Existence and Uniqueness of Knot Factorizations

Rasmus Hedegaard

Department of Mathematical Sciences
University of Copenhagen

MASTER'S PROJECT
Submitted *1st September, 2009*

Supervisor: Nathalie Wahl
Department of Mathematical Sciences
University of Copenhagen

Foreword

The purpose of this project is to study prime knots. In chapter 1, we start out by a quick review of the basic properties of knots, but as the main focus of the project lies on chapter 2, we will not prove anything in chapter 1, as doing so would make the report exceed its required bounds.
In chapter 2, we show that for any given knot, it is possible to construct a compact, connected, and orientable surface, whose boundary is that knot. We call such a surface a *Seifert surface* for the knot. Seifert surfaces are important in that the minimal genus of a Seifert surface for a knot turns out to be an invariant of that knot, and as such, we define the genus of a knot to be this minimum. Using knot genus, we prove that any knot can be factored as a sum of prime knots, which immediately begs the question if this factorization is unique. The last part of the report consists mostly of a quite long, but interesting proof, which answer this question in the affirmative.

Forord

Formålet med dette projekt er at studere primknuder. I kapitel 1 starter vi med en hurtig gennemgang af knuders elementære egenskaber, men da projektets fokus ligger i kapitel 2, vil vi ikke føre beviser i kapitel 1, da dette ville sprænge rapportens påkrævede rammer.
I kapitel 2 viser vi, at det for enhver givet knude er muligt at konstruere en kompakt, sammenhængende og orienterbar flade, hvis rand er den knude. Sådan en flade kaldes en *Seifert-flade* for knuden. Seifert-flader er vigtige i og med, at den mindste genus for en Seifert flade for en knude viser sig at være en invariant for knuden, og med grobund i dette definerer vi genus for en knude til at være dette minimum. Ved brug af genus for knuder viser vi, at en knude kan faktoriseres som en sum af primknuder, hvilket umiddelbart giver anledning til spørgsmålet, om denne faktorisering også er entydig. Sidste del af rapporten består primært af et langt, men ganske interessant bevis, som giver et bekræftende svar på dette spørgsmål.

Contents

Chapter 1

The knot concept and equivalence of knots

This chapter introduces the basic concepts of knot theory. As the main point of focus of this project is factorization of knots into prime components (all of which will be defined later), the results of this chapter are meant to be a survey of the results necessary for knot factorization, and as such, these introductory results will be stated without proof. The interested reader may find missing proofs or relevant references in [2, chapter 1] (unless otherwise mentioned).

1.1 The definition of knots

Informally, one can think of a knot as a piece of string that has been knotted somehow, and whose ends have been welded together. As such, it seems natural to regard knots as simple closed curves in the Euclidean space $\mathbb{R}^3$, but we choose instead to work with knots in the 3-sphere S^3. The symmetry of S^3 gives us certain advantages, such as the Schönflies theorems to be mentioned later, but as S^3 is the one-point compactification of $\mathbb{R}^3$, and as any knot in S^3 misses at least one point of S^3, most facts about knots in S^3 hold also in $\mathbb{R}^3$.

Definition 1. *A* knot *K in S^3 is an embedding $S^1 \to S^3$ of the 1-sphere into S^3. More generally, a* link *is an embedding of some finite collection of disjoint 1-sphere into S^3.*

The class of knots is thus contained in the class of links. However, we will mostly occupy ourselves with the notion of knots.

Certainly we need some notion of equivalence between knots, as to resemble our intuitive notion of when one knot can be changed (without cutting it, or without the knot passing through itself) to look like another. It turns out that we need something closely related to the following concept of isotopy. Let $I = [0, 1]$ denote the unit interval.

Definition 2. *Let X and Y be topological spaces. Two embeddings $f_0, f_1\colon X \to Y$ are said to be* isotopic*, if there exists an embedding*

$$F\colon X \times I \to Y \times I,$$

such that $F(x,t) = (f(x,t),t)$ for some map $f\colon X \times I \to Y$ with the property that $f(x,0) = f_0(x)$ and $f(x,1) = f_1(x)$. The map F is called a (level-preserving) isotopy *connecting f_0 and f_1.*

Being isotopic is indeed an equivalence relation, as one may check in the usual way. We will also use the simpler notation $f_t(x) = f(x,t)$, which, by a stroke of luck, is consistent with the boundary conditions of the definition. This concept of isotopy is, however, not what we need to describe equivalence of knots, as any part of the knot, where knotting takes place, may be shrunk continuously to a point, thus removing whatever special trait the knot had, see figure 1.1.

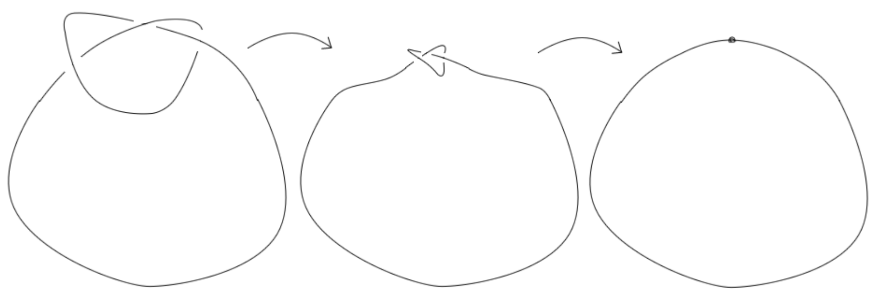

Figure 1.1: Any complications vanish at a point.

The failure of isotopies to serve as an appropriate notion of knot equivalence is of course due to the importance of *how* the knots are embedded in the ambient space, in our case S^3 . As such, we need a notion of isotopy, which takes the ambient space into consideration. In particular, we adjust the previous definition as follows.

Definition 3. *Let X and Y be topological spaces. Two embeddings $f_0, f_1\colon X \to Y$ are said to be* ambient isotopic*, if there is an isotopy (as in the previous definition)*

$$H\colon Y \times I \to Y \times I,$$

such that $H(y,t) = (h_t(y),t)$, with $f_1 = h_1 f_0$ and $h_0 = id_Y$.

Notice that when the maps f_0 and f_1 are ambient isotopic, they are also isotopic via the isotopy $F(x,t) = (h_t f_0(x),t)$. Notice also that while an isotopy is only required to move $f_0(X)$ to $f_1(X)$ without regard for the ambient space, an ambient isotopy is required to move the surrounding space along with $f_0(X)$.

Definition 4. *Two knots K_1 and K_2 are said to be* equivalent*, if they are ambient isotopic.*

From now on, we are going to abuse terminology, so whenever we use the word “knot”, we refer either to a single embedding of the 1-sphere into S^3, to the equivalence class of a knot, to the image in S^3

of a single embedding of the 1-sphere, or to an equivalence class of such images. Whichever one we mean is either going to be clear from the context or irrelevant.

It is clear that if K_1 and K_2 are equivalent knots, say through an ambient isotopy H, then the restriction $h_1\colon Y \setminus f_0(X) \to Y \setminus f_1(X)$ of the homeomorphism $h_1\colon Y \to Y$ is itself a homeomorphism (this is not the case for the more naive concept of isotopy). In other words, equivalent knots have homeomorphic complements, whence we already have our first invariant of knots: Given two knots, we can attempt to check if their complements are distinct, for example by using the fundamental group, because if they are, the knots are also distinct. In fact, knot complements are an even better invariant. As cited in [5, chapter 1], it is a result of Gordon and Luecke that if two knot complements are homeomorphic, then the corresponding knots are equivalent.

Definition 3 gives rise to unwanted pathology, at least when compared to the idea of a knot as a knotted piece of string. The knot of figure 1.2 is an example of this. It consists of an infinite number of similarly knotted sections that get smaller and smaller and ultimately converge to a point L. The pathology of this is, the weird appearance aside, that any proper subarc of this knot, which contains the limit point L, has a non-simply connected complement, in contrast to the intuitive idea of a knot.

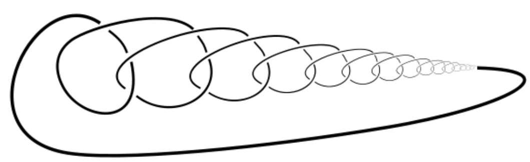

Figure 1.2: A wild knot.

We wish to eliminate such pathology, and work only with knots that fit our intuitive understanding. Regard S^3 as the boundary of the standard 4-simplex, and give S^3 the triangulation corresponding to that of the boundary of this 4-simplex. This divides S^3 into five 3-simplices, and in terms of barycentric coordinates, this gives us a notion of piecewise linearity on S^3, as we may use finer triangulations of S^3 at will.[1]

Definition 5. *A* tame knot *is a knot that is equivalent to a simple closed polygon in S^3. A knot, which is not tame, is called* wild*.*

Certainly any proper subarc of a tame knot has a simply-connected complement, as any such subarc is a succession of straight line segments, so tame knots certainly do not give rise to pathology as above. From now on, we work only with tame knots, and henceforth, whenever we use the word ‘knot”, we automatically mean a tame knot. Also, as we work in the realm of the triangulated S^3, we may from now on assume that any submanifold of S^3 is piecewise linear, i.e. that any submanifold is a subsimplex of S^3, see [4, chapter 1]. In the same spirit, we redefine our concept of knot equivalence, and agree that two knots are equivalent, if the isotopy of definition 3 is piecewise linear – in fact,

[1] Alternatively, we may regard S^3 as the one-point compactification of $\mathbb{R}^3$, whence S^3 inherits the affine structure of $\mathbb{R}^3$, but our approach gives an immediate triangulation of S^3, and opens the door to the world of piecewise linear topology.

from now on, any map is tacitly assumed to be piecewise linear. As this is important, we emphasize it as a standing assumption.

Standing assumption. *All submanifolds of S^3 are assumed to be subsimplices of S^3, and all maps are assumed to be piecewise linear.*

When we draw knots, we will for aesthetically reasons still draw them smoothly, and if anyone has a problem with this, one can think of a knot as consisting of so many small line segments that it is impossible to distinguish a drawing of the knot from a corresponding smooth drawing.

1.2 Other equivalences of knots

We have already defined two knots to be equivalent, if there is an ambient isotopy connecting them. As it turns out, there are other useful equivalent notions of knot equivalence, among which we will introduce two. One of the two requires the following preparatory definitions.

