Hard Unknots and Collapsing Tangles

Louis H. Kauffman and Sofia Lambropoulou

1 Introduction

Classical knot theory is about the classification, up to isotopy, of embedded closed curves in three-dimensional space. Two closed curves embedded in three-dimensional space are said to be *isotopic* if there is a continuous family of embeddings starting with one curve and ending with the other curve. This definition of isotopy captures our intuitive notion of moving a length of elastic rope about in space without tearing it. Given a closed loop of rope in space, it can sometimes be deformed in this way to a simple flat circle. In such a case we say that the loop is *unknotted*.

So it is not absurd if we sometimes ask: Is this knot diagram unknotted? The question is not only non-contradictory, it is often hard. Look ahead to Figures 5 and 6. The shapes illustrated are unknot diagrams, but most people's intuition does not allow recognizing the unknotting, although it will present itself if one makes a rope model and shakes it. Recognizing unknot diagrams is a starting point for all of knot theory.

Corresponding to the physical models, one can make mathematical models of knots by formalizing the diagrams that one naturally draws for them, as in Figure 2. The study based on the diagrams is called *combinatorial knot theory*. The diagrams are well defined as mathematical structures, namely, planar graphs with four edges incident to each vertex and with extra structure: indication at each vertex of which strand crosses over and which under.

This paper studies unknot diagrams and the collapsing of diagrams to knotted diagrams. We present an infinite collection of examples of unknot diagrams that are *hard*, in the sense that they cannot be simplified by Reidemeister moves without being made more complicated. We pinpoint the smallest hard unknot diagrams (Theorem 8), and we give a beautiful characterization of a certain class of unkots in terms of continued fractions and their convergents (Theorems 6 and 7). Our methods of proof are elementary and so we make the paper as self-contained as possible. Along the way, we develop and use John H. Conway's theory of *rational tangles*. Rational tangles allow algebraic operations, leading to connections with matrices and continued fractions. We extend our previous results [28] about collapsing of closures of tangle sums to the unknot to collapses to arbitrary knots and links. Thus we produce "hard" versions of knots and links as well as hard unknots and characterize these also in terms of continued fractions (Theorems 9 and 10).

Given a knot or link one can apply our methods of collapse in reverse, producing complex configurations that collapse back to that knot or link. This configuration can be allowed to undergo a crossing switch that locks it into a new knot or link that does not collapse. Thus we can create towers of knots and links such that each stage of the tower will collapse to the previous stage under a single crossing switch.

We envision this paper as useful for research knot theorists, for students beginning in knot theory and for scientists such as biologists who would like an entrance into knot theory that leads to connections with other fields. In particular, we have ended the paper with two sections about such connections. Section 12 is about the connection of our results with molecular biology, and it discusses how the form of the sequence of knots and links in a process of processive recombination stabilizes so that eventually extra twists are being added to the same place in the diagrammatic form of the produced knots and links. Along with this stabilization process, there is the general form of the recombination process itself that is related to our study of collapse. Our results apply to the process of successive recombination in DNA where a molecular configuration may undergo a crossing switch of DNA strands that locks it into a new topological configuration. Thus, in successive recombination towers of knots and links are created, each one a single crossing switch away from the next. The patterns of long sequences of such recombinations are beyond the reach of experiment, but can be understood mathematically. The results of this paper could apply to future experiments with DNA recombination. The reader may be interested in comparing our results with those of Isabel Darcy in her papers [7, 8].

Perhaps the most intriguing of our results is formulated in Theorem 7 which tells us that the numerator closure of the sum of two rational tangles is an unknot if and only if the continued fractions for the two tangles are convegents of one another. This means that one continued fraction is a one-step truncate of the other, and it means that if P/Q and R/S are the reduced fractions, then $|PS - QR| = 1$. Fractions with this relationship come up in many mathematical contexts. Section 13 makes connection with other mathematics via this very relationship. Here we contact Farey series, Pick's Theorem and the Ford circles in the hyperbolic plane. This section demonstrates how far-reaching are the mathematical processes involved in the study of fundamental questions in topology.

Acknowledgements. The first author thanks the National Science Foundation for support of this research under NSF Grant DMS-0245588. It gives both authors pleasure to acknowledge the hospitality of the Mathematisches Forschungsinstitut Oberwolfach, the University of Illinois at Chicago and the

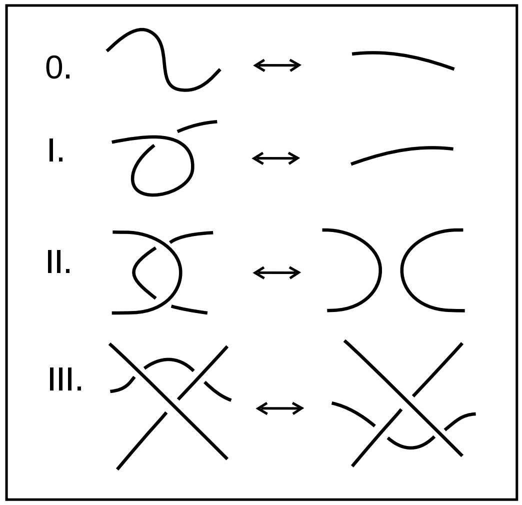


Figure 1: The Reidemeister Moves

National Technical University of Athens, Greece, where much of this research was conducted. We particularly thank Slavik Jablan for conversations and for helping us by using his computer program LinKnot.

2 The Culprit

Combinatorial knot theory got its start in the hands of Kurt Reidemeister [38], who discovered a set of moves on planar diagrams that capture the topology of knots and links embedded in three dimensional space. Reidemeister proved that if we have two knots or links in three-dimensional space, then they are isotopic if and only if corresponding diagrams for them can be obtained, one from the other, by a sequence of moves of the types shown in Figure 1.

The moves are local and they are performed on the diagrams in such a way that the rest of the diagram is left fixed. This means that when one searches for available Reidemeister moves on a diagram, one searches for one-sided, two-sided, and three-sided regions with the approriate crossing patterns as shown in Figure 1. It is important to note that knot and link diagrams are taken to be on the surface of a two dimensional sphere that is standardly embedded in three-dimensional space. Thus the outer region of a planar diagram is handled just like any other region in that diagram, and can be the locus of any of the moves. In this section, we discuss the unknotting of knot diagrams using the Reidemeister moves.

Definition 1. We define the *complexity* of a knot or link diagram K to be the number of crossings, $C(K)$, in the diagram. A Reidemeister move is said to

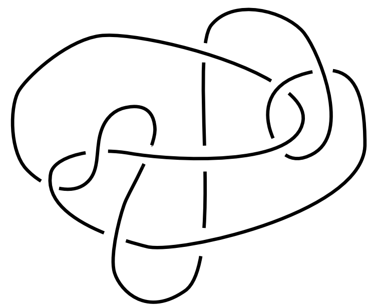


Figure 2: The Culprit

be *simplifying* if it reduces the crossing number of the diagram. We say that a diagram *can be simplified* (to the unknot) if there is a sequence of simplifying Reidemeister moves, combined with type III Reidemeister moves that transform the diagram to the unknot diagram. We shall call a diagram of the unknot *hard* if it has the following two properties.

1. There are no simplifying type I or type II moves on the diagram.

2. There are no type III moves on the diagram.

Hard unknot diagrams must be made more complex before they will simplify to the unknot, if we use Reidemeister moves.

Here is an example of a hard unknot diagram (originally due to Ken Millett [34]), in Figure 2. We like to call this diagram "The Culprit". The Culprit is a hard unknot diagram. It has 10 crossings, and in order to be undone, we definitely have to increase the number of crossings before decreasing them to zero. The reader can verify this for himself by checking each region in the diagram of the Culprit. A simplifying Reidemeister II-move can occur only on a two-sided region, but no two-sided region in the diagram admits such a move. Similarly, on the Culprit diagram there are no simplifying Redeimeister I-moves and there are no Reidemeister III-moves (note that a III-move does not change the complexity of the diagram). Culprit is just our term for a hard unknot, and we usually refer to Millett's example as "The Culprit" in this paper because it was our detective work in tracking the properties of this culprit that led to our results and questions.

View Figure 3 for an unknotting sequence for the Culprit. Notice that we undo it by swinging the arc that passes underneath most of the diagram outward, and that in this process the number of crossings in the intermediate diagrams increases. In the diagrams of Figure 3 the largest increase is to a diagram of 12 crossings. This is the best possible result for the initial diagram in the figure.

