is restricted to the primes.

3.5. If p is 2, other difficulties come up. Explain why no knot can be labeled mod 2. (Modulo 2, what does the crossing relationship say?)

3.6. Check that the theory of labelings applies to links of many components.

3.7. Show that the knots 4_1, 7_1, and 8_{16} are distinct by using mod 5 and mod 7 labelings. (Find mod 5 and mod 7 labelings of 8_{16}.)

4 Matrices, Labelings, and Determinants

Linear algebra simplifies the problem of labeling knot diagrams; just as important is the fact that, with the introduction of matrices, many new knot invariants appear. Some of these invariants are introduced here. These invariants are studied in greater depth in Chapter 7.

Here is an algebraic reduction of the problem. Given a knot diagram, label each arc of the diagram with a variable, say x_i. At each crossing a relation between the variables is defined: if arc x_i crosses over arcs x_j and x_k, then $2x_i - x_j - x_k = 0 \pmod p$. A knot can be labeled mod p if there is a mod p solution to this system of equations with not all x_i equal.

Whether or not a knot is colorable, or can be labeled mod p, has now been reduced to a problem of linear algebra, that of studying the solutions to a system of linear

equations. As usual in linear algebra, the use of matrices will simplify the problem.

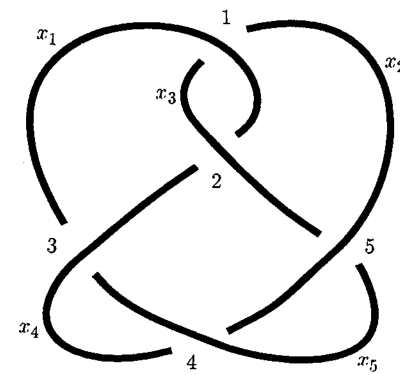


Figure 3.13

For example, the knot in Figure 3.13 is drawn with its arcs labeled and its crossings numbered. The corresponding system of equations that needs to be solved is given by the matrix below. The rows correspond to the equations determined by each crossing, the columns to the variables taken in order.

$$\begin{pmatrix} 2 & -1 & -1 & 0 & 0 \\ -1 & 0 & 2 & -1 & 0 \\ -1 & 0 & 0 & 2 & -1 \\ 0 & -1 & 0 & -1 & 2 \\ 0 & 2 & -1 & 0 & -1 \end{pmatrix}$$

Standard techniques of linear algebra apply to solving systems of equations mod p as well as for finding real or rational solutions. (Formally, for p prime the integers mod p form a field.) Unfortunately, the added condition in the present problem, that the solutions have at least two of the x_i distinct, introduces a few subtleties that need to be addressed before general results can be presented.

Two preliminary observations are needed. First note that setting each $x_i = 1$ yields a solution to the system of equations. Second, observe that any two solutions can be added together to yield another solution.

These remarks imply that if there is a solution with not all entries equal, there is such a solution with $x_n = 0$. (x_n could be replaced with any other x_i here.) Conversely, a *nontrivial* solution with $x_n = 0$ results in a labeling of the knot. Hence, a solution with not all x_i equal corresponds to a nontrivial solution to the system of equations determined by the original matrix with its last column deleted.

It is easier to work with problems related to square matrices, and fortunately the given problem can be reduced to this setting. This is done by showing that any one of the equations is a consequence of the others. In terms of the matrix, multiplying certain of the rows by -1 results in a matrix with its rows adding to 0.

The correct choice of -1's is not obvious; here is the algorithm: Orient the knot. At each crossing in the diagram put a dot to the right of the overcrossing, just before the crossing point. Now, count how many arcs of the diagram must be crossed by a path from the dot to a point in the plane far from the diagram. If an odd number of arcs are crossed, then multiply the corresponding row of the matrix by -1. It is fairly simple to show that

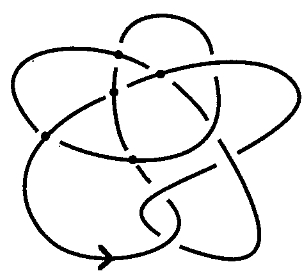


Figure 3.14

the sum of the rows is now trivial. In Figure 3.14 the crossings that correspond to rows that are multiplied by -1 are marked.

The following result summarizes the discussion above.

□ **THEOREM 4.** *There is an $n \times n$ matrix corresponding to a knot diagram with n arcs. Deleting any one column and any one row yields a new matrix. The knot can be labeled* mod p *if and only if the corresponding set of equations has a nontrivial* mod p *solution.*

Of course whether or not the system of equations has a nontrivial solution depends on the determinant of the matrix. A solution exists if the determinant is 0, or, working mod p, if the determinant is divisible by p. Furthermore, the number of solutions is determined by the mod p nullity of the matrix.

(The nullity of a matrix is the dimension of the kernel of the matrix, thought of as a linear transformation. More algorithmically, any square matrix with entries in a field, (mod p entries in the present case), can be diagonalized by performing row and column operations; that is, by adding multiples of rows or columns to other rows or columns respectively. The number of 0's on the diagonal (or entries divisible by p if working mod p) is the *nullity*. With more care, a square integer matrix can be diagonalized, using only integer row and column operations. Performing this integer diagonalization performs mod p diagonalizations for all p simultaneously. The exercises illustrate these procedures.)

□ **DEFINITION.** *The determinant of a knot is the absolute value of the determinant of the associated $(n-1) \times (n-1)$ matrix constructed above.*

□ **DEFINITION.** *The* mod p *rank of a knot is the* mod p *nullity of the associated* $(n-1)\times(n-1)$ *matrix constructed above.*

Of course, for these two definitions to give well-defined invariants, it must be proved that none of the choices involved, of either the knot diagram or the ordering of the labels on the arcs and crossings, affects the determinant or mod p rank of the associated matrix.

□ **THEOREM 5.** *The determinant of a knot and its* mod p *rank are independent of the choice of diagram and labeling.*

PROOF

There are two parts to the proof. The first is purely linear algebra, observing facts about the determinant and nullity of matrices. The second calculates the effect of the choice of labelings and the Reidemeister moves on the associated matrix.

As far as the linear algebra goes, a needed result states that if, for a square matrix, the sum of the rows and the sum of the columns is 0, then if a row and column are removed, the nullity (and the absolute value of the determinant) of the resulting matrix does not depend on which row and column were removed. A simpler result states that if the matrix is changed by adding a new row and column, each containing all 0's except for a single 1 on the diagonal, then the nullity and determinant are unaffected.