Definition 6. *Let K denote a knot, and let u be a straight line segment of K. Let Δ denote a triangle in S^3 that has u as an edge, and let v and w denote the two other edges of Δ, so that $\partial\Delta = u \cup v \cup w$. If $\Delta \cap K = u$, then the set defined by $K' = (K \setminus u) \cup v \cup w$ is another knot, and we say that K' is obtained from K by a* Δ-move *(which is to be read "*a triangle move*"), see figure 1.3. The reverse move is denoted by Δ^{-1}.*

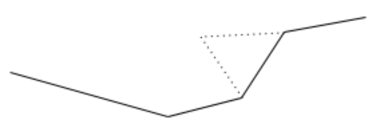

Figure 1.3: Illustrating a triangle move (and its reverse).

Definition 7. *Two knots K and L are said to be* combinatorially equivalent*, if one can be obtained from the other by a finite sequence of Δ- and Δ^{-1}-moves.*

The main theorem of this section is the following.

Theorem 1. *Let K and L be two knots in S^3. Then the following are equivalent.*
(1) There is an orientation-preserving homeomorphism $f\colon S^3 \to S^3$, such that $f(K) = L$.
(2) K and L are equivalent (via an ambient isotopy).
(3) K and L are combinatorially equivalent.

We shall later make use of this theorem: In some cases the more down-to-earth notion of combinatorially equivalence of (3) is preferred over the more technical ambient isotopy, and in other cases, it is convenient to use the homeomorphism of (1) because of its ability to preserve almost every interesting property. In fact, the theorem holds also for links, which we will make use of when introducing diagrams.

1.3 Knot sum and prime knots

We define now the most simple of all knots.

Definition 8. *The* unknot *is the knot that is the boundary of a disc in* S^3.

Two different 2-simplices in S^3 are ambient isotopic. Given a disc in S^3, we may use triangle moves along the boundary of the disc to collapse 2-simplices until only one is left. The boundary of this 2-simplex is thus equivalent to the boundary of the original disc. We see that the unknot is well-defined.

We can now define a binary sum operation on the set of knots, using the definition of [5]. We will need the concept of a *ball-arc pair*, which is simply a 3-ball B containing an arc α that meets ∂B only at its end-points. We say that the pair (B, α) is *trivial*, if it is pairwise homeomorphic to $(D \times I, \star \times I)$, where $\star$ is a point in the interior of the disc D. Recall that the 3-sphere S^3 can be obtained by gluing together two 3-balls along their boundary spheres.

Definition 9. *Let K_1 and K_2 be two knots. We define the knot $K_1 + K_2$ in the following fashion. Consider K_1 and K_2 as being in two different copies of S^3. From each copy of S^3, remove a 3-ball B whose boundary intersects the knot transversely at two points, and such that the ball-arc pair $(B, B \cap K)$ is trivial, and identify together the boundaries of the remaining balls such that their intersections with the knots match up. See figure 1.4.*

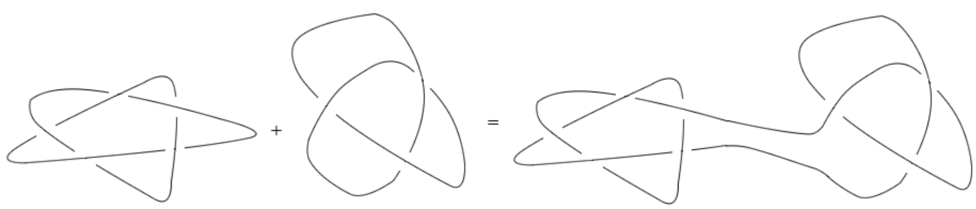

Figure 1.4: The sum of two knots.

It is not hard to imagine that knot addition is well-defined, as we may shrink, for example, the second knot, and drag this minuscule copy of the knot along the first knot until we find an appropriate place to reenlarge it. A quite similar argument shows commutativity of the sum operation, and it is also not hard to see that the operation is associative. To prove this rigorously, one would use the tools of piecewise linear topology, but we will not do so here.

As a sort of reverse process to knot sum, we emphasize also how to check if a given knot is a sum of two "lesser" knots, so let K be some knot. Suppose that Σ is some 2-sphere, which meets K transversely at two points. Then (by the Schönflies theorem, to appear in the very first part of chapter 2) Σ separates S^3 into two components B_1 and B_2, each of whose closure is a 3-ball. Each of B_1 and B_2 contains part of the knot K, say the arcs α_1 and α_2 respectively. By removing the pair (B_2, α_2) from S^3, and by attaching a trivial ball-arc pair in its place, we complete the arc α_1 to

obtain some knot K_1, and by a similar process we complete α_2 to obtain some knot K_2. Now we have that K_1 and K_2 live in two distinct copies of S^3, and we have in fact that $K = K_1 + K_2$, as one can easily see by applying definition 9: Remove from each of K_1 and K_2 the very same trivial ball-arc pairs that we attached to α_1 and α_2 to obtain K_1 and K_2. We say that the sphere Σ separates K_1 from K_2.
Certainly the unknot is a neutral element with respect to knot addition. We define prime knots in the obvious way as follows.

Definition 10. *A* prime knot *is a knot that cannot be written as a sum of two non-trivial knots.*

We will be a lot more concerned with prime knots in the next chapter.

1.4 Knot diagrams and Reidemeister moves

When drawing a picture of a knot, we will very often draw a picture of a so-called diagram for the knot instead of drawing the knot itself. The knots of figure 1.4 are examples of diagrams. Formally, a diagram for a link L is a subset of S^2 depicting the link in the following way. Let S^2 be some equatorial 2-sphere in S^3, for example the one determined by the first three coordinates of $\mathbb{R}^4$ in a standard embedding $S^3 \to \mathbb{R}^4$. Using Δ- and Δ^{-1}-moves, we may change the link so that it is in general position with respect to the standard projection $p\colon S^3 \to S^2$, which means that each line segment of L projects to a line segment of S^2 (i.e. no line segment maps to a point), that the projection of any two line segments of L have at most one point in common, which for disjoint line segments of L is not an end point, and that any point of S^2 belongs to at most two projected line segments of L.
So far the image $p(L)$ gives no information as to where the different strands of the links are positioned in space. Defining a crossing point of $p(L)$ to be a point, which is not an end-point, and which belongs to two projected line segments, we assign a label to each crossing point P telling us, which of the corresponding line segments of L is above the other. Letting $\mathcal{L}$ be the set of such labels, we call $(p(L), \mathcal{L})$ a *diagram* for the link. In a drawing, we illustrate at each crossing of $p(L)$ which strand is above the other, as in figure 1.4.
Instead of trying to transform a knot into an equivalent knot using one of the three equivalent methods of theorem 1, one can instead work with a diagram of a knot and try to transform this into another equivalent diagram, for an appropriate notion of "equivalent". Certainly one would like a pair of knots to be equivalent, if and only if their corresponding diagrams are equivalent, but first of all we define equivalence of diagrams.

Definition 11. *Two link diagrams are said to be* equivalent*, if and only if they are related by a finite sequence of the following so-called* Reidemeister moves *or their reverses, and an orientation-preserving homeomorphism of S^2.*

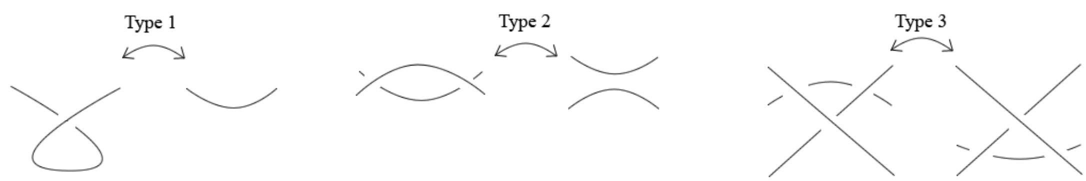

Figure 1.5: The three types of Reidemeister moves.

Thus two diagrams, which are related by just one Reidemeister move, are the same except in the one place where the change took place.
With this notion of equivalence of diagrams, we obtain our goal mentioned above, which for completeness is made explicit in the following proposition.

Theorem 2. *Two links are equivalent, if and only if their diagrams are equivalent.*

1.5 Linking numbers

As the final topic of this chapter, we introduce linking numbers, which will play a prominent role in the upcoming main chapter. We follow [5, page 11]. The linking number is a measure of "how linked" two components of a link are. In particular, suppose we are given a diagram D for a link L. Give each component of the link some orientation. This induces an orientation on the diagram D. We assign to each crossing of the diagram an integer in the set $\{+1, -1\}$ as outlined in figure 1.6. In particular, if at a crossing, one strand passes above the other in the manner of a right-hand-screw, the crossing attains the integer $+1$, and otherwise it attains the integer -1. Notice that when assigning this integer, we use not only the orientation of the link, but also the orientation of S^2.

Figure 1.6: Assigning either $+1$ or -1 to each crossing.

We define the linking number as follows.

Definition 12. *Let L be an oriented link with a diagram D, and let L_1 and L_2 be two different component of L. Then the* linking number *$lk(L_1, L_2)$ of L_1 and L_2 is half the sum of the integers assigned to the crossings, for which one strand is from L_1 and the other is from L_2.*

It is not hard to see that this definition is independent of the diagram D. In particular, as equivalent diagrams are related by a sequence of Reidemeister moves (and an orientation-preserving homeomorphism of S^2), it suffices to check that each Reidemeister move does not change the linking number, which is immediate by inspection (we need not worry about moves of type 1, as such moves affect only one component of L). Also, an orientation-preserving homeomorphism of S^2 certainly preserves all crossings of D. Notice that if we change the orientation of only one of the two component L_1 or L_2, then the linking number changes sign. Hence the linking number remains the same, if we change the orientation of both L_1 and L_2.

Chapter 2

Factorizations of knots

Unless otherwise mentioned, this chapter follows the work [5, chapter 2], but compared to this work, lots of details have been supplied. At some points in this chapter, we use the two-dimensional Schönflies theorem, which states that any embedding of S^1 into S^2 separates S^2 into two components, each of which has a disc D^2 as its closure. Later on, we use the three-dimensional analogue of this result, which we state here as a theorem.