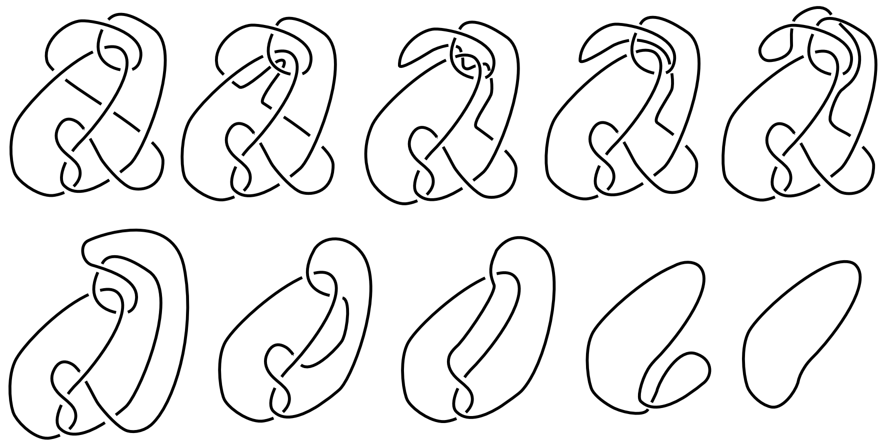


Figure 3: The Culprit Undone

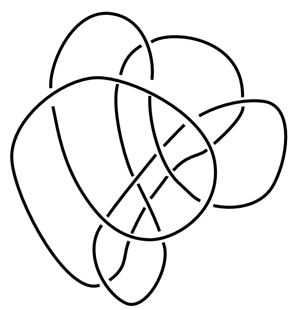


Figure 4: An Easy Unknot

In contrast to the Culprit, view Figure 4 for an example of an easy unknot. This one simplifies after Reidemeister III-moves unlock further moves of type I and type II. The reader can probably just look at the diagram in Figure 4 and see at once that it is unknotted.

It is an unsolved problem just how much complexity can be forced by a hard unknot diagram.

Definition 2. Let K be a hard unknot diagram. Let K' be a diagram isotopic to K such that K' can be simplified to the unknot. For any unknotting sequence of Reidemeister moves for K there will be a diagram K'_{max} with a maximal number of crossings. Let $Top(K)$ denote the minimum of $C(K'_{max})$ over all unknotting sequences for K. Let

$$R(K) = Top(K)/C(K)$$

be called the *recalcitrance* of the hard unknot diagram K. Very little is known about $R(K)$.

In the case of our Culprit, $Top(K) = 12$, while $C(K) = 10$. Thus $R(K) = 1.2$. A knot that can be simplified with no extra complexity has recalcitrance equal to 1. We shall have more to say about the recalcitrance later.

In [20] Hass and Lagarias show that there exists a positive constant c_1 such that, for each $n > 1$, any unknotted diagram K with n crossings can be transformed to the trivial knot diagram using less than or equal to $2^{c_1 n}$ Reidemeister moves. As a corollary to this result, they conclude that any unknotted diagram K can be transformed by Reidemeister moves to the trivial knot diagram through a sequence of knot diagrams each of which has at most $2^{c_1 n}$ crossings. In our language, this result says that

$$R(K) \leq \frac{2^{c_1 C(K)}}{C(K)}.$$

The authors of [20] prove their result for $c_1 = 10^{11}$, but remark that this is surely too large a constant. Much more work needs to be done in this domain. We should also remark that the question of the knottedness of a knot is algorithmically decideable, due to the work of Haken and Hemion [21]. This algorithm is quite complex, but its methods are used in the work of Hass and Lagarias.

In fact, we can now do considerably better in estimating $R(K)$. In [22] it is shown that if the diagram K represents the unknot and if K is presented with a height function in the plane (a Morse diagram) with $m(K)$ maxima and $c(K)$ crossings then one can unknot K by Reidemeister moves using a diagram K' with no more than $(c(K) + 2m(K) - 2)^2$ crossings. This means that

$$R(K) \leq \frac{(c(K) + 2m(K) - 2)^2}{c(K)}.$$

The problem of bounding the number of Reidemeister moves needed to unknot K is still difficult. A technique of working with large-scale moves is formalized in the work of Dynnikov [9] where he shows, using special grid diagrams and a complete set of moves distinct from Reidemeister moves, how to unknot diagrams without making the special diagrams more complex. Dynnikov's methods are used in proving the results in [22].

We mention the phenomenon shown in Figure 26. Here the first diagram illustrated is a hard unknot published by Goeritz [18] in 1934. We show how to unknot the Goeritz diagram in this figure by illustrating large-scale swing moves that are combinations of Reidemeister moves. Each swing move passes through diagrams that are larger than the original diagram before the swing is started. Note that in the figure there are a number of swing moves so that, using Reidemeister moves, the number of crossings in the diagram goes up and

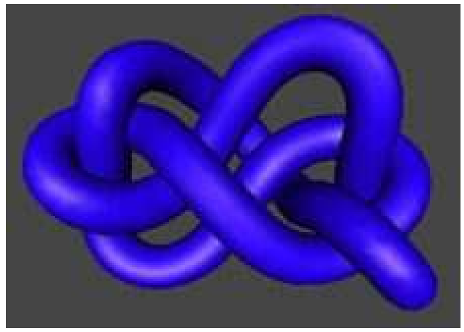


Figure 5: Three-Dimensional Unknot

down a number of times. Furthermore, right in the middle of the process we arrive at another, smaller, hard unknot diagram (here labeled H). In this way we discovered H, the smallest possible hard unknot diagram.

The next example is illustrated in Figure 5, a three-dimensional rendering of an unknot. This example is interesting psychologically, because it looks knotted to most observers. The three-dimensional picture is a frame from a deformation of this example in the energy minimization program KnotPlot [40]. It is illuminating to run the program on this example and watch the knot self-repel and undo itself. We leave the verification that this is an unknot as an exercise for the reader.

In Figure 6 we give an example of a hard unknot that is of a different type than the sort that we are considering in this paper. (It is not the closure of the sum of two rational tangles.) There are many knots and many hard unknots. We only scratch the surface of this subject.

For work on unknots related to braids see [35]. Another approach to detecting unknots is the use of invariants of knots and links. For example, it is conjectured that the Jones polynomial [24, 25] of a knot diagram is equal to one, only when that knot is unknotted.

3 Rational Tangles, Rational Knots and Continued Fractions

In this section we recall the subject of rational tangles and rational knots and their relationship with the theory of continued fractions. By the term "knots" we will refer to both knots and links, and whenever we really mean "knot" we shall emphasize it. Rational knots comprise the simplest class of knots. They are also known in the literature as Viergeflechte, four-plats or 2-bridge knots depending on their geometric representation. The theory of tangles was

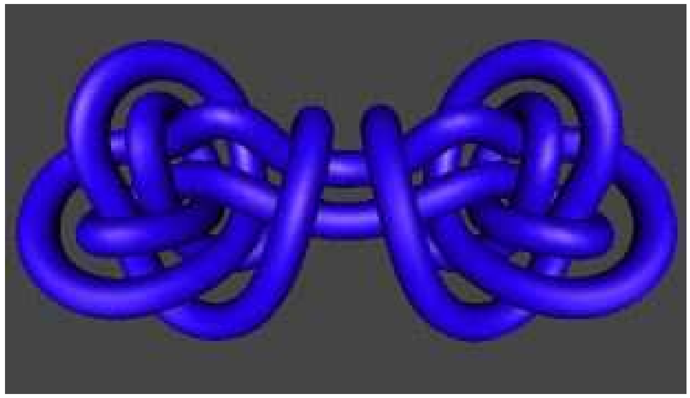


Figure 6: Another Hard Unknot

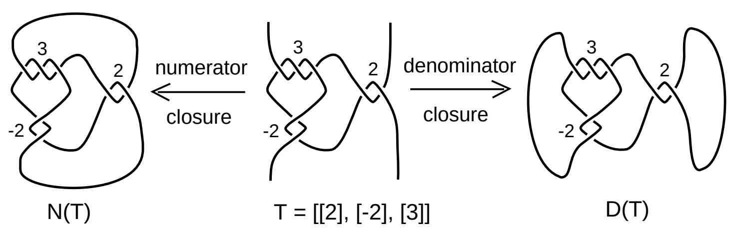


Figure 7: A Rational Tangle and its Closures to Rational Knots

introduced in 1967 by Conway [5] in his work on enumerating and classifying knots and links.