The rest of the argument checks the effect of the Reidemeister moves on the associated matrix. For example, Reidemeister move 2a introduces two new rows and two new columns. Two of the new columns result from splitting one of the old arcs into two, and hence the sum of those two columns has entries determined by the one old column. A few row and column operations show that the new matrix can be changed into the old, with two new rows and columns added, each of which has a single ± 1 in it. The full argument for this and the other Reidemeister moves, is left to the reader. □

TORSION INVARIANTS

The determinant and ranks are captured by stronger invariants. It is relatively easy to diagonalize a matrix when working mod p; any nonzero entry can be used to clear out a row and column. Diagonalizing over the integers is harder, though possible, as is proved in most modern algebra texts in the classification of abelian groups. The proof uses the Euclidean algorithm. The typical result states that a square integer matrix can be diagonalized so that each entry on the diagonal divides the next entry. If the matrix associated to a knot is diagonalized in this way, the resulting diagonal entries are called the *torsion invariants* of the knot. Their product is the determinant of the knot, and the number of entries which are divisible by p is the mod p rank of the knot.

The proof that these are well-defined knot invariants will not be given. The best approach relies on the theory of abelian groups. The matrix associated to a knot can be viewed as a presentation matrix for an abelian group. The various alterations in the matrix do not affect the group so determined, and the torsion invariants are just the torsion invariants of this group.

EXERCISES

4.1. For each knot with 6 or fewer crossings find the associated matrix, and its determinant. In each case, for what p is there a mod p labeling?

4.2. The knots 8_{18} and 9_{24} both have determinant 45. Check that one has a mod 3 rank of 1, while the other has a mod 3 rank of 2. The knots 8_8 and 9_{49} both have determinant 25. Compute their mod 5 ranks.

4.3. Prove the linear algebra results stated in the proof of Theorem 5.

4.4. Because the unknot has particularly simple diagrams, the arguments given above really need to be modified slightly. The two diagrams for the unknot that cause difficulties are the diagram with no crossings, and the diagram with exactly one crossing. What goes wrong in these cases? Why don't these problems occur in other situations? How would you correct for these minor problems? (Define the determinant and nullity of a 0×0 matrix to be 1.)

4.5. Prove that the determinant of a knot is always odd. (See Exercise 5 of the previous section, relating to mod 2 labelings. Also, this result does not apply for links of more than one component.)

4.6. Show that if a knot has mod p rank n, then the number of mod p labelings is $p(p^n - 1)$.

5 The Alexander Polynomial

In the previous section it was seen that the simple notion of colorability leads to a study of determinants of matrices. The following description of the Alexander polynomial greatly extends the use of matrices and determinants. In this case, rather than work

with entries that are integers the entries of the matrix are polynomials.

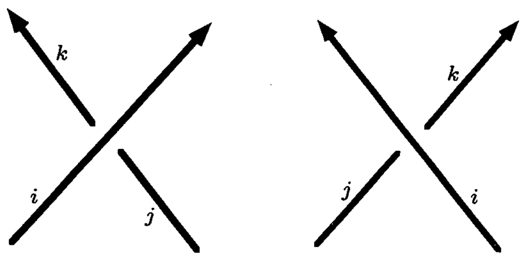
Alexander's original description was based on labeling the regions in the plane bounded by the arcs of the diagram, and Reidemeister was the first to give a presentation focusing on the arcs. Since then, many alternative definitions have been found. Chapter 10 provides a modern viewpoint, one that is quite simple, and that provides access to many new invariants.

To compute the Alexander polynomial of a knot, $A_K(t)$, first pick an oriented diagram for K. Number the arcs of the diagram, and separately number the crossings. Next, define an $n \times n$ matrix, where n is the number of crossings (and arcs) in the diagram, according to the following procedure:

If the crossing numbered ℓ is right-handed with arc i passing over arcs j and k, as illustrated in Figure 3.15a, enter a $1-t$ in column i of row ℓ, enter a -1 in column j of that row, and enter a t in column k of the row. If the crossing is left-handed, as illustrated in Figure 3.15b, enter a $1-t$ in column i of row ℓ, enter a t in column j and enter -1 in column k of row ℓ. All of the remaining entries of row ℓ are 0. (An exceptional case occurs if any of i, j, or k are equal. In this exceptional case, the sum of the entries described above is put in the appropriate column. For instance, if $j = k$ for some left-handed crossing, enter $-1+t$ in column j. What if $j = k$ at a right-handed crossing?)

□ **DEFINITION.** *The $(n-1) \times (n-1)$ matrix obtained by removing the last row and column from the $n \times n$ matrix just described is called an Alexander matrix of K. The determinant of the Alexander matrix is called the Alexander polynomial of K. (The determinant of a 0×0 matrix is defined to be 1.)*

Unfortunately, this polynomial depends on the choice of the original diagram as well as on the other choices involved in its description. That dependence is captured by the following theorem.



(a) (b)

Figure 3.15

□ **THEOREM 6.** *If the Alexander polynomial for a knot is computed using two different sets of choices for diagrams and labelings, the two polynomials will differ by a multiple of $\pm t^k$, for some integer k.*

For example, applying this procedure to the trefoil yields the polynomial $t^2 - t + 1$. Another set of choices might give $-t^4 + t^3 - t^2$. See below.

SKETCH OF PROOF

The argument is more detailed than, but quite similar to, the proof of Theorem 5. With some care, the reader should be able to check the effect of performing Reidemeister moves on the Alexander matrix. The complete proof includes one new difficult step; analyzing the effect of a change of orientation. It will be shown that the Alexander polynomial of the reverse of a knot K is obtained from

the Alexander polynomial of K by substituting t^{-1} for t and multiplying by an appropriate power of t, and perhaps multiplying by -1. (See Exercise 5.7.) Hence, the independence of the Alexander polynomial on orientation follows from its symmetry; replacing t with t^{-1} returns the same polynomial multiplied by some power of t. This symmetry property will be discussed in Chapter 6. (Alexander was unable to find a proof; a complete argument was first given by Seifert.)

EXAMPLES

The trefoil knot provides the simplest example of a knot with nontrivial Alexander polynomial. Figure 3.16 indicates a labeling of the arcs and crossings. The associated matrix is:

$$\begin{pmatrix} 1-t & -1 & t \\ t & 1-t & -1 \\ -1 & t & 1-t \end{pmatrix}$$

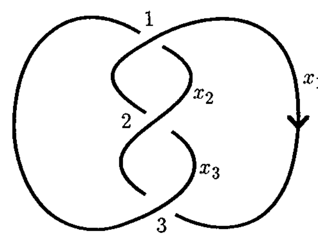


Figure 3.16

Deleting the bottom row and the last column gives a 2×2 Alexander matrix with determinant $t^2 - t + 1$.