Theorem 3 (Three-dimensional Schönflies)**.** *Let $\varphi\colon S^2 \to S^3$ be a piecewise linear embedding. Then $S^3 \setminus \varphi(S^2)$ has two components, each of whose closure is a piecewise linear ball B^3.*

The piecewise linear condition cannot be neglected. The infamous Alexander's horned sphere is an example of an embedding of S^2 into S^3 with the property that the exterior of the sphere is not even simply-connected, see e.g. [3, section 2.B].

2.1 Seifert's algorithm

As for the integers with the usual product, any knot can be written uniquely as a sum of prime knots, ignoring, of course, the order of the summands. In the course of proving this, we shall move up in dimension, and instead of studying knots themselves, we shall study surfaces that are intimately connected to the knots. Essentially we are only going to use the following definition in the case of knots, but as talking about the broader class of links only gives rise to a very minor change in the proof of theorem 4, we are going to do so.

Definition 13. *A* Seifert surface *for a given link L is a connected, compact, and orientable surface, whose boundary is the link L.*

Certainly the unknot has a Seifert surface, namely a disc. It may be unclear what other links have a Seifert surface, but we have in fact the following surprising theorem.

Theorem 4. *Any link L has a Seifert surface.*

Proof. Let D be a diagram for the link L. Give each component of L an orientation, and let the diagram D inherit these orientations. We consider each crossing of D, and change the diagram as shown in figure 2.1.

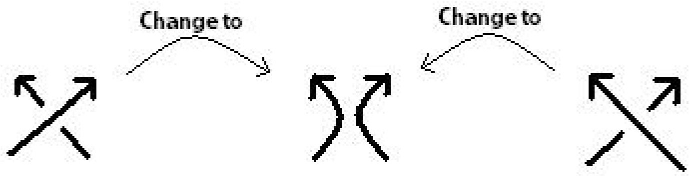

Figure 2.1: Eliminating all crossings of D.

Specifically, we eliminate each crossing of the diagram D in the only manner possible such that the orientations remain consistent. The resulting diagram $\widehat{D}$ is then a disjoint union of oriented simple closed curves, as illustrated in figure 2.2, which shows the method applied to the so-called figure-eight knot.

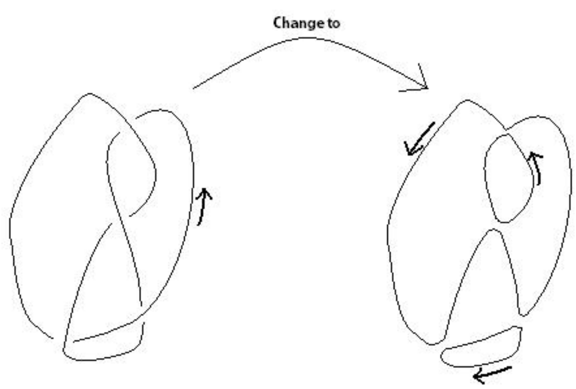

Figure 2.2: A diagram D for the figure-eight knot and the corresponding diagram $\widehat{D}$.

The disjoint circles of $\widehat{D}$ are called *Seifert circles* for the diagram D. Note that $\widehat{D}$ locally looks the same as the diagram D, except at the finite number of crossings of D.
Each Seifert circle bounds a disc. Should two or more circles happen to be nested, we consider the corresponding discs to be at different heights in S^3, in order to maintain disjointness. The union of these discs constitute a (most likely disconnected) surface $\widehat{\mathcal{S}}$ in S^3, and we wish to alter this surface to obtain another surface $\mathcal{S}$, whose boundary is the link L. A crossing of the diagram D corresponds to two discs in $\widehat{\mathcal{S}}$, namely the pair of discs that arose from removing that particular crossing (conversely, a pair of discs may correspond to more than one crossing). For such a pair of

discs, join them together at their boundaries by a small strip with a half-twist, taking care of twisting the strip such that the twist correctly mimics the corresponding crossing of the link, see figure 2.3. As shown in that figure, we position the twisted strip at the position of the corresponding crossing. Doing this for each crossing of D, we obtain a surface $\mathcal{S}$, whose boundary is the link L, as illustrated in figure 2.4. By construction, the surface $\mathcal{S}$ is compact, and we argue now that it is also orientable.

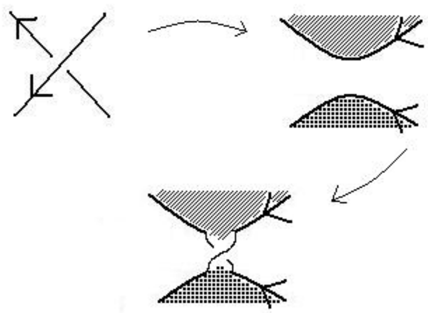

Figure 2.3: Connecting discs in $\widehat{\mathcal{S}}$ by half-twisted strips.

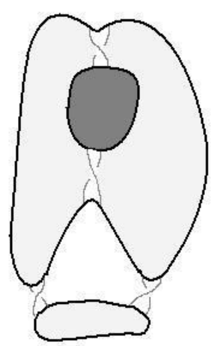

Figure 2.4: Seifert surface for the figure-eight knot. The nested discs are at different heights.

We start out by giving each disc of the surface $\widehat{\mathcal{S}}$ an orientation, and then we argue that we can attach the half-twisted strips in a consistent manner (orientation-wise), and hence obtain an orientation of the entire surface $\mathcal{S}$.

The surface $\widehat{\mathcal{S}}$ can be oriented in 2^n different ways, as we for each of the n disjoint discs can choose among two different orientations. Among these 2^n orientations of $\widehat{\mathcal{S}}$, we choose one orientation, which will be seen to induce an orientation on $\mathcal{S}$. Denote the two possible orientations for each disc of $\widehat{\mathcal{S}}$ by **a** and **b**. Recall that each disc has a Seifert circle as its boundary, and that each Seifert circle inherited an orientation from the diagram D. If the boundary of a disc has a counter-clockwise orientation, then give this disc orientation **a**, and if the boundary has a clockwise orientation, give this disc orientation **b**. We say that a pair of discs in $\widehat{\mathcal{S}}$ are *neighboring discs*, if in the surface $\mathcal{S}$, there is a twisted band connecting the two discs. We shall argue that nested neighboring discs have been given the same orientation, and that non-nested neighboring discs have been given different orientations. This is actually easy to see, because if we have a pair of nested discs as in figure 2.5 (there may be more twisted bands between them, but this is unimportant here), and if, say, the innermost disc has some orientation, say **a**, then as the twisted bands faithfully relays the orientation on the boundary, the boundary circle of the outer disc will also have a counter-clockwise orientation, whence the outer disc also has orientation **a**.

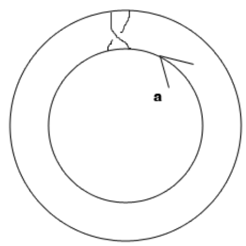

Figure 2.5: Nested neighboring discs are given the same orientation.

Notice that it does not matter which way the twisted band twists, and notice also that we could have drawn the same conclusion, if we had known the orientation of the outer disc. In the exact same manner, we can argue that non-nested neighboring discs have been given opposite orientations, but we leave the argument, and let figure 2.6 speak for itself.

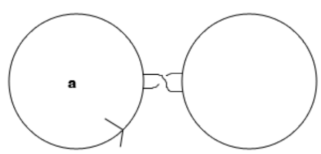

Figure 2.6: Non-nested neighboring discs have been given opposite orientations.

It is now a simple matter to check that neighboring discs have, in this way, been given consistent orientations. We argue pictorially as in figure 2.7, where we slide the letter "S" from one disc, along the twisted band, to a neighboring disc, and we see that giving nested discs the same orientation is consistent, and that giving non-nested discs opposite orientations is also consistent, see figure 2.7. As neighboring discs have been oriented consistently, we have exhibited an orientation of $\mathcal{S}$, so $\mathcal{S}$ is orientable. As we are working with links, the surface $\mathcal{S}$ may be disconnected. It this is so, let $c_1, \ldots, c_k$ denote the k components of $\mathcal{S}$. Connect component c_1 to c_2 by a long thin tube. If the orientation of $\mathcal{S}$ remains consistent along this tube, we do nothing. Otherwise, we switch the orientation of c_2. Continue inductively by connecting c_i to c_{i+1} by a long thin tube, switching the orientation of c_{i+1} if necessary. Now the resulting surface is not only compact and orientable, but also connected. □

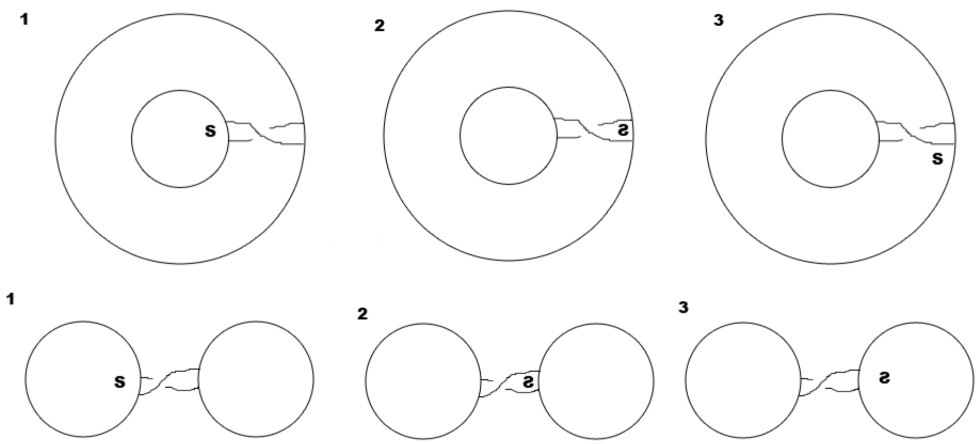

Figure 2.7: Neighboring discs have been given consistent orientations.