A 2-*tangle* is a proper embedding of two unoriented arcs and a finite number of circles in a 3-ball B^3, so that the four endpoints lie in the boundary of B^3. A *tangle diagram* is a regular projection of the tangle on a cross-sectional disc of B^3. By "tangle" we will mean "tangle diagram". A *rational tangle* is a special case of a 2-tangle obtained by applying consecutive twists on neighbouring endpoints of two trivial arcs. Such a pair of arcs comprise the [0] or $[\infty]$ tangles (see Figure 8), depending on their position in the plane. We shall say that such a tangle is in *twist form*. For an example see Figure 7. Conway defined the rational knots as "numerator" or "denominator" closures of the rational tangles. See Figure 7. Conway [5] also defined *the fraction* of a rational tangle to be a rational number or ∞, obtained via a continued fraction that is associated with the tangle. We discuss this construction below.

We are interested in tangles up to isotopy. Two tangles, T, S are *isotopic*, denoted by $T \sim S$, if and only if they have identical configurations of their four endpoints, and they differ by a finite sequence of the Reidemeister moves

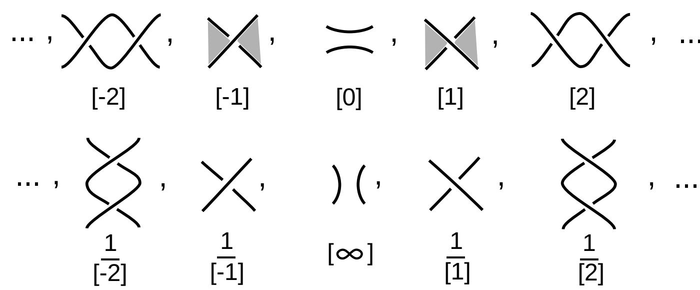


Figure 8: The Elementary Rational Tangles and the Types of Crossings

which take place in the interior of the projection disc. Of course, each twisting operation used in the definition of a rational tangle changes the isotopy class of the tangle to which it is applied. Rational tangles are classified by their fractions by means of the following theorem, different proofs of which are given in [36], [3, 4], [19] and [26].

Theorem 1 (Conway, 1975). *Two rational tangles are isotopic if and only if they have the same fraction.*

3.1 The Tangle Fraction

We shall now recall from [26] the main properties of rational tangles and of continued fractions, which illuminate the classification of rational tangles. The elementary rational tangles are displayed as either horizontal or vertical twists, and they are enumerated by integers or their inverses, see Figure 8.

The crossing types of 2-tangles (and of unoriented knots) follow the *checkerboard rule*: shade the regions of the tangle in two colors, starting from the left outside region with grey, and so that adjacent regions have different colors. Crossings in the tangle are said to be of "positive type" if they are arranged with respect to the shading as exemplified in Figure 8 by the tangle $[+1]$, i.e. they have the region on the right shaded as one walks towards the crossing along the over-arc. Crossings of the reverse type are said to be of "negative type" and they are exemplified in Figure 8 by the tangle $[-1]$. The reader should note that our crossing type conventions are the opposite of those of Conway in [5]. Our conventions agree with those of Ernst and Sumners [12, 13], [43] which, in turn, follow the standard conventions of biologists.

In the class of 2-tangles we have the operations *addition* and *multiplication*, as illustrated in Figure 9, which are denoted by "+" and "$*$" respectively. These operations are well-defined up to isotopy. In order to read out a rational tangle we transcribe it as an algebraic expression using horizontal and vertical twists. For example, Figure 10 illustrates the rational tangle in twist form: $[1] + ([1] * [3] * \frac{1}{[-3]}) + [1]$, and also an isotopic form of the same tangle with expression

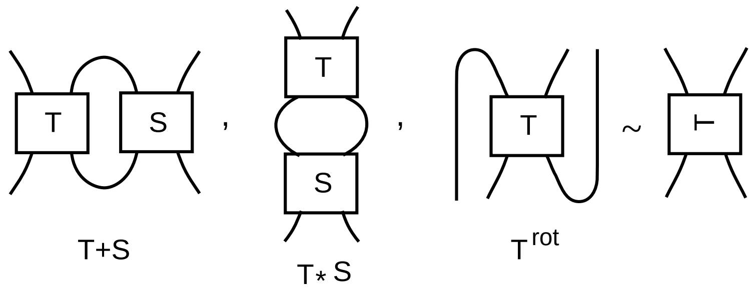


Figure 9: Addition, Multiplication and Rotation of 2-Tangles

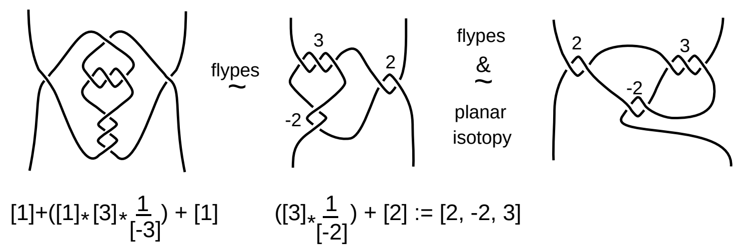


Figure 10: A Rational Tangle in Twist Form Converted to Its Standard Form and to Its 3-Strand-Braid Representation

$(([3] * \frac{1}{[-2]}) + [2])$. The decomposition of this last tangle is shown in Figure 11.

Note that addition and multiplication do not, in general, preserve the class of rational tangles (but every rational tangle can be constructed from the [0] tangle or the $[\infty]$ tangle by using these two operations). For example, the 2-tangle $\frac{1}{[3]} + \frac{1}{[3]}$ is not rational. The sum (product) of two rational tangles is rational if and only if one of the two consists in a number of horizontal (vertical) twists.

The *mirror image* of a tangle T, denoted $-T$, is T with all crossings switched. For example, $-[n] = [-n]$ and $-\frac{1}{[n]} = \frac{1}{[-n]}$. Then, the *subtraction* is defined as $T - S := T + (-S)$. The *rotation* of T, denoted T^{rot}, is obtained by rotating T on its plane counterclockwise by 90^0. The *inverse* of T is defined to be $-T^{rot}$. Thus, inversion is accomplished by rotation and mirror image. Note that T^{rot} and the inverse of T are in general not isotopic to T and they are order 4 operations. But for rational tangles the inversion is an operation of order 2. (This follows from *the remarkable property of rational tangles that they are isotopic to their horizontal or vertical flip in space by* 180^0. Note that a flip switches the endpoints of the tangle and, in general, a flipped tangle is not isotopic to the original one.) For this reason we shall denote the inverse of a rational tangle T by $1/T$, and hence the rotation of the tangle T will be denoted by $-1/T$. This explains the notation for the tangles $\frac{1}{[n]}$.

As we said earlier, there is a fraction associated to a rational tangle which characterizes its isotopy class (Theorem 1). In Figure 11, we illustrate how one obtains the fraction by using the decomposition of a rational tangle into sums and products. Notice, in that figure, how the tangle labeled x is rotated by 90^0 (becoming $-1/x$) and is then recognizable as $-1/x = -1/[3] + [2]$. The arithmetical operations for computing the tangle fraction are in direct parallel to this "topological arithmetic". Thus the fraction associated with x is $F(x) = -1/(-1/3 + 2)$ and, putting this together, the fraction for the tangle in the figure is $F(T) = 2 - 1/(-1/3 + 2)$.

In fact, the fraction is defined for any 2-tangle and always obeys the following three rules:

1. $F([\pm 1]) = \pm 1$.
2. $F(T + S) = F(T) + F(S)$.
3. $F(T^{rot}) = -1/F(T)$.

If $T = [\infty]$, we assign $F([\infty]) := \infty = \frac{1}{0}$, as a formal expression.