Consider a harder example, the $(2, n)$-torus knot, shown in Figure 3.17. If the diagram is labeled as was done for the trefoil, the Alexander polynomial is given as the determinant of the $(n-1) \times (n-1)$ matrix

$$\begin{pmatrix} 1-t & -1 & 0 & 0 & \cdots & 0 \\ t & 1-t & -1 & 0 & \cdots & 0 \\ 0 & t & 1-t & -1 & \cdots & 0 \\ & & & \vdots & 1-t & -1 \\ 0 & \cdots & \cdots & 0 & t & 1-t \end{pmatrix}$$

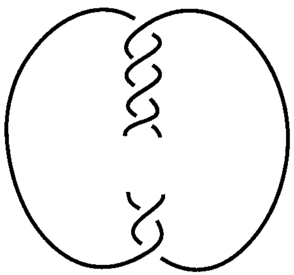


Figure 3.17

Clearly, to compute the exact determinant here would take a fairly detailed inductive argument. (The result turns out to be $(t^n+1)/(t+1)$.) Without actually computing the determinant it is easily proved that for different positive n the Alexander polynomials are distinct. Note first that the coefficient of the lowest degree term, the constant term, is the determinant of the matrix obtained by setting $t = 0$. The result is 1. The highest degree term is found by taking the determinant of the matrix containing only the t terms of the matrix above; that is remove all the $\pm 1's$. The resulting determinant is t^{n-1}.

Hence the Alexander polynomial of the $(2,n)$-torus knot is of degree exactly $n-1$. In particular, these knots form an infinite family of distinct knots, all of which are distinguished by the Alexander polynomial.

EXERCISES

5.1. Compute the Alexander polynomial for several knots in the appendix.

5.2. Relate the value of the Alexander polynomial of a knot evaluated at -1 to the determinant of the knot, defined in the previous section.

5.3. Check that Reidemeister move 1a does not change the Alexander polynomial.

5.4. It is possible to construct knots with the same polynomial, but which can be distinguished by their mod p ranks for some p. Compute the polynomials of 8_{18} and 9_{24}

to check that they are identical. In Exercise 4.2 of this chapter these knots were distinguished using the mod 3 ranks.

5.5. Show that the knot in Figure 3.18 has Alexander polynomial 1. (This is one of only two knots with 11 or fewer crossings that has trivial polynomial, other than the unknot.) Use Exercise 5.2 to argue that the knot cannot be distinguished from the unknot using labelings. Stronger algebraic techniques (Chapter 5) or combinatorial tools (Chapter 10) can be used to prove it is nontrivial.

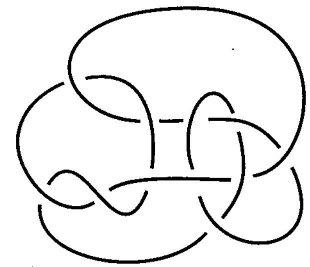


Figure 3.18

5.6. Prove that a knot and its mirror image, as illustrated in Figure 3.19, have the same polynomial. (Hint: Label the mirror image in the obvious way, but reverse its orientation.)

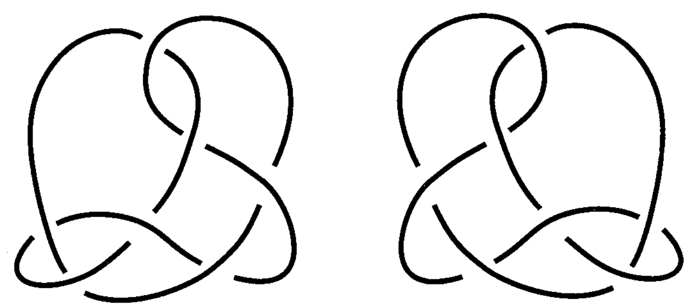


Figure 3.19

5.7. Show that the Alexander polynomial of K with its orientation reversed is obtained from the polynomial of K by substituting t^{-1} for t, and multiplying by the appropriate power of t, and perhaps changing sign.

CHAPTER 4: GEOMETRIC TECHNIQUES

Consider the surface drawn in Figure 4.1. It is built from a disk by attaching two twisted bands. Note that the boundary, or edge, of the surface is a knotted curve. In fact, the boundary is a trefoil knot.

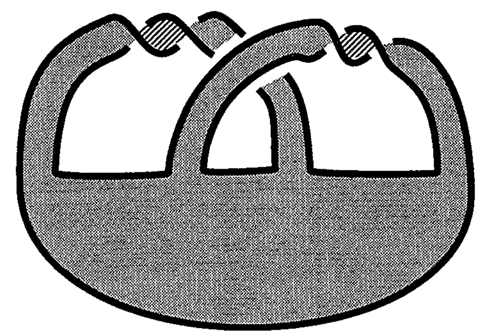


Figure 4.1

By studying the surface it is possible to learn more about the trefoil knot. In general, the term *geometric techniques* refers to the methods of knot theory that are based on working with surfaces. The use of these methods is motivated by a theorem stating that for every knot there is some surface having that knot as its boundary. An important application, on which this chapter ends, is the prime decomposition theorem for knots.

The first section of this chapter presents the basic definition of surface. The discussion corresponds closely to that of Chapter 2 where knot is defined. Naturally the definition is more technical. For a knot the interest is entirely in its placement in space; a surface has additional structure which is independent of its placement. For instance, the surface in Figure 4.1 is clearly different from a disk. The concept of internal, or *intrinsic*, properties of surfaces is made precise with the notion of homeomorphism, that is also described in Section 1.

Section 2 presents the fundamental theorems concerning surfaces. These results completely classify surfaces in terms of intrinsic properties. Once this internal structure of surfaces is understood the focus can shift to the placement of surfaces in space and to the knotted boundaries of surfaces. Section 3 begins the application of surface theory to knot theory; it is proved that every knot is the boundary of some surface. Sections 4 and 5 address the prime decompostion theorem, with Section 4 devoted to building the tools of the proof and Section 5 outlining the details of the argument.

1 Surfaces and Homeomorphisms

As with knots, it is possible to define a surface using the notion of differentiability. Again, a simpler working definition can be given using polyhedra.