The process of constructing the surface $\mathcal{S}$ as described in this proof is known as *Seifert's algorithm*. Notice that if we apply the algorithm to a diagram D for a *knot* K, then regardless of the chosen orientation of D, the algorithm returns the same underlying surface $\mathcal{S}$ (i.e. neglecting the orientation of $\mathcal{S}$). This is because both strands of the crossings in figure 2.1 will change directions, as we change the orientation of the diagram D, and hence the way we alter the diagram to obtain $\widehat{D}$ is unchanged, and we ultimately end up with the same underlying surface as for original orientation of the diagram. For a link with at least two components, the resulting underlying surface is, however, dependent on how we choose an orientation for the diagram. For example we have in figure 2.8 chosen two different orientations for a given diagram of the so-called *unlink*, and we see that they give rise to different sets of Seifert circles.

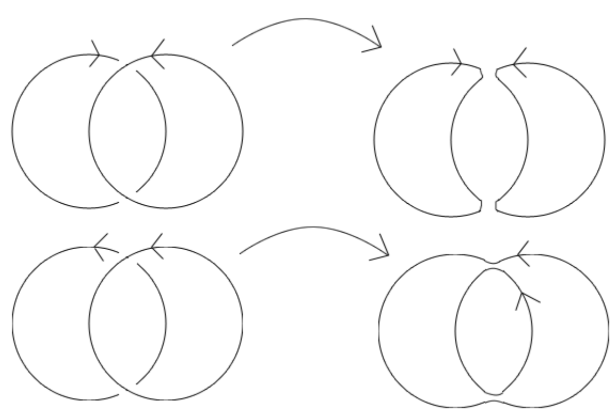

Figure 2.8: Similar diagrams for the unlink gives rise to different sets of Seifert circles.

From now on we concentrate mostly on knots. It will turn out to be profitable to consider the possible genera of Seifert surfaces for a given knot. Specifically we define the genus of a knot in terms of the possible genera of the corresponding Seifert surfaces as follows.

Definition 14. *The* genus $g(K)$ *of a knot* K *is defined by*

$$g(K) = \min\{g(\mathcal{S}) | \mathcal{S} \textit{ is a Seifert surface for } K\},$$

where $g(\mathcal{S})$ *is the genus of the surface* $\mathcal{S}$.

For example the unknot has genus 0. The genus is an invariant of knots, as we may see as follows. Let K_1 and K_2 be equivalent knots, and let F_1 be a minimal genus Seifert surface for K_1. By theorem 1, there is an orientation-preserving homeomorphism $f\colon S^3 \to S^3$, such that $f(K_1) = K_2$. But $f(F_1)$ is then a Seifert surface for K_2 of the same genus as K_1, which proves $g(K_2) \leq g(K_1)$. The other inequality follows similarly.
The outcome of Seifert's algorithm applied to a knot depends, of course, very much on which diagram we choose to represent the knot. The least we can hope for is thus that there exists *some* diagram for the knot, such that Seifert's algorithm applied to this diagram gives a minimal genus Seifert surface. As cited in [1, pp. 105–106], this turns out to be too much to hope for. In fact, letting $g_c(K)$ denote the minimal genus of a Seifert surface of a knot K obtainable from Seifert's algorithm, an Israeli mathematician named Yoav Moriah constructed in 1987 an infinite family of knots $\{K_i\}_{i\in\mathbb{N}}$, for which the difference $g_c(K_i) - g(K_i)$. We call $g_c(K)$ the *canonical genus* of the knot K. An even more concrete result was obtained nine years later by M. Kobayashi and T. Kobayashi, who constructed an infinite family of knots with arbitrarily high genus, and with the property that $g_c(K) = 2g(K)$ for every K in the family.
We finish off this section with a small remark, which is important enough to state as a proposition.

Proposition 1. *If* K *is a knot with genus* $g(K) = 0$*, then* K *is the unknot.*

Proof. Let $\mathcal{S}$ be a Seifert surface for K. As $\mathcal{S}$ is a compact, connected, and orientable surface with one boundary circle, the classification of surfaces dictates that $\mathcal{S}$ is a disc, see e.g. [6, chapter 11]. By definition 8, K is then the unknot. □

2.2 Existence of knot factorizations

Using the concept of genus of a knot, we prove in this section that any knot can be written as a sum of prime knots. This result will follow easily from the following all-important theorem.

Theorem 5. *The genus of a knot is additive, so for any two knots* K_1 *and* K_2*, we have*

$$g(K_2 + K_2) = g(K_1) + g(K_2).$$

Proof. We prove first that $g(K_1+K_2) \leq g(K_1)+g(K_2)$. Let K_1 and K_2 be situated in the same copy of S^3, and let F_1 and F_2 be Seifert surfaces for K_1 and K_2, respectively. We may assume F_1 and F_2 to be disjoint (as we may assume that the knots K_1 and K_2 are so far apart in S^3 to allow room for disjoint Seifert surfaces). Assume also that the there is some 2-sphere Σ in S^3, which separates K_1

from K_2.

We argue first that each of the surfaces F_1 and F_2 does not separate S^3, so let $F \in \{F_1, F_2\}$. As F is connected, compact, and orientable with one boundary component, F is homeomorphic to a standard surface $\widetilde{F}$ of some genus g with a small open disc removed, see [6, theorem 11.1]. Giving $\widetilde{F}$ its standard CW-structure, it is well-known that $\widetilde{F}$ is homotopy equivalent to its 1-skeleton, that is, $\widetilde{F}$ is homotopy equivalent to some graph G (for example, the torus minus a small open disc is homotopy equivalent to its 1-skeleton, a wedge of two circles). Let $\widetilde{H}$ denote reduced singular (co)homology. Applying Alexander duality ([3, theorem 3.44]), we thus obtain

$$\widetilde{H}_0(S^3 \setminus F) \simeq \widetilde{H}^2(F) \simeq \widetilde{H}^2(\widetilde{F}) \simeq \widetilde{H}^2(G) = H^2(G) = 0,$$

where the last equality $H^2(G) = 0$ for instance follows from the fact that since G is a CW-complex of dimension 1, then $H_2(G) = 0$, and since first homology groups of graphs are always free groups (being the abelianization of the fundamental group, which is a free product of k copies of $\mathbb{Z}$, where k is the number of edges of G not contained in a maximal tree in G), then $H^2(G) = \mathrm{Hom}(H_2(G), \mathbb{Z}) = \mathrm{Hom}(0, \mathbb{Z}) = 0$. The relevant references for this last argument are [3, prop 1A.2, lemma 2.34, thm. 2A.1, thm. 3.2]. As $\widetilde{H}_0(S^3 \setminus F) = 0$, the set $S^3 \setminus F$ is path-connected.

We can then via the Mayer-Vietoris sequence applied to the pair $(S^3 \setminus F_1, S^3 \setminus F_2)$ argue that the complement $S^3 \setminus (F_1 \cup F_2) = (S^3 \setminus F_1) \cap (S^3 \setminus F_2)$ is also path-connected. Indeed, we have the following portion of the sequence:

$$\cdots \to H_1(S^3) \to H_0\left(S^3 \setminus (F_1 \cup F_2)\right) \to H_0(S^3 \setminus F_1) \oplus H_0(S^3 \setminus F_2) \to H_0(S^3) \to 0.$$

Certainly $H_1(S^3) = 0$, and $H_0(S^3) \simeq \mathbb{Z}$, and we just argued that $H_0(S^3 \setminus F_1) \oplus H_0(S^3 \setminus F_2) \simeq \mathbb{Z} \oplus \mathbb{Z}$, whence the above portion of the sequence reduces to

$$0 \to H_0\left(S^3 \setminus (F_1 \cup F_2)\right) \to \mathbb{Z} \oplus \mathbb{Z} \to \mathbb{Z} \to 0.$$

This sequence is split, and so the complement $S^3 \setminus (F_1 \cup F_1)$ is path-connected, as claimed.

We may thus find a path α from some boundary point $P \in K_1 = \partial F_1$ to some boundary point $Q \in K_2 = \partial F_2$, such that except for its end-points, all of α is contained in $S^3 \setminus (F_1 \cup F_2)$. The setup is shown in figure 2.9, where on the left-hand side, we see a Seifert surface for the so-called trefoil knot, and on the right-hand side, we see a Seifert surface for the knot, which in the traditional ordering of prime knots has come to be known as 5_2, see e.g. [5, table 1.1].[1]

[1] Here the number "5" refers to the fact that the least number of crossings of a diagram for the knot is five, and we say that the knot has *crossing number* 5. The subscripted number "2" refers to the knot being the second knot in the traditional ordering with crossing number 5, where being second is an arbitrary choice. Apparently not widely spread yet, the knot 5_2 is also sometimes referred to as the 3-twist knot.

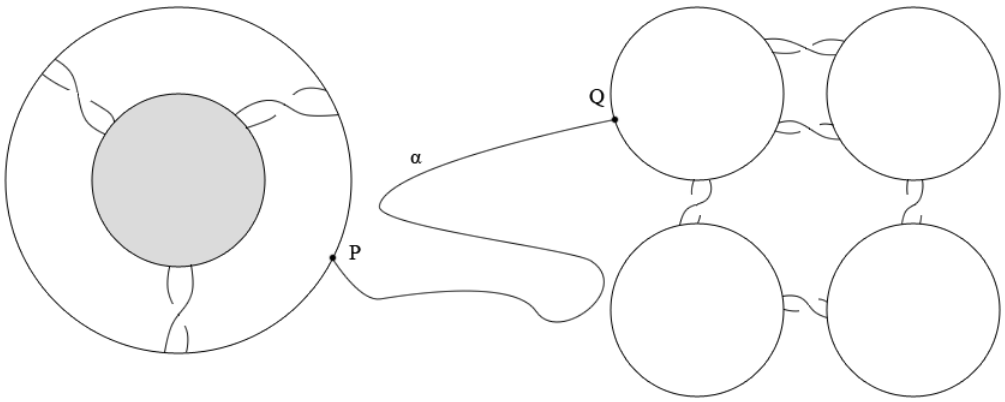

Figure 2.9: Connecting the boundaries of F_1 and F_2 by a path α.

We may also assume that the path α intersects once the 2-sphere Σ separating K_1 from K_2, because if α intersects Σ more than once, as in the first picture of figure 2.10, then we first replace the unwanted part (the dotted line in the second picture) of α inside of Σ by an arc on Σ (for example the shortest geodesic arc between the two intersection points, as shown by a thick curve in the second picture), and then we shrink Σ slightly to obtain only one point of intersection between α and Σ, as shown in the third picture of figure 2.10.