These rules suffice for computing the fraction inductively for rational tangles. In this way, they serve as the definition of the tangle fraction for rational tangles. Read Figure 11 again and note that the computation follows exactly the rules above. Generalizing, the reader can easily deduce the formula below:

$$T * \frac{1}{[n]} = \frac{1}{[n] + \frac{1}{T}}$$

Indeed, rotate $T * \frac{1}{[n]}$ by 90^0 and note that it becomes $-[n] - \frac{1}{T}$. The original tangle is the negative reciprocal of this tangle. This formula implies that the two operations: addition of $[+1]$ or $[-1]$ and inversion between rational tangles suffice for generating the whole class of rational tangles. As for the fraction, we have the corresponding formula:

$$F(T * \frac{1}{[n]}) = \frac{1}{n + \frac{1}{F(T)}}.$$

For example, in Figure 11, the fraction associated with the tangle T is

$$F(T) = 2 + \frac{1}{-2 + \frac{1}{3}}.$$

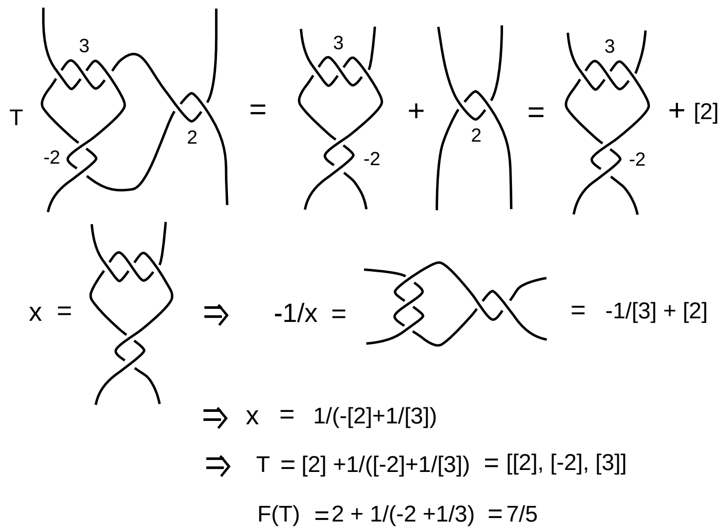


Figure 11: Finding the Fraction

Starting from a rational tangle in twist form, one can always bring the twists alternately to the right and to the bottom of the tangle using only isotopy moves on subtangles, called *flypes*. See Figure 12 for an abstract illustration of flypes. Note that in a flype the endpoints of the subtangle remain fixed; this is not true for a flip of the tangle in space. One can then associate to the rational tangle a vector of integers $(a_1, a_2, \ldots, a_n)$, where the first entry denotes the place where the tangle starts untwisting and the last entry where it begins to twist. For example the tangle of Figure 7 corresponds to the vector $(2, -2, 3)$. We then say that the rational tangle is in *standard form* and we denote it by $T = [[a_1], [a_2], \ldots, [a_n]]$. We also call this standard form the *continued fraction form* of the tangle, since it can literally be written in the form of a continued fraction involving the elementary integer tangles $[a_i]$.

From the above discussion it follows that for a rational tangle in standard form,

$$T = [[a_1], [a_2], \ldots, [a_n]] = [a_1] + \cfrac{1}{[a_2] + \cdots + \frac{1}{[a_{n-1}] + \frac{1}{[a_n]}}}$$

for $a_1 \in \mathbb{Z},\ a_2, \ldots, a_n \in \mathbb{Z} - \{0\}$, its fraction is the numerical continued fraction

$$F(T) = [a_1, a_2, \ldots, a_n] := a_1 + \cfrac{1}{a_2 + \cdots + \frac{1}{a_{n-1} + \frac{1}{a_n}}}.$$

Since fractions and rational tangles correspond, we will, from now on, use the notation $[a_1, \ldots, a_n]$ to denote both the continued fraction and a rational tangle in continued fraction form.

Figure 12: The Flype Moves

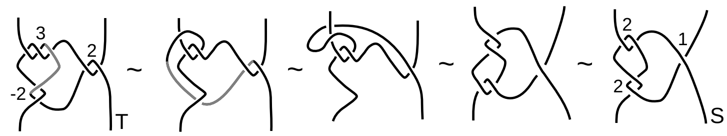


Figure 13: Reducing to Alternating Form Using the Swing Moves

The subject of continued fractions is of perennial interest to mathematicians. See for example [30], [37], [33], [45]. Here we only consider continued fractions of the above type, with all numerators equal to 1. As in the case of rational tangles, we allow the term a_1 to be zero. Clearly, the two simple algebraic operations *addition of* $+1$ *or* -1 and *inversion* generate inductively the whole class of continued fractions, starting from zero.

If a rational tangle T changes by an isotopy, the associated continued fraction form may also change, while the fraction remains invariant. For example, $[2,-2,3] = [1,2,2] = \frac{7}{5}$, see Figure 13. Indeed, the rational tangles $[2,-2,3]$ and $[1,2,2]$ are isotopic as shown by the swing moves of Figure 13. A *swing move* is an isotopy move in which we swing an arc joining two consecutive crossings of opposite type across a subtangle box. See the first move in Figure 13. For the isotopy type of a rational tangle T with fraction $\frac{p}{q}$ we shall use the notation $[\frac{p}{q}]$.

The key to the exact correspondence of fractions and rational tangles lies in the construction of a canonical alternating form for the rational tangle. A tangle is said to be *alternating* if it is isotopic to a tangle where the crossings alternate from under to over as we go along any component or arc of the weave. Similarly, a knot is alternating if it possesses an alternating diagram. Using the swing moves and an induction argument, one shows that rational tangles (and rational knots) are alternating. We shall say that the rational tangle $S = [\beta_1, \beta_2, \ldots, \beta_m]$ is in *canonical form* if S is alternating and m is odd. From the above, S alternating implies that the β_i's are all of the same sign. It turns out that the canonical form for S is unique. In Figure 13 we bring our working rational tangle $T = [2,-2,3]$ to its canonical form $S = [1,2,2]$.

On the other hand, by Euclid's algorithm and keeping all remainders of the same sign, one can show that every continued fraction $[a_1, a_2, \ldots, a_n]$ can be transformed to a unique canonical form $[\beta_1, \beta_2, \ldots, \beta_m]$, where all β_i's are positive or all negative integers and m is odd. For example,

$$[2,-2] = [1,1,1] = \frac{3}{2}.$$

There is also an algorithm that can be applied directly to the initial continued fraction to obtain its canonical form, which works in parallel with the algorithm for the canonical form of rational tangles. Indeed, we have:

Proposition 1. *The following identity is true for continued fractions and it is also a topological equivalence of the corresponding tangles:*

$$[..., a, -b, c, d, e, ...] = [..., (a-1), 1, (b-1), -c, -d, -e, ...].$$

This identity gives a specific inductive procedure for reducing a continued fraction to all positive or all negative terms. In the case of transforming to all negative terms, we can first flip all signs and work with the mirror image. Note also that

$$[..., a, b, 0, c, d, e, ...] = [..., a, b+c, d, e, ...]$$

will be used in these reductions.

Proof. The identity in the Proposition follows immediately from the formula:

$$a + 1/(-b) = (a-1) + 1/(1 + 1/(b-1)).$$

If a and b are positive, this formula allows the reduction of negative terms in a continued fraction. □

For the study of rational knots it is easier to use another way of representing an abstract rational tangle in standard form, illustrated in Figure 10. This is the 3-*strand-braid representation*. As illustrated in Figure 10, the 3-strand-braid representation is obtained from the standard representation by planar rotations of the vertical sets of crossings, thus creating a lower row of horizontal crossings. By the checkerboard shading convention, the signs of the crossings in the lower row do not change under these rotations.

Note that the set of twists of a rational tangle may be always assumed *odd*. Indeed, let n be even and let the left-most twist $[a_1]$ be on the upper part of the 3-strand braid representation. Up to the ambiguity (upper or lower row) of the right-most crossing, the vector associated to a rational tangle is *unique*, i.e. $(a_1, a_2, \ldots, a_n) = (a_1, a_2, \ldots, a_n - 1, 1)$, if $a_n > 0$, and $(a_1, a_2, \ldots, a_n) = (a_1, a_2, \ldots, a_n + 1, -1)$, if $a_n < 0$.

Remark 1. We have omitted here the proof of the invariance of the fraction. The interested reader can consult [5], [19], [26] for various proofs of this fact. One definition of the fraction of a general 2-tangle is:

$$F(T) = i < N(T) > / < D(T) >,$$

where $i^2 = -1$ and $< L > = < L > (A)$ denotes the bracket polynomial [25] of a link L with $A = i^{1/2}$. Here $N(T)$ and $D(T)$ are the numerator and denominator closures of the tangle T, as illustrated in Figure 7.