Any 3 noncollinear points in 3-space, p_1, p_2, and p_3, form the vertices of a unique triangle. That triangle is

defined to be the set of points

$$\{xp_1 + yp_2 + zp_3 \mid x + y + z = 1,\ x,\ y,\ z \geq 0\},$$

where each p_i is thought of as a vector in R^3. The union of a finite collection of triangles is called a *polyhedral surface* if: (1) each pair of triangles is either disjoint or their intersection is a common edge or vertex, (2) at most two triangles share a common edge, and (3) the union of the edges that are contained in exactly one triangle is a disjoint collection of simple polygonal curves, called the *boundary* of the surface. This third condition rules out such possibilities as a surface being the union of exactly two triangles meeting at a vertex. (In this case the union of the edges contained in exactly one triangle would be all six edges; these form two unknots meeting in the common vertex—they are not disjoint.) Figure 4.2 illustrates a simple polyhedral surface, a planar square with a square hole in its center. It is illustrated as the union of a collection of triangles.

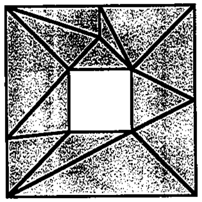


Figure 4.2

Surfaces will be drawn smoothly. Any smooth surface can be closely approximated by a polyhedral surface, but as the number of triangles required can be extremely large, it is easier to leave that *triangulation* out of the illustration. The details of the relationship between smooth and polyhedral surfaces is part of the foundational material of geometric topology.

ORIENTATION

The intuitive approach to orientability states that a surface is orientable if it is two-sided. The Möbius band is the standard example of a nonorientable surface. In calculus, a surface is called orientable if there is a nowhere vanishing vector field normal to the surface. For polyhedral surfaces there is a simple definition which corresponds to both the intuitive idea and the formal definition given in calculus.

□ **DEFINITION.** *A polyhedral surface is orientable if it is possible to orient the boundary of each of its constituent triangles in such a way that when two triangles meet along an edge, the two induced orientations of that edge run in opposite directions.*

A surface can be *triangulated*, that is, described as the union of triangles, in many different ways, and the definition of orientability appears to depend on the choice of triangulation. However, whether or not a surface can be oriented is actually independent of the choice of triangulation.

HOMEOMORPHISM

A notion of deformation of polyhedral surfaces can be given in much the same way as was done for knots. An important observation is that, although one surface might not be deformable into a second surface, the two might be intrinsically the same; that is, they are indistinguishable without reference to how they sit in space. For example, the number of boundary components of a surface is intrinsic; an inhabitant of the surface could determine this number.

However, whether or not the boundary is knotted can only be seen from a three-dimensional perspective.

This idea of intrinsic equivalence is formally defined as *homeomorphism*. Surfaces F and G in 3-space are called homeomorphic if there is a continuous function with domain F and range G which is both one-to-one, and onto. For polyhedral surfaces there is an alternative definition. Note that there are many ways that a triangle can be subdivided into smaller triangles; a few such subdivisions are illustrated in Figure 4.3. Triangulations of surfaces can similarly be subdivided so as to yield finer triangulations.

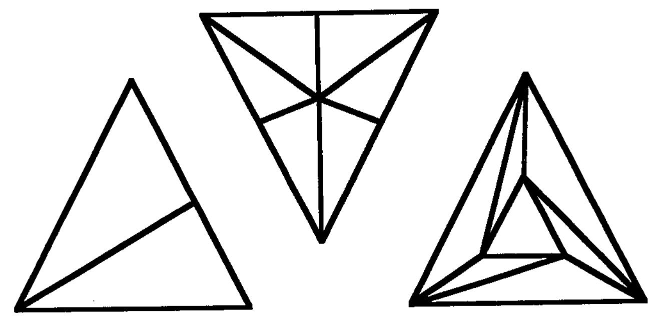


Figure 4.3

□ **DEFINITION.** *Polyhedral surfaces are called homeomorphic if, after some subdivision of the triangulations of each, there is a bijection between their vertices such that when three vertices in one surface bound a triangle the corresponding three vertices in the second surface also bound a triangle.*

Determining whether or not two surfaces are homeomorphic can be difficult. It might first come as a sur-

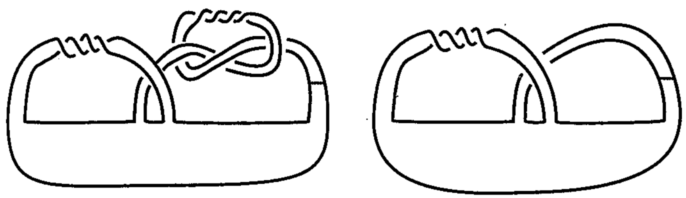


Figure 4.4

prise that the surfaces illustrated in Figure 4.4 are homeomorphic. (In the illustrations surfaces will usually not be shaded any more.)

A homeomorphism from one to the other is given by the map that cuts the first along the dotted line, unknots and untwists the band, and then reattaches it. The map is easily seen to be one-to-one and onto. Continuity follows from the fact that points that are close together on the original band are mapped to close points on the image band. Notice that this homeomorphism does not preserve the knot type of the boundary! In a case such as this it would be extremely complicated to write the map down explicitly in terms of coordinates. Triangulating the surfaces and finding the bijection would be completely unmanageable. In the next section tools are developed that greatly simplify the use of surfaces.

EXERCISES

1.1. Show that the boundary of the surface illustrated in Figure 4.1 is the trefoil knot.

1.2. The surface in 4.1 is homeomorphic to the same surface with the bands untwisted. Why? By comparing their boundaries, show that the surface with its bands twisted cannot be deformed into the one with untwisted bands.

1.3. Given a knot diagram, it is possible to construct a surface by "checkerboarding" the plane. Figure 4.5 shows this for two diagrams of the trefoil. Each surface was constructed by darkening in alternate regions of the plane determined by the knot projection. The first surface in 4.5 is nonorientable. (If you start on the top of the surface and travel around it once, you have gone through three twists, and hence finish on the other side.) The other surface is orientable. Redraw it using two colors to distinguish the two sides. Which of the diagrams for knots of 7 or fewer crossings in the Appendix result in orientable surfaces when checkerboarded?

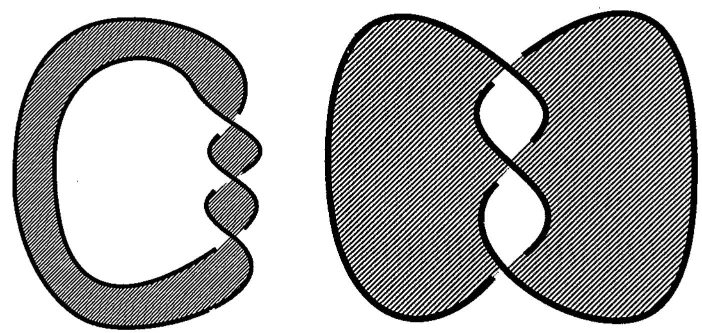


Figure 4.5

2 The Classification of Surfaces

Several connected orientable surfaces without boundary are illustrated in Figure 4.6. Associated to these surfaces is an integer called the genus of the surface, which roughly counts the number of holes. It turns out that for *any* oriented surface there is an associated number called the genus.