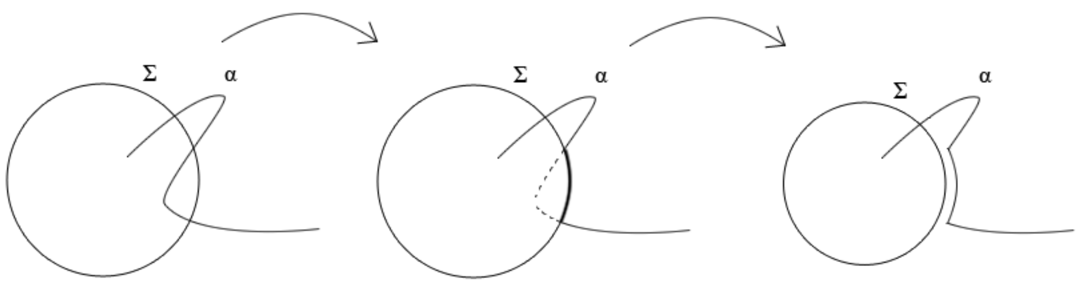

Figure 2.10: The path α intersects once a 2-sphere Σ separating K_1 from K_2

Make a thin strip around α, thin enough to intersect Σ only where α intersects Σ, and if necessary, give the strip a half-twist to match orientations of the two Seifert surfaces F_1 and F_2. The boundary of $F_1 \cup F_2$ connected by this thin strip is now the knot $K_1 + K_2$, as Σ intersects the knot transversely at two points, and as Σ is a 2-sphere separating K_1 from K_2. The surface $F_1 \cup F_2$ connected by the thin strip is then a Seifert surface for the sum $K_1 + K_2$, and its genus is clearly the sum of the genera of F_1 and F_2. As F_1 and F_2 were chosen to be minimal genus Seifert surfaces, we have thus

$$g(K_1 + K_2) \leq g(K_1) + g(K_2).$$

For the other inequality , suppose that F is a minimal genus Seifert surface for $K_1 + K_2$, and let Σ be a 2-sphere exhibiting the knot $K_1 + K_2$ as the sum of K_1 and K_2, i.e. Σ meets the knot transversely at two points, and it separates K_1 from K_2. The 2-sphere Σ separates $K_1 + K_2$ into two arcs α_1 and α_2, and for whichever arc β in Σ connecting the two points of intersection in $\Sigma \cap (K_1 + K_2)$, then $\alpha_1 \cup \beta$ and $\alpha_2 \cup \beta$ are the knots K_1 and K_2, respectively, as illustrated in figure 2.11.

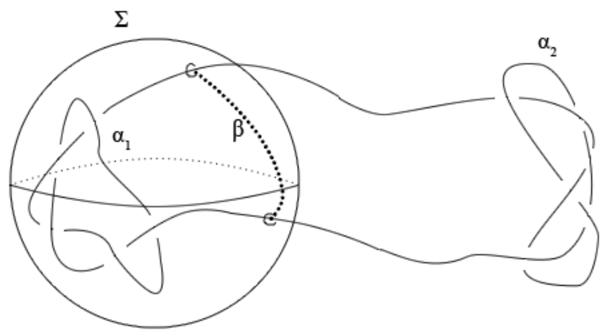

Figure 2.11: The 2-sphere Σ separates K_1 from K_2.

As described in chapter 1, we may assume that the inclusion of the two surfaces Σ and F in S^3 is piecewise linear, for example by choosing a very fine triangulation of S^3 and by perturbing Σ and F slightly, such that we may regard both as sub-complexes of S^3. As such, we may assume that Σ is in general position with respect to F, so that the intersection $F \cap \Sigma$ is a 1-dimensional manifold, see [4, chapter 1]. In fact, since F has the knot $K_1 + K_2$ as boundary, the sphere Σ intersects the boundary of F only at the two places, where the knot leaves Σ, and by general position, the intersection $F \cap \Sigma$ thus consists of some finite collection of simple closed curves together with one arc β joining the two points, where the knot $K_1 + K_2$ punctures Σ. The immediate plan is to eliminate each of these simple closed curves of intersection, and we do this by changing F.

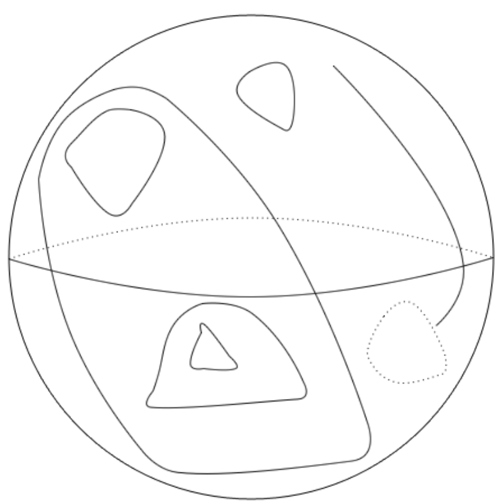

Figure 2.12: The curves of intersection between the surfaces F and Σ shown on Σ.

By the 2-dimensional Schönflies theorem, each simple closed curve of $F \cap \Sigma$ separates Σ into two disc components, only one of which contains the arc β, see figure 2.12. We agree that the *inside* of a simple closed curve C in $F \cap \Sigma$ is the component of the complement $\Sigma \setminus C$, which does not contain the arc β.

Choose such a simple closed curve C, which is innermost on Σ, meaning that its inside contains none of the other simple closed curve of $F \cap \Sigma$. We may do this, as there are only finitely many curves in $F \cap \Sigma$. We perform an elementary 1-surgery on the 2-manifold F by removing from F a small annular neighborhood of the curve C, and by replacing it by two discs. I.e. we replace a small cylinder $S^1 \times D^1$ around C by two discs $D^2 \times S^0$ (see e.g. [5, chapter 12] for the basic rules of surgery). We name the resulting surface $\widehat{F}$, see figure 2.13. Notice that by doing this first for an

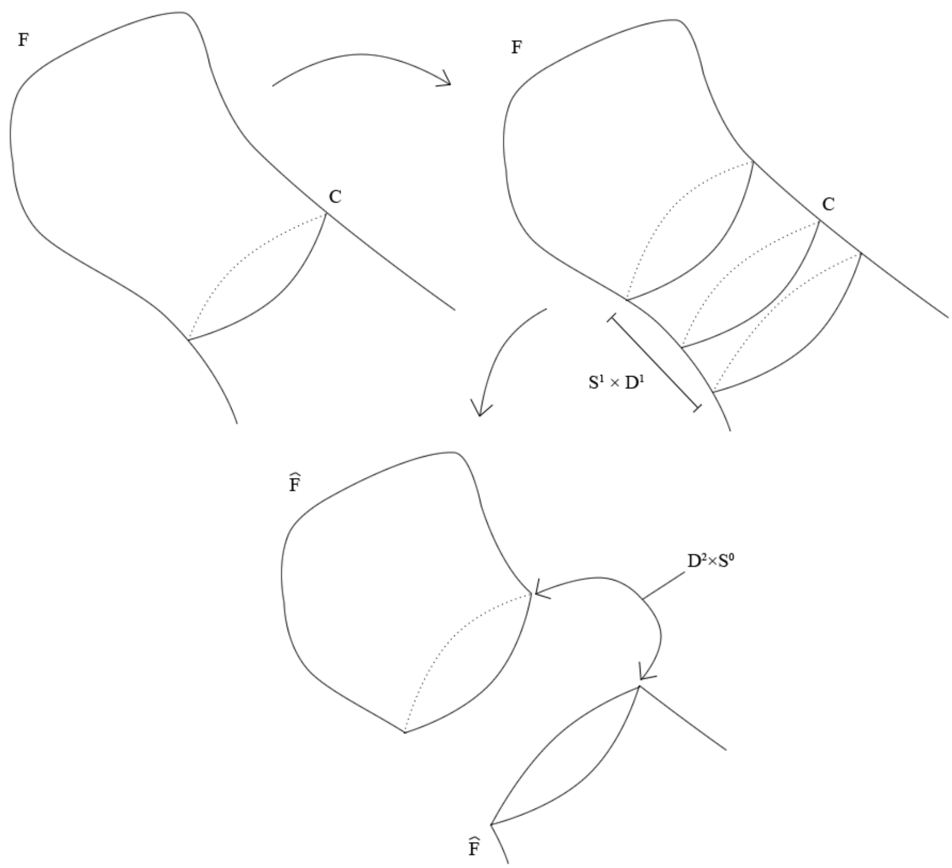

Figure 2.13: Replacing a small cylinder on F around the curve C by two discs.

innermost circle on Σ, the surgery on F will not affect the other circles in $F \cap \Sigma$: If necessary, we choose the cylinder $S^1 \times D^1$ around C so small that it does not intersect the other curves of $F \cap \Sigma$, whence its removal will not affect the other circles of $F \cap \Sigma$. This is not necessarily possible, if we had chosen some arbitrary circle to begin with.

The surface $\widehat{F}$ is disconnected, because otherwise the surgery would have had the effect of removing a hollow handle, and thus reducing the genus of F. But $\widehat{F}$ also has the knot $K_1 + K_2$ as boundary,

and as $\widehat{F}$ then also would have been connected, compact and orientable, it would have been a Seifert surface for $K_1 + K_2$ of lower genus than F, contradicting the choice of F. So $\widehat{F}$ is disconnected, and we throw away the part of $\widehat{F}$ that does not contain the knot $K_1 + K_2$, and we repeat the above process of eliminating circles in $F \cap \Sigma$, but this time with $\widehat{F}$ in place of F, and so on.
Ultimately we end up with a Seifert surface F' for $K_1 + K_2$ of the same genus as F, and with the property that the intersection $F' \cap \Sigma$ consists only of the arc β. As $\alpha_1 \cup \beta \simeq K_1$ and $\alpha_2 \cup \beta \simeq K_2$, this means that Σ separates F' into two pieces F'_1 and F'_2, which are Seifert surfaces for K_1 and K_2 respectively. As the sum of the genera of F'_1 and F'_2 equals the genus of F', and as F'_1 and F'_2 are not necessarily minimal genus Seifert surfaces for K_1 and K_2, we obtain thus

$$g(K_1) + g(K_2) \leq g(F'_1) + g(F'_2) = g(F') = g(K_1 + K_2),$$

as we set out to prove. □

This theorem immediately yields a lot of interesting corollaries.