3.2 Rational Knots and Continued Fractions

By joining with simple arcs the two upper and the two lower endpoints of a 2-tangle T, we obtain a knot called the *Numerator* of T, denoted $N(T)$. Joining with simple arcs each pair of the corresponding top and bottom endpoints of T we obtain the *Denominator* of T, denoted by $D(T)$, see Figure 7. Note that $N(T) = D(T^{rot})$ and $D(T) = N(T^{rot})$. As we shall see in the next section, the numerator closure of the sum of two rational tangles is still a rational knot. But the denominator closure of the sum of two rational tangles is not necessarily a rational knot. For example, consider the denominator of the sum $\frac{1}{[3]} + \frac{1}{[3]}$.

A rational knot is defined to be the numerator of a rational tangle. Given two different rational tangle types $[\frac{p}{q}]$ and $[\frac{p'}{q'}]$, when do they close to isotopic rational knots? More than one rational tangle can yield isotopic rational knots. The corresponding equivalence relation between the rational tangles is reflected by an arithmetic equivalence of their corresponding fractions. Rational knots and links are one of the few cases in knot theory where we have a complete classification story. Indeed, we have the following theorem:

Theorem 2 (Schubert, 1956). *Suppose that rational tangles with fractions $\frac{p}{q}$ and $\frac{p'}{q'}$ are given. Here p and q are relatively prime, similarly for p' and q'. If $K(\frac{p}{q})$ and $K(\frac{p'}{q'})$ denote the corresponding rational knots obtained by taking numerator closures of these tangles, then $K(\frac{p}{q})$ and $K(\frac{p'}{q'})$ are isotopic if and only if*

1. *$p = p'$ and*
2. *either $q \equiv q' \bmod p$ or $qq' \equiv 1 \bmod p$.*

Different proofs of Theorem 2 are given in [41], [3], [27]. The proof in [27] is the only published combinatorial proof of Schubert's Theorem.

Schubert classified rational knots by finding canonical forms via representing them as 2-bridge knots. In [27] we gave a new combinatorial proof of Theorem 2, by posing the question: given a rational knot diagram, at which places may one cut it so that it opens to a rational tangle? We then pinpoint two distinct categories of cuts that represent the two cases of the arithmetic equivalence of Schubert's theorem. The first case corresponds to the *special cut*, as illustrated in Figure 14. The two tangles $T = [-3]$ and $S = [1] + \frac{1}{[2]}$ are non-isotopic by the Conway Theorem (Theorem 1), since $F(T) = -3 = 3/(-1)$, while $F(S) = 1 + 1/2 = 3/2$. But they have isotopic numerators: $N(T) \sim N(S)$, the left-handed trefoil. Now $-1 \equiv 2 \bmod 3$, confirming Theorem 2. See [27] for a complete analysis of the special cut.

The second case of Schubert's equivalence corresponds to the *palindrome cut*, an example of which is illustrated in Figure 15. Here we see that the tangles

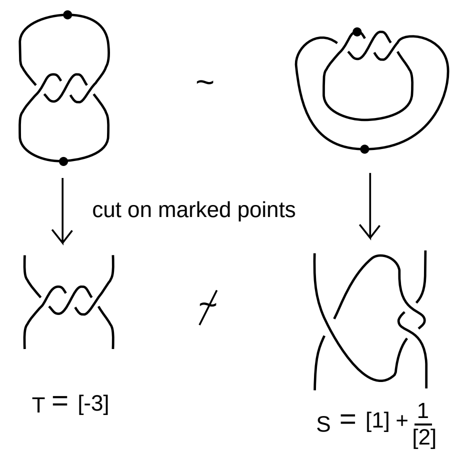


Figure 14: An Example of the Special Cut

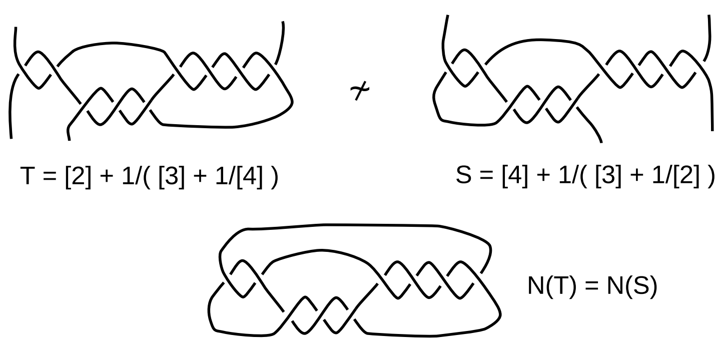


Figure 15: An Instance of the Palindrome Equivalence

$$T = [2,3,4] = [2] + \frac{1}{[3] + \frac{1}{[4]}}$$

and

$$S = [4,3,2] = [4] + \frac{1}{[3] + \frac{1}{[2]}}$$

both have the same numerator closure. Their corresponding fractions are

$$F(T) = 2 + \frac{1}{3 + \frac{1}{4}} = \frac{30}{13} \quad \text{and} \quad F(S) = 4 + \frac{1}{3 + \frac{1}{2}} = \frac{30}{7}.$$

Note that $7 \cdot 13 \equiv 1 \bmod 30$, again confirming Theorem 2.

In the general case if $T = [a_1, a_2, \ldots, a_n]$, we shall call the tangle $S = [a_n, a_{n-1}, \ldots, a_1]$ *the palindrome of* T. Clearly these tangles have the same numerator. In order to check the arithmetic in the general case of the palindrome cut we need to generalize this pattern to arbitrary continued fractions and their palindromes (obtained by reversing the order of the terms).

Theorem 3 below is a known result about continued fractions. See [27], [42] or [29]. We shall give here our proof of this statement. For this we will first present a way of evaluating continued fractions via 2×2 matrices (compare with [17], [33]). This method of evaluation is crucially important for the rest of the paper. We define matrices $M(a)$ by the formula

$$M(a) = \begin{pmatrix} a & 1 \\ 1 & 0 \end{pmatrix}.$$

These matrices $M(a)$ are said to be the *generating matrices* for continued fractions, as we have:

Theorem 3 (The Palindorome Theorem). *Let* $\{a_1, a_2, \ldots, a_n\}$ *be a collection of* n *integers, and let*

$$\frac{P}{Q} = [a_1, a_2, \ldots, a_n]$$

and

$$\frac{P'}{Q'} = [a_n, a_{n-1}, \ldots, a_1].$$

Then $P = P'$ *and* $QQ' \equiv (-1)^{n+1} \bmod P$.

Moreover, for any sequence of integers $\{a_1, a_2, \ldots, a_n\}$ *the value of the corresponding continued fraction*

$$\frac{P}{Q} = [a_1, a_2, \ldots, a_n]$$

is given through the following matrix product

$$M = M(a_1)M(a_2)\ldots M(a_n)$$

via the identity

$$M = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix}$$

where this matrix also gives the evaluation of the palindrome continued fraction

$$[a_n, a_{n-1}, \ldots, a_1] = \frac{P}{Q'}.$$

Proof. Let
$$\frac{R}{S} = [a_2, a_3, \ldots, a_n].$$
Then
$$\frac{P}{Q} = [a_1, a_2, \ldots, a_n] = a_1 + \frac{1}{\frac{R}{S}} = a_1 + \frac{S}{R} = \frac{Ra_1 + S}{R}.$$
By induction we may assume that
$$M(a_2)M(a_3)\ldots M(a_n) = \begin{pmatrix} R & S' \\ S & V \end{pmatrix}.$$
Hence
$$M(a_1)M(a_2)\ldots M(a_n) = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} R & S' \\ S & V \end{pmatrix} = \begin{pmatrix} a_1R + S & a_1S' + V \\ R & S' \end{pmatrix}.$$
This proves by induction that
$$M = M(a_1)M(a_2)\ldots M(a_n) = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix}.$$
To see the result about the palindrome continued fraction (and the terms P and Q' in the matrix), let M^T denote the transpose of M. Then
$$M^T = M(a_n)M(a_{n-1})\cdots M(a_1),$$
from which the statement about the palindrome follows by a repeat of the above calculation. Note also that $Det(M) = (-1)^n$ since M is a product of n matrices of determinant equal to -1, and $Det(M) = PU - QQ'$. Hence
$$PU - QQ' = (-1)^n.$$
This last equation implies the congruence stated at the beginning of the Theorem, and completes the proof of the Theorem. □

4 The Return of the Culprit

In order to see the Culprit in a way that allows us to understand why it is unknotted, we shall use the language and technique of the theory of tangles as explained in the previous section.