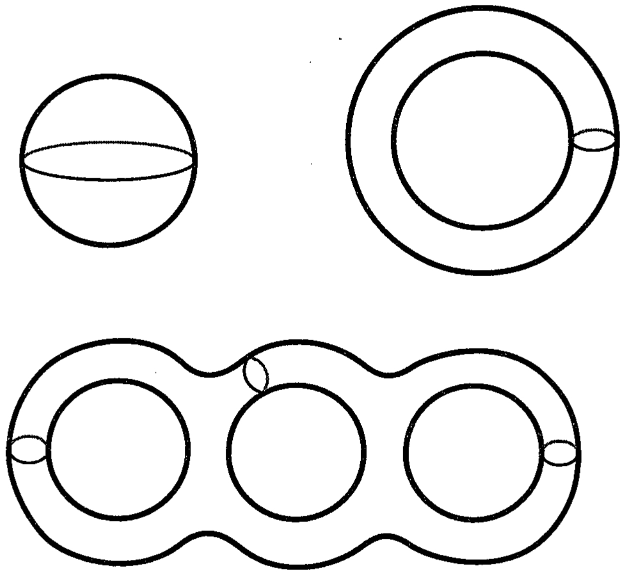


Figure 4.6

A theorem, called *the classification of surfaces*, implies that connected oriented surfaces *without boundary* are homeomorphic if and only if the they have the same genus. (Recall once again that homeomorphic surfaces need not be deformable into each other in 3-space.) A more general classification of surfaces applies to surfaces with boundary.

EULER CHARACTERISTIC AND GENUS

The Euler characteristic is an easily defined invariant of a polyhedral surface. Its definition is stated in terms of

a specific triangulation, and a basic result, usually proved using algebraic topology, says that its value is independent of choice of triangulations. Consequently, the Euler characteristics of homeomorphic surfaces are equal. The Euler characteristic and genus are difficult to compute from the definitions alone. The following results greatly simplify their calculation.

□ **DEFINITION.** *If a polyhedral surface S is triangulated with F triangles, and there are a total of E edges and V vertices in the triangulation, then the Euler characteristic is given by* $\chi(S) = F - E + V$.

For example, in the octahedron illustrated below, there are 8 faces, 12 edges, and 6 vertices. Therefore its Euler characteristic is $8 - 12 + 6 = 2$.

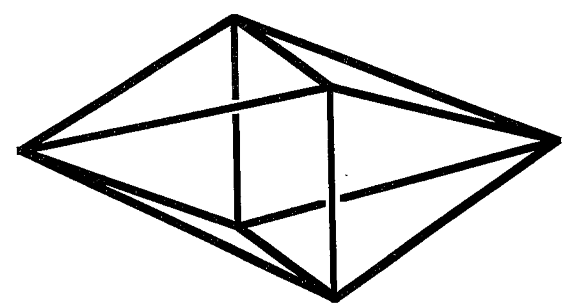


Figure 4.7

The genus of a surface is defined in terms of its Euler characteristic. Initially, the definition appears to introduce unnecessary algebra, but many simplifications will derive from it.

□ **DEFINITION.** *The genus of a connected orientable surface S is given by*

$$g(S) = \frac{2 - \chi(S) - B}{2},$$

where B is the number of boundary components of the surface.

□ **THEOREM 1.** *If two surfaces intersect in a collection of arcs contained in their boundary, the Euler characteristic of the union is the sum of their individual Euler characteristics minus the number of arcs of intersection.*

PROOF
The basic idea of the proof is simple. Suppose that each arc of intersection is a single edge of a triangle on each surface. Then the triangulations of the surfaces piece together to give a triangulation of the union. The count that is used to compute the Euler characteristic of each surface separately gets a contribution of 1 from each edge of intersection (-1 for the edge, and $+2$ for its endpoints.) Hence for the sum of the two Euler characteristics there is a contribution of $+2$ from each edge of intersection. However, in the union there is a contribution of only $+1$ from each edge. The result follows.

If each arc is not a single edge of a triangle, it can be arranged to be the union of edges, after subdividing. Again it turns out that the contribution of each arc toward the total Euler characteristic is $+1$, and the rest of the argument is the same. □

EXAMPLE
Many of the surfaces that arise are formed as disks with twisted bands added. (See Figures 4.1 and 4.4.) As the Euler characteristic of a disk is 1 (compute it for a single triangle), and a band is just an elongated disk, the Euler characteristic of a single disk with bands added is

$$(1 + \#(\text{bands})) - 2(\#(\text{bands})) = 1 - \#\text{bands}.$$

(Each band contributes two arcs of intersection.) If the surface is formed by adding bands to a collection of disjoint disks, the resulting surface has Euler characteristic $(\#\text{disks}) - (\#\text{bands})$.

□ **COROLLARY 2.** *If two connected orientable surfaces intersect in a single arc contained in each of their boundaries, the genus of the union of the two surfaces is the sum of the genus of each.*

PROOF
Express the Euler characteristic in terms of the genus and apply Theorem 1. Note that one boundary component is lost in forming the union. Exercise 2.3 asks for the details. □

Theorem 3 follows from a calculation similar to that of Theorem 1:

□ **THEOREM 3.** *If a connected orientable surface is formed by attaching bands to a collection of disks, then the genus of the resulting surface is given by*

$$(2 - \#\text{disks} + \#\text{bands} - \#\text{boundary components})/2.$$

One more result of this sort will be needed later on.

□ **THEOREM 4.** *If two surfaces intersect in a collection of circles contained in the boundary of each, the Euler characteristic of their union is the sum of their Euler characteristics.*

PROOF

The argument is similar to that of Theorem 1. In computing the Euler characteristic of a surface, each boundary component contains an equal number of edges and vertices of the triangulation. Hence, it contributes 0 to the total Euler characteristic. The same is true for the union. □

CLASSIFICATION THEOREMS

In knot theory the main interest in surfaces concerns those with boundary. Hence, the statements of the classification theorems are restricted to this setting. The first part of the classification gives a family of standard models for surfaces. The second gives the homeomorphism classification of these models.