Corollary 1. *The unknot is the only knot that has an additive inverse.*

Proof. If U denotes the unknot, and $U = K_1 + K_2$ for some knots K_1 and K_2, then $0 = g(U) = g(K_1) + g(K_2)$, whence $g(K_1) = g(K_2) = 0$. By proposition 1, this implies $K_1 = K_2 = U$. □

The following two corollaries are clear.

Corollary 2. *Let K be a non-trivial knot. Then whenever $m \neq n$ are distinct integers, we have*

$$\sum_{j=1}^{m} K \neq \sum_{j=1}^{n} K.$$

In particular summing a non-trivial knot with itself gives rise to infinitely many different knots.

Corollary 3. *Knots of genus 1 are prime.*

As a finally consequence, we prove that knots can indeed be factored into prime knots.

Theorem 6. *Any knot can be expressed as a sum of prime knots.*

Proof. We use strong induction on the genus. The unknot is the empty sum, by convention, and knots of genus 1 are prime, so let K be a knot with $g(K) > 1$, and suppose that the theorem holds true for knots of lesser genus. If K is prime, we are done. If not, factor K into a sum of two non-trivial knots, which by additivity of genus have strictly smaller genera than that of K. By induction, we can factor the two summands into prime knots, whence the theorem follows. □

2.3 Uniqueness of knot factorizations

We are now get ready to prove the key theorem of uniqueness, a theorem whose proof is rather long, but one which will almost immediately yield uniqueness of prime factorizations of knots.

Theorem 7. *Let K be a knot, and suppose that K can be expressed as the two different knot sums $K = P + Q$ and $K = K_1 + K_2$, where P is a prime knot, and Q, K_1, and K_2 are some other (not necessarily prime) knots. Then either*

$$(i)\quad K_1 = P + K_1' \text{ for some knot } K_1', \text{ and } Q = K_1' + K_2, \text{ or}$$
$$(ii)\quad K_2 = P + K_2' \text{ for some knot } K_2', \text{ and } Q = K_1 + K_2'.$$

Proof. Let Σ be a 2-sphere meeting the knot $K = K_1 + K_2$ transversely at two points, and such that the 2-sphere separates K_1 from K_2, as in the definition of knot sum. Furthermore, let B be a 3-ball, whose boundary ∂B demonstrates K as the sum $K = P + Q$. Then B intersects K in an arc α, and by replacing the complement of the pair (B, α) by a trivial ball-arc pair, the pair becomes (S^3, P). We will furthermore assume that Σ and B are in general position with respect to each other, so that the intersection $\Sigma \cap \partial B$ consists of a finite collection of disjoint simple closed curves, see figure 2.14.

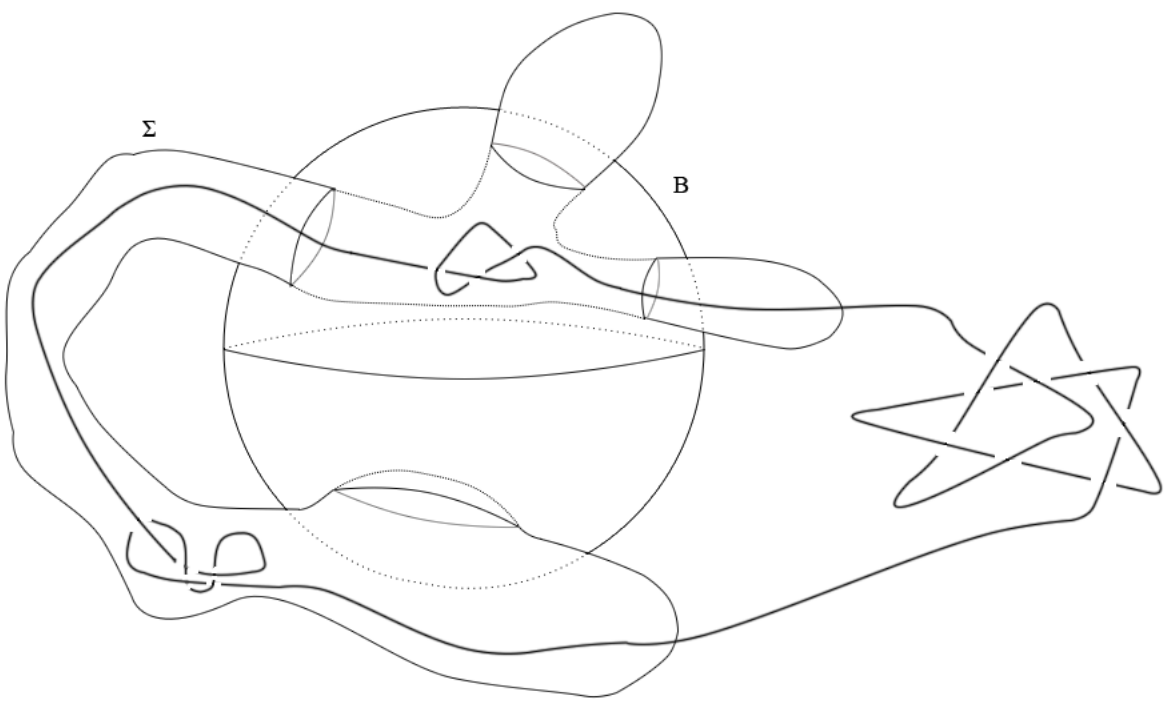

Figure 2.14: An example of how B and Σ can intersect.

Similar to what happened in the proof that genus is additive, theorem 5, the aim is to simplify the intersection $\Sigma \cap B$, and in a moment, we start out by considering the curves in $\Sigma \cap \partial B$.

If the intersection $\Sigma \cap B$ is empty, then the theorem holds true, since we would then have that the ball B is completely contained in one of the two components of the complement $S^3 \setminus \Sigma$, say the component containing K_1. Replace the component of $S^3 \setminus \Sigma$ that does not contain K_1 by a trivial ball-arc pair. Then the knot has become K_1, and ∂B partitions K_1 as $K_1 = P + K_1'$ for some knot K_1'. Similarly, we may have replaced (B, α) by a trivial ball-arc pair to change the knot into Q, and then Σ exhibits Q as the sum $Q = K_1' + K_2$, see figure 2.15.

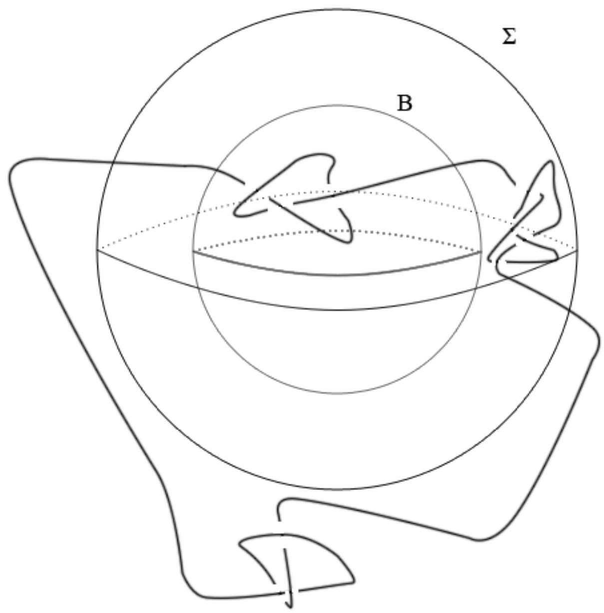

Figure 2.15: The case where B and Σ does not intersect.

From now on, we strive to simplify $B \cap \Sigma$, and if at some point this intersection is empty, we are done by the above argument.

Assume that $\Sigma \cap \partial B$ is non-empty. Each simple closed curve on Σ is a copy of S^1, and hence as $\Sigma \cap K$ is just two points, each oriented simple closed curve in $\Sigma \setminus K$ has linking number 0, +1 or −1 with K, see figure 2.16. Consider first the components of $\Sigma \cap \partial B$, whose linking number with K is zero. As illustrated, each such simple closed curve C has both points of $\Sigma \cap K$ on one side of it in Σ, and we shall consider this side of C in Σ as being the outside. Choose among all simple closed curves of linking number zero in $\Sigma \cap \partial B$ a curve C, which is innermost on Σ in the sense that in contains no other such curves on its inside. The curve C thus bounds a disc D on Σ with $D \cap \partial B = \partial D$, by the choice of C. The curve $C = \partial D$ also bounds a disc D' on ∂B with the property that $D' \cap K = \emptyset$, because otherwise the two sides of C in ∂B would contain a point each of $\partial B \cap K$, and then the knot would not have linking number 0 with the curve C, see figure 2.17 for an illustration.

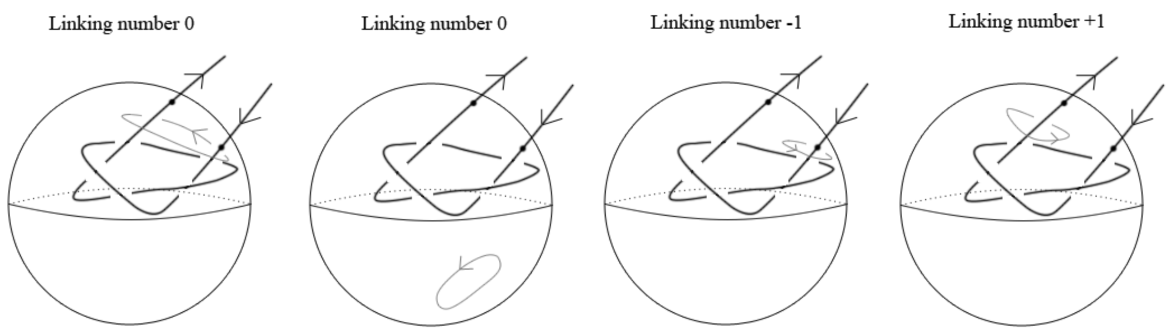

Figure 2.16: The possible cases showing how a simple closed curve on Σ can link with the knot K.