View Figure 16. In Figure 16 we have a drawing C of the Culprit and next to it we have a drawing C' of the result of part of Figure 3, where the undercrossing arc in C has been moved (making the complexity rise) until it has been drawn outside the rest of the diagram. Concentrate on C'. Notice that we can cut C' into two pieces, as shown in Figure 16. These two pieces, A and B, are rational tangles, and this cutting process shows that C' is the numerator closure

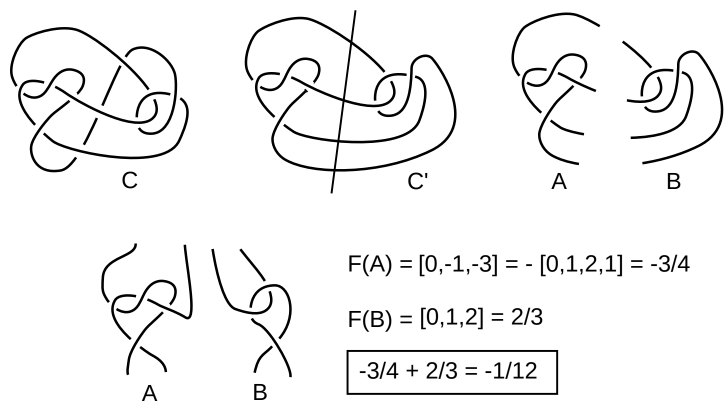


Figure 16: The Culprit Analysis

of the tangle sum of A and B. This is written $C' = N(A+B)$. Each rational tangle T has a rational fraction $F(T)$ that tells all about it. In this case,

$$F(A) = \frac{1}{-1 + \frac{1}{-3}} = -3/4$$

and

$$F(B) = \frac{1}{1 + \frac{1}{2}} = 2/3.$$

We know that the numerator of $A + B$ is unknotted and we would like to understand why it is unknotted. We notice that the sum of the fractions of A and B is

$$F(A) + F(B) = -3/4 + 2/3 = -1/12.$$

Thus the numerator of the sum of the fractions of A and B is -1. Does this -1 imply the unknottedness of the numerator of $A + B$? Well, the answer is that it does, and that will be the subject of much of the rest of the paper. See Theorem 6.

5 Collapsing to Unknots and Unlinks

In this section we first prove that the numerator of the sum of two rational tangles is a rational knot or link. We identify the knot or link, and use the result to characterize in Theorem 6 those unknots and unlinks that arise as numerators of the sum of two rational tangles.

Theorem 4 (Addition of Rational Tangles). *Let* $\frac{P}{Q} = [a_1, a_2, \ldots, a_n]$ *and* $\frac{R}{S} = [b_1, b_2, \ldots, b_m]$*. Let* $A = [\frac{P}{Q}]$ *and* $B = [\frac{R}{S}]$ *be the corresponding rational*

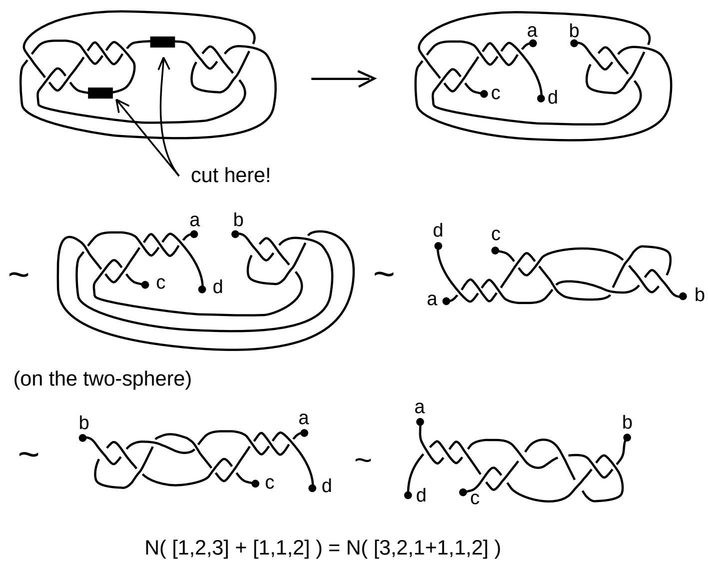


Figure 17: The Numerator of a Sum of Rational Tangles is a Rational Link

tangles. Then the knot or link $N(A+B)$ is rational. In fact

$$N(A+B) = N([a_n, a_{n-1}, \ldots, a_2, a_1 + b_1, b_2, \ldots, b_m]).$$

Proof. View Figure 17. In this figure we illustrate a special case of the Theorem. The general case is clear from this illustration. □

Definition 3. Given continued fractions $\frac{P}{Q} = [a_1, \ldots, a_n]$ and $\frac{R}{S} = [b_1, \ldots, b_m]$, let

$$[a_1, \ldots, a_n] \sharp [b_1, \ldots, b_m] := [a_n, \ldots, a_2, a_1 + b_1, b_2, \ldots, b_m].$$

If

$$\frac{F}{G} = [a_n, \ldots, a_2, a_1 + b_1, b_2, \ldots, b_m],$$

we shall write

$$\frac{P}{Q} \sharp \frac{R}{S} = \frac{F}{G}.$$

By Theorem 4, $N([\frac{P}{Q} \sharp \frac{R}{S}]) = N([\frac{P}{Q}] + [\frac{R}{S}])$.

Remark 2. The reader should note that $N([\frac{P}{Q}]+[\frac{R}{S}])$ is *not*, in general, isotopic to $N([\frac{P}{Q} + \frac{R}{S}]) = N([\frac{PS+QR}{QS}])$. This is the whole point of the previous Theorem. When we add two rational tangles we do not in general get a rational tangle, but the closure of this sum is a rational knot. This rational knot is the closure of a rational tangle distinct from $[\frac{P}{Q} + \frac{R}{S}]$. A good example is $P/Q = [3, 2, 1] = 10/3$ and $R/S = [2, 2, 1] = 7/3$. Then $P/Q + R/S = 51/9$, while $(P/Q)\sharp(R/S) =$

$[1, 2, 3+2, 2, 1] = 51/35$. We have $N([10/3]+[7/3]) = N([51/35])$, which is not isotopic to $N([51/9])$ by the Schubert Theorem. Nevertheless, the numerator of $(P/Q)\sharp(R/S)$ is equal to $PS + QR$. This is the import of the next Theorem.

Theorem 5. *If P/Q has matrix*

$$M = M(\vec{a}) = M(a_1)\ldots M(a_n)$$

and R/S has matrix

$$N = M(\vec{b}) = M(b_1)\ldots M(b_m),$$

then $[a_1, \ldots, a_n]\,\sharp[b_1, \ldots, b_m]$ has matrix

$$M\,\sharp\,N := M^T N^E,$$

where N^E denotes the matrix obtained by interchanging the rows of N. This gives an explicit formula for $[a_1, \ldots, a_n]\,\sharp\,[b_1, \ldots, b_m]$. This formula can be used to determine not only when $N([\frac{P}{Q}]+[\frac{R}{S}])$ is unknotted but also to find its knot type as a rational knot via Schubert's Theorem. In particular, we find that

$$N([\frac{P}{Q}]+[\frac{R}{S}]) = N([\frac{PS+QR}{Q'S+UR}]) = N([\frac{Num(P/Q+R/S)}{Num(Q'/U+R/S)}])$$

where $|PU - QQ'| = 1$. In other words, if we take the numerator closure of the sum of two rational tangles, this knot or link is equal to the numerator of a rational tangle whose arithmetical numerator is the artihmetical numerator of the sum of the fractions of the original two rational tangles.