□ **THEOREM 5.** (*Classification I*) *Every connected surface with boundary is homeomorphic to a surface constructed by attaching bands to a disk.*

SKETCH OF PROOF

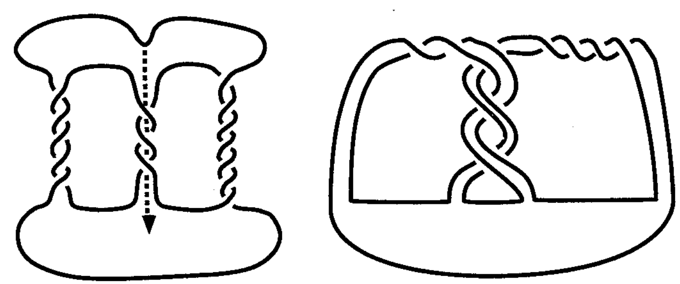
The proof of this theorem is technical, and the details appear in the references. Here is the overall idea. Fix a triangulation of the surface. A small neighborhood of each vertex forms a disk. Thin neighborhoods of the edges form bands joining the disks together. Hence, a neighborhood of the edges is homeomorphic to a union of disks with bands added. Two steps remain. The more difficult one shows that adding the faces has the same effect as not attaching certain of the bands. The other one shows that the number of disks can be reduced to one, and is detailed in the exercises. □

□ **THEOREM 6.** (*Classification II*) *Two disks with bands attached are homeomorphic if and only if the following three conditions are met:*

(1) *they have the same number of bands,*
(2) *they have the same number of boundary components,*
(3) *both are orientable or both are nonorientable.*

EXAMPLE

The surface in Figure 4.8a consists of two disks joined together by three twisted bands. The boundary is the (5,−3,7)-pretzel knot. If that surface is deformed by pushing in a narrow strip through the center band, the resulting surface can be further deformed to appear as in Figure 4.8b.



(a) (b)

Figure 4.8

EXERCISES

2.1. Use Theorem 3 to compute the genus of the surface illustrated in Figure 4.9 below.

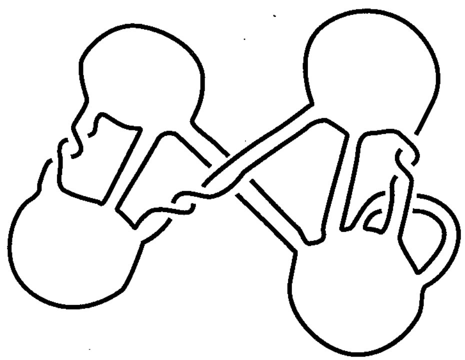


Figure 4.9

2.2. Provide the details of the proof of Theorem 3.

2.3. Prove Corollary 2.

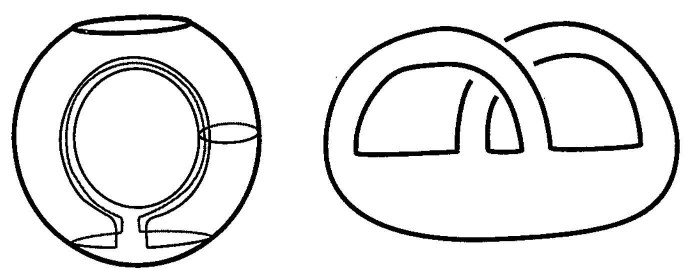
2.4. Use Theorem 5 to prove that the only genus 0 surface with a single boundary component is the disk.

2.5. Generalize the construction illustrated in Figure 4.8 to arbitrary pretzel knots. For what values of p, q, and r, is the surface orientable?

2.6. By the classification of surfaces, the punctured torus in Figure 4.10a can be deformed into a disk with bands attached. Find a deformation into the disk with two bands illustrated in Figure 4.10b. (The punctured torus has a subsurface, which is outlined. Your deformation should consist of two steps. First, deform the entire surface onto the subsurface; then, deform the subsurface to appear as the disk with bands added.)

2.7. If a surface consists of two disks with a single band joining them, it is homeomorphic to a single disk with no bands attached. Based on such an observation argue that any connected surface which is built by adding bands to a collection of disks can in fact be built starting with only one disk. (This observation is of practical importance: The surfaces that knots bound will initially be constructed from several disks. Calculations of knot invariants coming from surfaces are much easier if the surface is described using only one disk.)

2.8. Prove that the genus of a surface is nonnegative by using induction on the number of bands.



(a) (b)

Figure 4.10

2.9. Prove that the genus of an orientable surface is an integer. (Apply induction on the number of bands, and check the effect of adding an (oriented) band on the number of boundary components.

2.10. Prove that every connected orientable surface is homeomorphic to a surface of the type illustrated in Figure 4.11. (Compute the genus and number of boundary components, and then apply Theorem 6.)

Figure 4.11

3 Seifert Surfaces and the Genus of a Knot The main theorem of this section states that every knot is the boundary of an orientable surface. Consequently, geometric methods can be applied to the general study of knots and not just to particular examples.

□ **THEOREM 7.** *Every knot is the boundary of an orientable surface.*

PROOF
The proof consists of an explicit construction first described by Seifert. An orientable surface with a given knot as its boundary is now called a *Seifert surface* for the knot.

The construction begins by fixing an oriented diagram for the knot. Beginning at an arbitrary point on an arc, trace around the diagram in the direction of the orientation. Any time a crossing is met, change arcs along which you trace, but do so in such a way that the tracing continues in the direction of the knot. If at some point you start retracing your path, go to an untraced portion of the

diagram and begin tracing again. Figure 4.12 illustrates the result of this procedure for a particular knot.

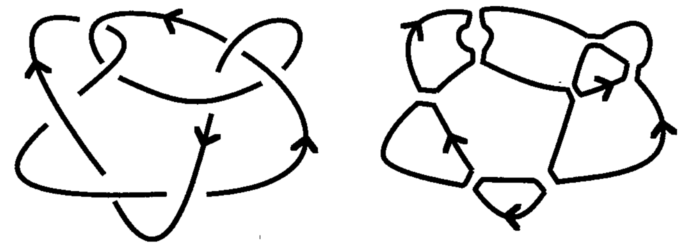


Figure 4.12

The result of this procedure is a collection of circles, called *Seifert circles*, drawn over the diagram. These circles can now be used to construct an orientable surface, as follows.

Each of the circles is the boundary of a disk lying in the plane. If any of the circles are nested, lift the inner disks above outer disks, according to the nesting.

To form the Seifert surface connect the disks together by attaching twisted bands at the points corresponding to crossing points in the original diagram. These bands should be twisted to correspond to the direction of the crossing in the knot. Figure 4.13 illustrates the final surface if this algorithm is applied to the knot in Figure 4.12.