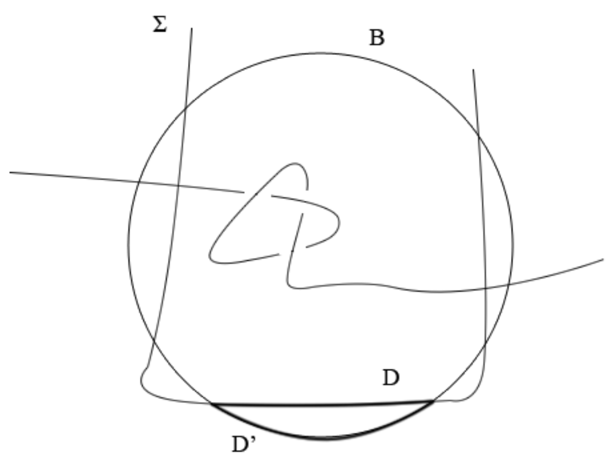

Figure 2.17: The disc D in $\Sigma \cap \partial B$ has linking number 0 with K, and thus so has D'.

The union of the two discs $D \cup D'$ bounds a ball in B by the three-dimensional Schönflies theorem. We wish to alter B by removing this ball, but without changing anything essential. Formally, we remove a small regular neighborhood in B of the ball bounded by $D \cup D'$, small enough not to intersect the knot K (which we may do, as the knot does not intersect D or D'). The result is shown in figure 2.18, compare with 2.17. We have thus changed our setup in such a way that the intersection $\Sigma \cap \partial B$ now has one less component than before. Denoting the removed regular neighborhood around the ball by V, we have a homeomorphism of pairs $(B \setminus V, \alpha) \simeq (B, \alpha)$, and thus we could as well have taken this new 3-ball $B \setminus V$ as B to begin with. Continuing like this, we may assume that we remove all the simple closed curves of $\Sigma \cap \partial B$ that have linking number zero with the knot K, and thus only the simple closed curves of linking number $+1$ or -1 remain.
We leave now the intersection $\Sigma \cap \partial B$, and consider instead $\Sigma \cap B$. If $\Sigma \cap B$ has a component, which is a disc, then we choose such a disc D. As D is a component of $\Sigma \cap B$, then the intersection $D \cap K$ consists of just one point. Otherwise the boundary ∂D of the disc would have been a simple closed curve of $\Sigma \cap \partial B$ with both points of $\Sigma \cap \partial B$ on one side of it in Σ, contrary to the fact that we have

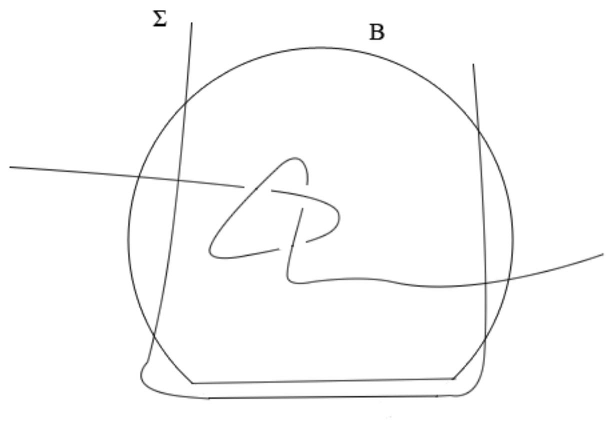

Figure 2.18: The ball B has been isotoped to miss the disc D.

eliminated any such curve. As P is a prime knot, one of the two sides of D in B is a trivial ball-arc pair, see figure 2.19. Indeed, remove the complement of B in S^3, and replace it by a trivial ball-arc pair. The knot has thus become P, by definition of B. Then as both sides of D in B are ball-arc pairs, one of those pairs must be trivial, as we would otherwise have a non-trivial decomposition of P. Remove from B a small regular neighborhood of the ball-part of this trivial ball-arc pair, small

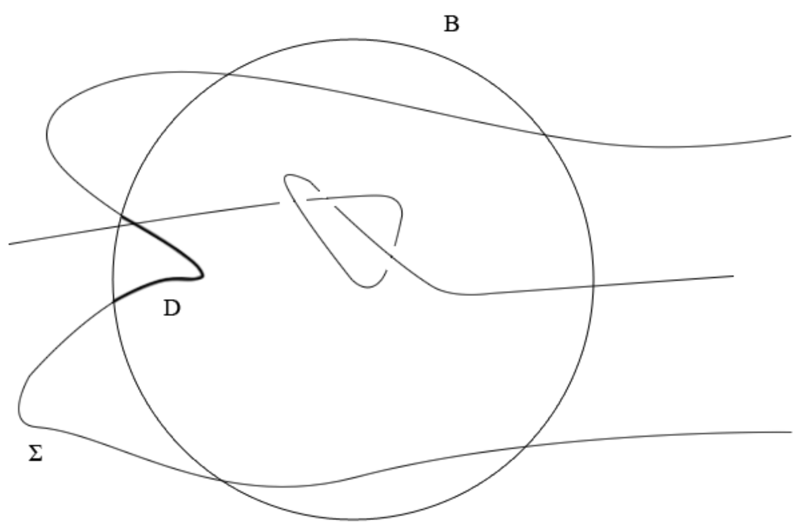

Figure 2.19: A disc D in $\Sigma \cap B$ whose boundary circle is of non-zero linking number.

enough not to change the knotted part of the arc α. Then B with this neighborhood removed is another 3-ball containing α, for which the homeomorphic type of (B, α) is unchanged, so we may as well assume that we had chosen such a ball B to begin with. Note that by eliminating D from

$\Sigma \cap B$, we may have been lucky to remove more from $\Sigma \cap B$ than just D. In particular, we have removed everything of $\Sigma \cap B$ on the trivial side of D. By repeated application of this process, we may now assume that $\Sigma \cap B$ has no disc components.

Now $\Sigma \cap B$ has been reduced to some finite collection of disjoint annuli, because firstly, we have just removed any disc components of this intersection, and secondly, if there was a component of $\Sigma \cap B$ with strictly more than two boundary circles, then as the knot K intersects ∂B exactly twice, all but two of these boundary circles would have linking number zero with K, contrary to the fact that we have already removed such boundary circles from $\Sigma \cap \partial B$.

Let A be an annulus component of $\Sigma \cap B$. As we have removed all simple closed curves in $\Sigma \cap \partial B$ of linking number 0 with the knot, K exits B at the ends of the annulus. Declaring the part of $B \setminus A$ containing the knot K to be the inside in B of the annulus, we may choose A to be an outermost annulus on B. Then ∂A bounds an annulus A' in ∂B, and by the choice of A, we have that $A' \cap \Sigma = \partial A'$, see figure 2.20. As the reader will see shortly, this annulus may also have a different appearance than that of figure 2.20.

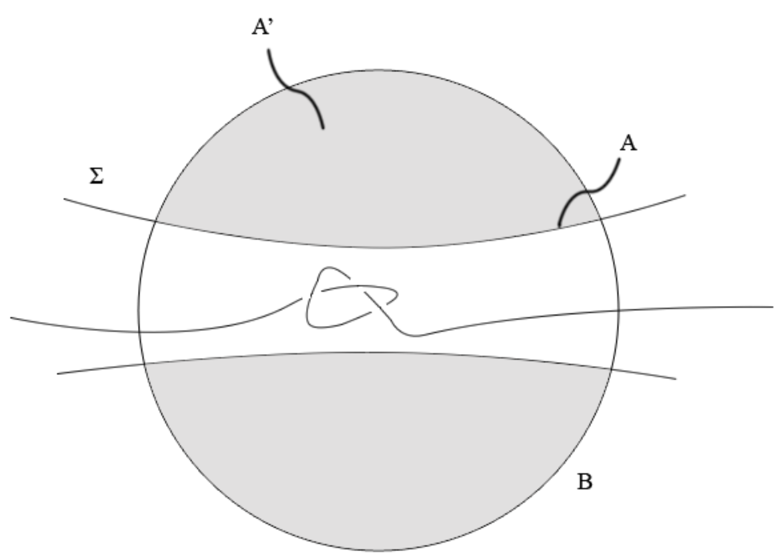

Figure 2.20: An example of how the two annuli A and A' may look like.

The two annuli A and A' have a common boundary, whereas the union $A \cup A'$ is a torus in B. Let M denote the part of B bounded by this torus, i.e. the component of the complement $B \setminus (A \cup A')$ not containing the knot K. We will see in a little while what M may look like. Let Δ denote the closure of one of the two components of $\partial B \setminus A'$, see figure 2.21.

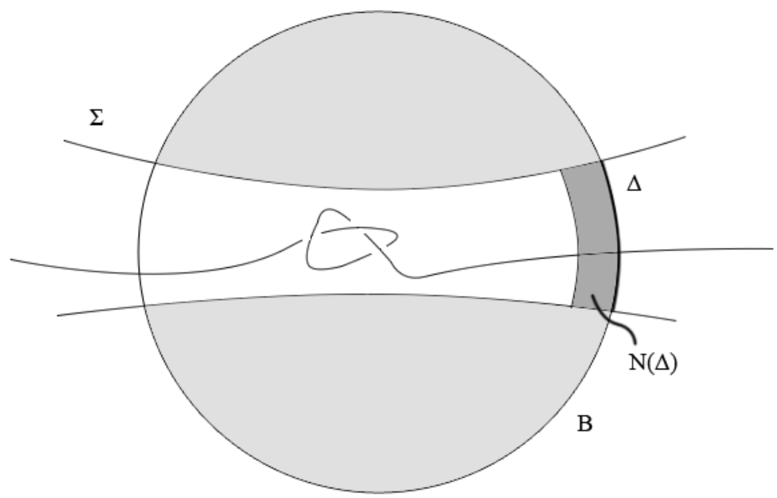

Figure 2.21: The disc Δ and its regular neighborhood $N(\Delta)$.