Proof. We know from Theorem 3 that

$$M(\vec{a}) = M(a_1)\ldots M(a_n) = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix},$$

and that

$$M(\vec{b}) = M(b_1)\ldots M(b_m) = \begin{pmatrix} R & S' \\ S & V \end{pmatrix}.$$

Let

$$F/G = [a_n, \ldots, a_2, a_1+b_1, b_2, \ldots, b_m] = \frac{P}{Q}\sharp\frac{R}{S}.$$

Then we have, by Theorem 4, that

$$N([\frac{P}{Q}]+[\frac{R}{S}]) = N([\frac{F}{G}]),$$

and

$$M(\vec{c}) = M(a_n)\ldots M(a_2)M(a_1+b_1)M(b_2)\ldots M(b_m) = \begin{pmatrix} F & G' \\ G & W \end{pmatrix}.$$

Now note the identity

$$\left(\begin{array}{cc} a_1 & 1 \\ 1 & 0 \end{array}\right)\left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right)\left(\begin{array}{cc} b_1 & 1 \\ 1 & 0 \end{array}\right) = \left(\begin{array}{cc} a_1+b_1 & 1 \\ 1 & 0 \end{array}\right).$$

Thus

$$M(\vec{c}) = M(\vec{a})^T \left(\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right) M(\vec{b}) = M(\vec{a})^T M(\vec{b})^E$$

where M^E denotes the matrix obtained from M by interchanging its two rows. In particular, this formula implies that

$$\left(\begin{array}{cc} F & G' \\ G & W \end{array}\right) = \left(\begin{array}{cc} P & Q \\ Q' & U \end{array}\right)\left(\begin{array}{cc} S & V \\ R & S' \end{array}\right) = \left(\begin{array}{cc} PS+QR & PV+QS' \\ Q'S+UR & Q'V+US' \end{array}\right).$$

Thus

$$N([P/Q]+[R/S]) = N([(PS+QR)/(Q'S+UR)]) = N([\frac{Num(P/Q+R/S)}{Num(Q'/U+R/S)}])$$

where $|PU-QQ'|=1$. This completes the proof of the Theorem. □

Theorem 6 (Unknot/Unlink Theorem). *Let $\frac{P}{Q}=[a_1,\ldots,a_n]$ and $\frac{R}{S}=[b_1,\ldots,b_m]$ be as in Theorem 5. Then $N([\frac{P}{Q}]-[\frac{R}{S}])$ is unknotted if and only if $PS-QR=\pm 1$, that is, PS and QR are consecutive integers. For links we have that $N([\frac{P}{Q}]-[\frac{R}{S}])$ is unlinked if and only if $P/Q=R/S$.*

Proof. It follows from Theorem 5 and the Schubert Theorem that $N([\frac{P}{Q}]-[\frac{R}{S}])$ is unknotted if and only if $PS-QR=\pm 1$. By the same token, for a link we have that $N([\frac{P}{Q}]-[\frac{R}{S}])$ is unlinked if and only if $PS-QR=0$. Since Q and S are non-zero, it follows that $P/Q=R/S$. □

Remark 3. Suppose, as in Theorem 6, that $|PS-QR|=1$ so that $N([\frac{P}{Q}]-[\frac{R}{S}])$ is an unknot. Then it follows from Theorem 6 that $N([\frac{Q}{P}]-[\frac{S}{R}])$ is also an unknot. Thus, the pair of fractions $(\frac{P}{Q},\frac{R}{S})$ yields an unknot if and only if the pair of inverse fractions $(\frac{Q}{P},\frac{S}{R})$ yields an unknot. In fact, the reader will enjoy proving the following identity for any rational tangles A and B :

$$N(\frac{1}{A}+\frac{1}{B}) = -N(A+B)$$

where the minus sign denotes taking the mirror image of the diagram $N(A+B)$, and equality is isotopy. Thus, for any pair of fractions $(\frac{P}{Q},\frac{R}{S})$ (not necessarily forming an unknot)

$$N([\frac{P}{Q}]-[\frac{R}{S}]) = -N([\frac{Q}{P}]-[\frac{S}{R}]).$$

Remark 4. Our proof of this Theorem is elementary, based on rational tangle calculus. The conclusions of the Theorem can be reached by more sophisticated means as the reader will see in the papers [31, 32]. These papers are concerned with the question of the unknotting number of rational knots and links, and they solve the problem for unknotting number one.

We end this section with a way to produce infinitely many unknots. Consider a rational number, its corresponding continued fraction, and its matrix representation:

$$P/Q = [a_1, \ldots, a_n]$$

with

$$M = M(\vec{a}) = M(a_1) \ldots M(a_n) = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix}.$$

Note that since the determinant of this matrix is $(-1)^n$, we have the formula $PU - QQ' = (-1)^n$ from which it follows that

$$P/Q - Q'/U = (-1)^n/QU.$$

Hence, by Theorem 6, the rational knot

$$N([P/Q] - [Q'/U])$$

is unknotted and, as we shall see later, is a good candidate to produce a hard unknot.

6 Continued Fractions, Convergents and Lots of Unknots

In this section we'll show that we can do even better than the example above for producing infinitely many unknots. Remarkably, the fraction Q'/U (at the end of the last section) has an interpretation as the truncation of our continued fraction $[a_1, \ldots, a_n]$.

Lemma 1. *Let $P/Q = [a_1, \ldots, a_n]$ with corresponding matrix $M = M(\vec{a}) = M(a_1) \ldots M(a_n) = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix}$. Then $Q'/U = [a_1, \ldots, a_{n-1}]$. Conversely, if $R/S = [a_1, \ldots, a_{n-1}]$ is a reduced fraction, then $|R| = |Q'|$ and $|S| = |U|$. So, we can choose signs for R and S so that*

$$\begin{pmatrix} P & R \\ Q & S \end{pmatrix} = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix}.$$

Proof. To see this formula, let

$$N = M(a_1) \ldots M(a_{n-1}) = \begin{pmatrix} R & S' \\ S & V \end{pmatrix},$$

so that

$$R/S = [a_1, \ldots, a_{n-1}].$$

Then

$$\begin{pmatrix} P & Q' \\ Q & U \end{pmatrix} = M(a_1) \ldots M(a_{n-1})M(a_n) = NM(a_n)$$

$$= \begin{pmatrix} R & S' \\ S & V \end{pmatrix} \begin{pmatrix} a_n & 1 \\ 1 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} Ra_n + S' & R \\ Sa_n + V & S \end{pmatrix}.$$

This shows that $Q'/U = R/S = [a_1, \ldots, a_{n-1}]$, as claimed. For the converse, note that the pairs $(P, Q), (Q', U), (R, S)$ are pairs of relatively prime integers. Note also that $Q'/U = R/S$. Thus $|R| = |Q'|$ and $|S| = |U|$. $\square$

Definition 4. One says that $[a_1, \ldots, a_{n-1}]$ is a *convergent* of $[a_1, \ldots, a_{n-1}, a_n]$. We shall say that two fractions P/Q and R/S are *convergents* if the continued fraction of one of them is a convergent of the other.

We have, so far, proved the following result.

Proposition 2. *Let $P/Q = [a_1, \cdots, a_{n-1}, a_n]$ and let $Q'/U = [a_1, \cdots, a_{n-1}]$, a convergent of P/Q. Then*

$$N([P/Q] - [Q'/U]) = N([a_1, \cdots, a_{n-1}, a_n] - [a_1, \cdots, a_{n-1}])$$

is an unknot.

Remark 5. Another, more direct way to see this result is illustrated in Figure 18. In this figure we show $N([2, 3, 4] - [2, 3])$ with the second tangle written as a 180^0 turn of the standard form for $[2, 3]$ (this is isotopic to the standard form since we are working with rational tangles). In this form, we see clearly that the braidings of the two tangles cancel in the back of the 2-sphere to form an unknot.

The reader should note that

$$N([a_1, \ldots, a_n] - [a_1, \ldots, a_{n-1}]) = N([a_1, \ldots, a_n]\sharp[-a_1, \ldots, -a_{n-1}])$$

$$= N([a_n, \ldots, a_2, a_1 - a_1, \ldots, -a_{n-1}]) = N([a_n, \ldots, a_2, 0, -a_2, \ldots, -a_{n-1}])$$

$$= N([a_n, \ldots, a_2 - a_2, \ldots, -a_{n-1}]) = \cdots$$

$$= N([a_n, a_{n-1} - a_{n-1}]) = N([a_n, 0]) = U$$

where U is the unknotted circle. Thus we see directly that a continued fraction and the negative of its convergent will yield an unknot when we take the numerator of the sum of the corresponding tangles.