It should be clear that the resulting surface has the original knot as its boundary, that it is orientable is not hard to prove either. (See Exercise 3.3.) Many different surfaces can have the same knot as boundary; stated differently, a knot can have many Seifert surfaces. □

□ **DEFINITION.** *The genus of a knot is the minimum possible genus of a Seifert surface for the knot.*

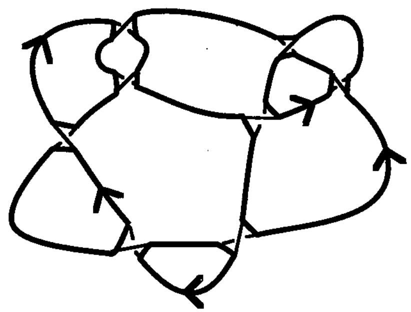


Figure 4.13

For example, Figure 4.1 shows that the trefoil bounds a surface of genus 1. On the other hand, it cannot bound a surface of genus 0, that is a disk, because then it would be unknotted, which is not the case.

A warning is called for here. It can be quite difficult to compute the genus of a knot. The genus of the surface produced by Seifert's algorithm depends on the diagram used, and, more importantly, Seifert's algorithm will not always yield the minimum genus surface! Even with this difficulty the genus is a powerful tool for studying knots.

EXERCISES

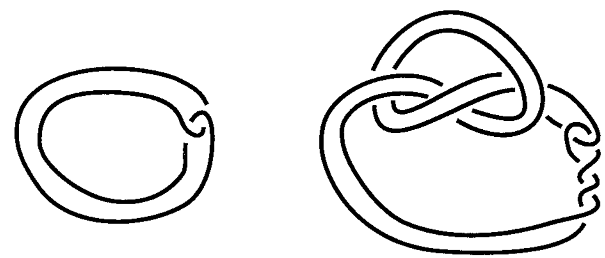
3.1. The knot in Figure 4.1 bounds a surface of genus 1, as drawn. What genus surface results if Seifert's algorithm is used to construct a Seifert surface starting with the diagram of the knot given in Figure 4.1?

3.2. Does the surface constructed by Seifert's algorithm depend on the choice of orientation of the knot? What if the procedure was used on a link instead?

3.3. Why does Seifert's algorithm always produce an orientable surface?

3.4. In applying Seifert's algorithm, a collection of Seifert circles is drawn. Express the genus of the resulting surface in terms of the number of these Seifert circles and the number of crossings in the knot diagram.

3.5. A double of a knot K is constructed by replacing K with the curve illustrated in Figure 4.14a. Figure 4.14b illustrates a double of the trefoil knot. The number of twists between the two parallel strands is arbitrary. Show that doubled knots have genus at most 1.



(a) (b)

Figure 4.14

4 Surgery on Surfaces

As discussed before,, Seifert surfaces can be very complicated. This section presents *surgery*, a method for simplifying surfaces. All the surfaces that occur later are orientable, and only that case will be described.

Underlying the constructions that follow are two observations. The first is that if two surfaces intersect along intervals, or circles, contained in their boundaries, then the union of the surfaces is again a surface. In the previous section the effect of such constructions on the Euler characteristic and genus was studied. Secondly, note that if one surface is contained in another, and the boundaries are disjoint, then removing the interior of the smaller from the other surface results in a new surface. For example, removing a disk from the interior of a surface results in a surface with one more boundary component. (This construction is sometimes called puncturing the surface.)

SURGERY

The process of cutting out pieces of a surface and pasting on other surfaces forms the basic operation of surgery. The initial set-up is the following. F is a surface in 3-space and D is a disk in 3-space. The interior of D is disjoint from F and the boundary of D lies in the interior F. This is all illustrated in Figure 4.15.

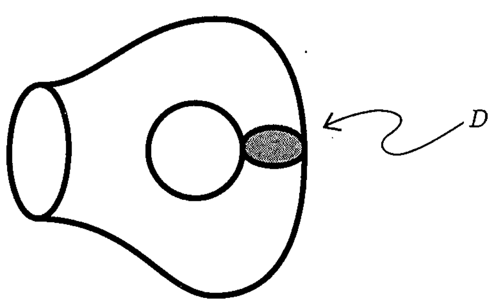


Figure 4.15

The construction of a new surface proceeds as follows. Remove a strip, or annulus, on F along the circle where

F and D meet. The new surface has two more boundary components than F. To each of these boundary components attach a disk which is parallel to the disk D. F has been transformed into a new surface by removing one annulus and adding two disks.

□ **DEFINITION.** *This procedure is referred to as performing surgery on F along D.*

The effect of surgery on the surface in Figure 4.15 is illustrated below. Note that if the boundary of D had been a different curve on F, then the surface that results from surgery might have had two components. In such cases the curve is called *separating*.

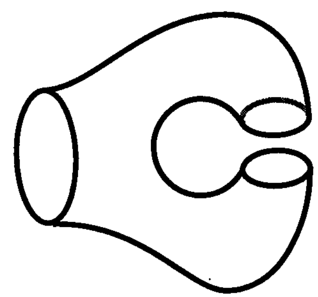


Figure 4.16

What is the effect of surgery on the genus of F? There are two cases to consider. In the first case the new surface has one component. In the second it has two.

□ **THEOREM 8.** *If surgery on a connected orientable surface, F, results in a connected surface, F', then* $\text{genus}(F') = \text{genus}(F) - 1$. *If surgery results in a surface with two components, F' and F'', then* $\text{genus}(F) = \text{genus}(F') + \text{genus}(F'')$.

PROOF

The proof proceeds by computing the effect of the two steps in surgery on the Euler characteristic of the surface. The Euler characteristic of an annulus is 0. Therefore, by Theorem 4, removing the annulus has no effect on the Euler characteristic of the surface.

The Euler characteristic of a disk is 1, so by Theorem 4 the effect of adding on the two disks is to increase the Euler characteristic by 2. Hence, the overall effect of surgery is to increase the Euler characteristic by 2. It follows from the formula for the genus of a connected surface that the genus is then decreased by 1.

In the case that the original surface F is split into two surfaces, F' and F'', the calculation is as follows. Let B, B', and B'' be the number of boundary components of F, F', and F'', respectively. Note that $B = B' + B''$. Hence:

$$\begin{aligned}\text{genus}(F') + \text{genus}(F'') &\\ &= (2-\chi(F') - B')/2 + (2 - \chi(F'') - B'')/2 \\ &= (4 - \chi(F') - \chi(F'') - B)/2 \\ &= (4 - (\chi(F) + 2) - B)/2 \\ &= \text{genus}(F).\end{aligned}$$

□

5 Connected Sums of Knots and Prime Decompositions

The connected sum of knots has already appeared in the exercises. It is now time formally to define this construction. The theory of prime knots and the prime decomposition theorem can then be presented.