Then Δ is a disc, and as the knot exits ∂B through each end of the annulus A, and as Δ and A have the same boundary, the knot will intersect the disc Δ exactly once. We thicken Δ slightly into the complement $B \setminus M$, or in technical lingo, we choose some small regular neighborhood $N(\Delta)$ of Δ in the closure of $B \setminus M$. We choose the neighborhood so small that it intersects the knot in an unknotted arc, i.e. such that $(N(\Delta), N(\Delta) \cap \alpha)$ is a trivial ball-arc pair, see again figure 2.21. The 3-manifold $M \cup N(\Delta)$ is a ball, since its boundary is a 2-sphere. We see in figure 2.22 two cases, the first in which the ball-arc pair $(M \cup N(\Delta), N(\Delta) \cap \alpha)$ is trivial, as we may push one end of $N(\Delta)$ through the cavity of $M \cup N(\Delta)$ without knotting the small trivial arc $N(\Delta) \cap \alpha$, and the second in which the pair is a copy of the pair (B, α), since in this case, we may also push $N(\Delta)$ through the cavity, but this time the small arc $N(\Delta) \cap \alpha$ will be knotted like α on the way.

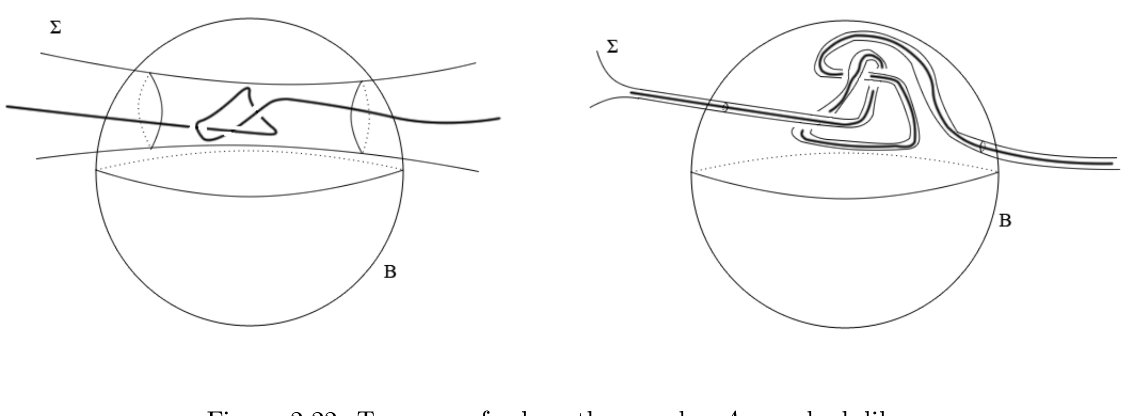

Figure 2.22: Two cases for how the annulus A may look like.

In fact, as P is prime, these two cases for the ball-arc pair $(M \cup N(\Delta), N(\Delta) \cap \alpha)$ are the only possible cases, for if the annulus A instead had the appearance as that of figure 2.23 (in which case the torus ∂M is known as a ***swallow follow torus***, because it starts out by swallowing the knot, and ends up following it), then we could push $N(\Delta)$ through the following part of the torus, but stop immediately before the swallowing part. Produce P in the usual way by replacing the complement of B by a trivial ball-arc pair. Then M, together with this new elongated neighborhood $N(\Delta)$, comprise a non-trivial ball-arc pair with the followed part of α. The complement of $M \cup N(\Delta)$ clearly also gives rise to a non-trivial ball-arc pair, contradicting the primeness of P.

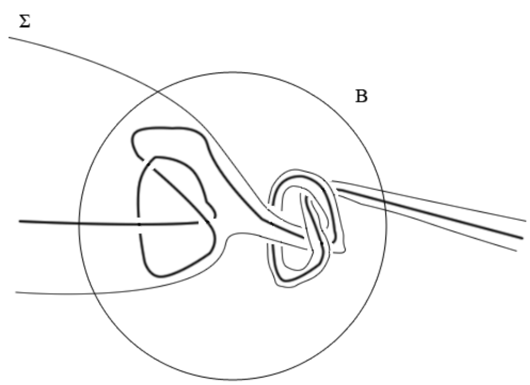

Figure 2.23: The impossible case of Σ swallowing half the arc α.

In the first case, where the pair $(M \cup N(\Delta), N(\Delta) \cap \alpha)$ is trivial, the set M is a solid torus, and as in the previous cases, we may remove a small regular neighborhood of M to obtain another ball B that exhibits the knot K as the sum $P + Q$ in the same way as before, and such that the homeomorphic type of (B, α) is unchanged.

As for the second case, if we isotope B to lie inside of the annulus A as to miss this part of Σ, we change the homeomorphic type of the pair (B, α), thus changing with the initial setup, where B separates P from Q. This case is thus very different from all the other cases of removing discs and annuli, and we need to think of something else.

The 2-sphere Σ separates K_1 from K_2, and the complement of Σ in S^3 has two components. Suppose without loss of generality that the component of $S^3 \setminus \Sigma$ which contains M also contains the arc of K corresponding to K_1. Reproduce K_1 by removing the other side of Σ (the side containing the arc α, possibly among other things), and by gluing on a trivial ball-arc pair in its place. As the part of Σ that followed the arc α was trivial to begin with, the knot K_1 will contain the arc α. The set B is still a ball, and the pair (B, α) is unchanged, so we have now that ∂B separates P from some knot K_1' in K_1, i.e. that $K_1 = P + K_1'$.

To show that Q then consists of the part of K that is not P, i.e. that $Q = K_1' + K_2$, we go back to the situation where K is the knot that Σ and ∂B separates, as in the second picture of figure 2.22. In this case, recall that the boundary of the set M is a torus, whence we may remove the interior of M, and replace it by a solid torus $S^1 \times D^2$. We glue the boundary of this solid torus to ∂M in such

a way that the boundary of each meridional disc of $S^1 \times D^2$ is glued onto a closed curve of ∂M that intersects $\partial\Delta$ exactly once, i.e. a curve that travels through the knotted hole along its boundary, and which at each end of the hole is connected by an arc following the boundary of B. This setup is not likely to be embeddable in $\mathbb{R}^3$, and we will not attempt a drawing.
Now $(S^1 \times D^2) \cup N(\Delta)$ is a ball: Since each $S^1 \times S^1$ of $S^1 \times D^2$ meets Δ, by construction, and as $N(\Delta)$ is a small regular neighborhood of Δ in the complement of $S^1 \times D^2$, the boundary of $(S^1 \times D^2) \cup N(\Delta)$ is a sphere. Replacing M by $S^1 \times D^2$ like this thus changes B to another ball B', but contrary to before, (B', α) is now a trivial ball-arc pair, as α is now a trivial arc going through the hole of $S^1 \times D^2$. Changing B to B', changes the complement of this ball from $S^3 \setminus B$ to $S^3 \setminus B'$, but the closure of the latter set is still a ball, so changing B to B' changes S^3 to a new copy of S^3. In conclusion, we may reproduce the knot Q by removing the side of $\partial B'$ not containing Q, and by replacing it with a trivial ball-arc pair, but (B', α) is already such a trivial ball-arc pair, and so the knot has already become Q. Similarly, we may produce K_1 by replacing the ball component of $S^3 \setminus \Sigma$ that contains the arc corresponding to K_2 by a trivial ball-arc pair. As before, $\partial B'$ partitions K_1 into a sum, but now one side of $\partial B'$ is already a trivial ball-arc pair, so K_1 has been reduced to K_1'. Thinking of Σ as decomposing the knot Q, the side of Σ that used to contain the arc corresponding to K_2 still corresponds to this arc, because as well as now as before, the part of Σ inside of B follows the arc α: Indeed, note that Σ may pass through B' more than one time, but as we have made sure that the outermost annulus A of $\Sigma \cap B$ is one which follows the knot, any other "‘passing through" annulus of $\Sigma \cap B$ must also follow the knot, so that all passing through annuli bound a trivial ball-arc pair inside of B, whence these parts of Σ in B' contribute nothing to K_2. It may also happen that Σ enters B', starts following the arc α, but regrets it half-way and loops back on itself, exiting B' the same way it came in, but this is easily seen to contribute to K_2 only trivially. We already argued that the other side of Σ corresponds to K_1', so in conclusion $Q = K_1' + K_2$, which we set out to prove. The other option of the theorem occurs of course, if we exchange the roles of K_1 and K_2. □

We use the theorem to prove the following cancelation property.

Corollary 4. *Let P be a prime knot, and say $P + Q = K_1 + K_2$ for some knots Q, K_1, and K_2. Assume that $P = K_1$. Then also $Q = K_2$.*

Proof. Applying theorem 7, we know that one of two things can happen. The first is that $K_1 = P + K_1'$ for some knot K_1', and that $Q = K_1' + K_2$. By assumption $P = K_1$, whence $K_1 = K_1 + K_1'$, so that the genus of K_1' is zero, implying that K_1' is the unknot. The result follows. The second possibility is that $K_2 = P + K_2'$ for some knot K_2', and $Q = K_1 + K_2'$. Using again $P = K_1$, we obtain $Q = K_1 + K_2' = P + K_2' = K_2$, as claimed. □

Finally, we prove what we had really been aiming at all along.

Theorem 8. *Any knot K can be factored uniquely as a sum of prime knots (up to a ordering of the factors).*

Proof. We have already proved existence, so suppose that we can write K as

$$K = P_1 + \cdots + P_m = Q_1 + \cdots + Q_n,$$

where all P_i and Q_j are prime. We prove the result by induction on m. If $m = 0$, then as no non-trivial knot has an additive inverse, we have also $n = 0$, so assume that $m \geq 1$. By repeated application of theorem 7, the knot P_1 is a summand of either $Q_1, Q_2, \ldots, Q_{n-1}$ or Q_n, say Q_j. Then certainly $P_1 = Q_j$, as Q_j is prime. By corollary 4, P_1 may be canceled from both sides of the equation, whence the required result follows by the induction hypothesis. □

Bibliography

[1] Adams, C. C. *The Knot Book.* American Mathematical Society, 2004.

[2] Burde, G., and Zieschang, H. *Knots.* de Gruyter, 1985.

[3] Hatcher, A. *Algebraic Topology.* Cambridge, 2005.

[4] Hempel, J. *3-manifolds.* American Mathematical Society, 2004.

[5] Lickorish, W. R. *An Introduction to Knot Theory.* Springer, 1997.

[6] Massey, W. S. *Algebraic Topology: An Introduction.* Springer, 1991.