Suppose that a sphere in 3-space intersects a knot, K, in exactly two points, as illustrated in Figure 4.17. This splits the knot into two arcs. The endpoints of either of those arcs can be joined by an arc lying on the sphere. Two knots, K_1 and K_2, result.

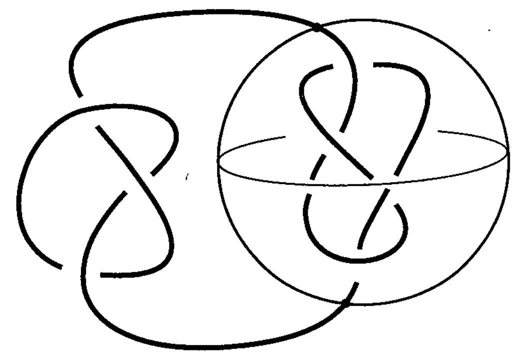


Figure 4.17

□ **DEFINITION.** *In the situation above K is called the connected sum of K_1 and K_2, denoted $K = K_1 \# K_2$.*

Given two knots, K_1 and K_2, it is easy to construct a knot K such that $K = K_1 \# K_2$. Surprisingly, K is not determined by K_1 and K_2. Examples illustrating the difficulty are hard to construct, but the nature of the problem appears with a discussion of orientation.

If the original knot K is oriented, then both K_1 and K_2 are naturally oriented. Conversely, if K_1 and K_2 are oriented knots it is possible to find a unique oriented knot K such that $K = K_1 \# K_2$ as oriented knots. To come up with a well-defined operation for which the equivalence classes of K_1 and K_2 determines the equivalence class of $K_1 \# K_2$ it is actually necessary to work with oriented knots. For instance, it can be shown that if an oriented knot K is distinct from its reverse, then the oriented connected sum of K with itself is distinct from the oriented connected sum of K with its reverse; that is, distinct even if orientations are ignored.

With connected sum carefully defined, the notion of prime knot can now be introduced, along with the prime decomposition theorem for knots.

□ **DEFINITION.** *A knot is called prime if for any decomposition as a connected sum, one of the factors is unknotted.*

□ **THEOREM 9.** (*Prime Decomposition Theorem*) *Every knot can be decomposed as the connected sum of nontrivial prime knots. If* $K = K_1 \# K_2 \# \cdots \# K_n$, *and* $K = J_1 \# J_2 \# \cdots \# J_m$, *with each* K_i *and* J_i *nontrivial prime knots, then* $m = n$, *and, after reordering, each* K_i *is equivalent to* J_i.

The proof of the existence of a prime decomposition follows immediately from the additivity of knot genus, to be proved below, using induction on the genus of the knot: if a knot decomposes as a nontrivial connected sum, then each factor has lower genus than the original knot; genus 1 knots are prime because 1 is not the sum of positive integers. The uniqueness of decompositions will not be proved here. The complete proof is similar to the proof of additivity of genus, as it involves the careful manipulation of surfaces in 3-space, in particular the families of spheres that split the knot into a connected sum. However, the argument is quite long and detailed.

□ **THEOREM 10.** (*Additivity of knot genus*) *If* $K = K_1 \# K_2$ *then* $genus(K) = genus(K_1) + genus(K_2)$.

PROOF

The proof that the genus of the connected sum is at most the sum of the genera of the summands is easy. Minimal genus Seifert surfaces for K_1 and K_2 can be pieced together along an arc to form a Seifert surface for the connected sum. By Corollary 2, the genus of that surface is the sum of the genus of each piece. It remains to show that the surface is a minimal genus Seifert surface for the connected sum.

The argument that the genus of the connected sum is at least the sum of the genera goes as follows. Figure 4.17 illustrates the connected sum of K_1 and K_2 along with a separating sphere S. Let F be a minimal genus Seifert surface for the connected sum. The surface is not drawn as there is initially no information as to how it sits in space relative to K_1, K_2, and S. It will be shown that there is a second surface, G, of the same genus as F, which can be described as the union of Seifert surfaces for K_1 and K_2, meeting in a single interval of their boundaries. It follows from Corollary 2 that the genus of G is the sum of the genera of those two surfaces, and is hence at least the sum of the minimal genera of Seifert surfaces of those knots. The approach is to work with the intersection of F and S. F intersects S in a collection of arcs and circles on S. (Initially, this might not be quite true. For instance, the intersection could contain some isolated points. However, moving F slightly will eliminate any such unexpected intersections.)

In addition, it should be clear that the only arc of intersection on S runs from the two points on S that intersect K. Now one works with the circles of intersection, using surgery to eliminate them one by one.

Consider an innermost circle of intersection. That is, pick one of the circles on S that bounds a disk on S containing no points of intersection of F and S in its interior. Surgery can be performed on F along this disk to construct a new surface bounded by K. If the new surface is connected, then it is a Seifert surface for K, which, by Theorem 8, has lower genus than did F, contradicting the minimality assumption on the genus of F. Hence, surgery results in a disconnected surface. Remove the component that does not contain K. The remaining surface has genus less than or equal to that of F (Theorem 8 again), and by the minimality assumption it actually has the same genus as F. In addition, this new surface will have fewer circles of intersection with S; the circle along which the surgery was done is no longer on the surface.

Repeating this construction, a surface G results that meets S only in an arc. Hence G is formed as the union of Seifert surfaces for K_1 and K_2 that intersect in a single arc, as desired.

This argument is often referred to as a *cut-and-paste* argument, because it consists of cutting out portions of the surface and pasting in new pieces of the surface. Another name for this type of geometric construction is an *innermost circle* argument. This type of argument is typical of geometric proofs in knot theory, and in geometric topology. □

As described earlier, the existence of prime decompositions follows from the additivity of knot genus; as a knot is decomposed as a connected sum, the genus of the factors decreases. The uniqueness follows from a much more careful cut-and-paste, innermost circle proof. The additivity of genus has the following immediate consequence.

□ **COROLLARY 11.** *If K is nontrivial, there does not exist a knot J such that $K\#J$ is trivial.*

EXERCISES

5.1. Give a proof of the final corollary.

5.2. Use the connected sum of 3 distinct knots to find an example of a knot which can be decomposed as a connected sum in two different ways.

5.3. Prove that a genus n knot is the connected sum of at most n nontrivial knots.

5.4. Fill in the details of the proof of the existence of prime decompositions using the additivity of genus.

5.5. Use the genus to give a simple proof that there are an infinite number of distinct knots. As a much harder problem, can you find an infinite number of distinct prime knots? (Later, once more efficient means are developed to compute Alexander polynomials, this too will become a simple exercise.)