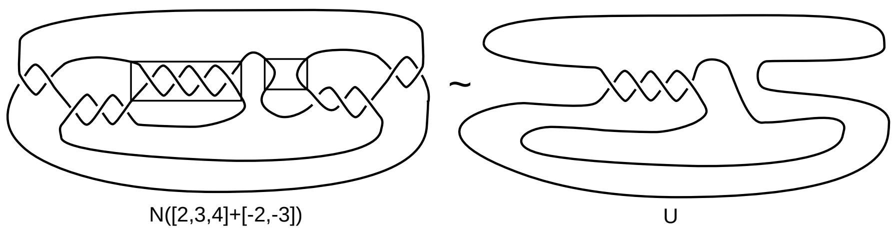


Figure 18: Braid Cancellation to an Unknot

We see from the above that the two consecutive integers PU and QQ' produce two continued fractions $P/Q = [a_1, \ldots, a_n]$ and $Q'/U = [a_1, \ldots, a_{n-1}]$ so that the second fraction is a convergent of the first. *The property of one fraction being a convergent of the other is in fact, always a property of fraction pairs produced from consecutive integers.* For example, consider the integers 90 and 91. We have $90 = 10 \times 9$ and $91 = 13 \times 7$. Thus

$$13 \times 7 - 10 \times 9 = 1.$$

Hence

$$13/9 - 10/7 = 1/63,$$

so that $N([13/9] + [-10/7])$ is an unknot by Theorem 6. We see that

$$13/9 = [1, 2, 4] = [1, 2, 3, 1]$$

and

$$10/7 = [1, 2, 3]$$

so that 13/9 and 10/7 are convergents. We will now prove this statement about convergents. For the next Lemma see also [15].

Lemma 2. *Let P and Q be relatively prime integers and let s and r be a pair of integers such that $Ps - Qr = \pm 1$. Let $R = r + tP$ and $S = s + tQ$ where t is any integer. Then $\{R, S\}$ comprises the set of all solutions to the equation $PS - QR = \pm 1$. Similarly, all solutions of the equation $PS - QR = \mp 1$ are given in the form $R = -r + tP$ and $S = -s + tQ$.*

Proof. Without loss of generality we can assume that $Ps - Qr = 1$. We leave it to the reader to formulate the case where $Ps - Qr = -1$. Certainly R and S as given in the statement of the Lemma satisfy the equation $PS - QR = 1$. Suppose that R and S is some solution to this equation. Then it follows by taking the difference with the equation $Ps - Qr = 1$ that $P(S - s) = Q(r - R)$. Since P and Q are relatively prime, it follows at once from this equation and the uniqueness of prime factorization of integers that $R = r + tP$ and $S = s + tQ$ for some integer t. The last part of the Lemma follows by the same form of reasoning. This completes the proof. $\square$

Lemma 3. *Let P and Q be relatively prime integers and let $P/Q = [a_1, \ldots, a_n]$ be a continued fraction expansion for P/Q. Let also $r/s = [a_1, \ldots, a_{n-1}]$ be a convergent for $[a_1, \ldots, a_n]$. Let $R = r + tP$ and $S = s + tQ$ where t is any integer. Then $R/S = [a_1, \ldots, a_n, t]$, and thus P/Q is a convergent of R/S.*

Proof. By Lemma 1, we have

$$M(\vec{a}) = M(a_1) \ldots M(a_n) = \begin{pmatrix} P & r \\ Q & s \end{pmatrix}.$$

Moreover, we have:

$$M(\vec{a})M(t) = \begin{pmatrix} P & r \\ Q & s \end{pmatrix} \begin{pmatrix} t & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} Pt + r & P \\ Qt + s & Q \end{pmatrix}.$$

By Lemma 2 for $\{R = r + tP, S = s + tQ\}$ we have $PS - QR = (-1)^{n+1}$. So $R/S = [a_1, \ldots, a_n, t]$, and P/Q is a convergent of R/S. Thus all such solutions correspond to continued fractions of which P/Q is a convegent. This completes the proof. □

Finally, we can sum up by pointing out that we have proved the following result.

Theorem 7 (Unknotting Convergents). *Let P/Q and R/S be reduced fractions. Then*

$$N([P/Q] - [R/S])$$

is an unknot if and only if P/Q and R/S are convergents.

Remark 6. In this sense Theorems 6 and 7 show us that the simplest method of collapse to the unknot is the only way that the collapse can occur when we take the numerator of the sum of two rational tangles. The reader will enjoy verifying that this class of unknots can be obtained from a closed loop of rope by consecutive twisting of adjacent strands of the rope, followed by a separation of the rope into the numerator sum of two rational tangles. Such processes occur naturally in situations involving DNA topology as described later in this paper.

Infinite Continued Fractions. Note that if we start with an infinite continued fraction

$$[a_1, a_2, a_3, \ldots]$$

and define the *truncations*

$$P_n/Q_n = [a_1, \ldots, a_n],$$

then we have

$$P_n/Q_n - P_{n-1}/Q_{n-1} = (-1)^n/Q_n Q_{n-1}.$$

This equation is usually used to analyze the convergence of the truncations of the continued fraction to a real number. Here we see a *single infinite continued fraction*, whether or not convergent, *producing an infinite family of unknot diagrams*:

$$K_n = N([P_n/Q_n] - [P_{n-1}/Q_{n-1}]).$$

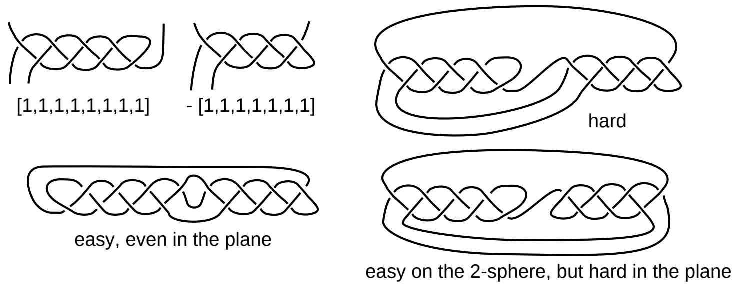


Figure 19: $N([f_8/f_7] - [f_7/f_6]) = N([34/21] - [21/13])$

Surely, the simplest instance of this phenomenon occurs with the infinite continued fraction for the Golden Ratio:

$$\frac{1+\sqrt{5}}{2} = [1, 1, 1, \ldots].$$

Here the convergents are $f_n/f_{n-1} = [1, 1, 1, \ldots, 1]$ (with n 1's) and f_n is the n-th Fibonacci number where $f_0 = 1, f_1 = 1$ and $f_{n+1} = f_n + f_{n-1}$ for $n \geq 1$. The matrices are

$$M(1)^n = \begin{pmatrix} f_n & f_{n-1} \\ f_{n-1} & f_{n-2} \end{pmatrix}.$$

Thus we have

$$f_n/f_{n-1} - f_{n-1}/f_{n-2} = (-1)^n/f_{n-1}f_{n-2}.$$

In this instance, this formula shows that the sequence of truncations is convergent (to the Golden Ratio).

In Figure 19 we illustrate a hard unknot based on the eighth and seventh truncations of the Golden Ratio continued fraction. This example is produced by using the standard closure of $N([f_8/f_7] - [f_7/f_6]) = N([34/21] - [21/13])$. By this same pattern, we can extract from the infinite continued fraction for the golden ratio an infinite family of hard unknots. Note the unavailability of Reidemeister moves on the first diagram. In Figure 19 the first diagram illustrates the numerator of the sum of the two tangles $[f_8/f_7] = [1, 1, 1, 1, 1, 1, 1, 1]$ and $-[f_7/f_6] = [-1, -1, -1, -1, -1, -1, -1]$ with each tangle in 3-strand braid form with the integer 1-twist on the left. In this form the diagram is a hard unknot, and this generalizes to infinitely many examples of the same type. The second example has one tangle turned by 180^0 so that the integer 1-twists face one another, giving an easy unknot. The third example has one tangle turned by 180^0 so that both integer 1-twists face into the outer region. This diagram is an easy unknot on the 2-sphere, since there is a Reidemeister II-move in the outer region. If we were to restrict moves to the plane, then this Reidemeister move would not be available. The third diagram is a hard unknot in the plane.

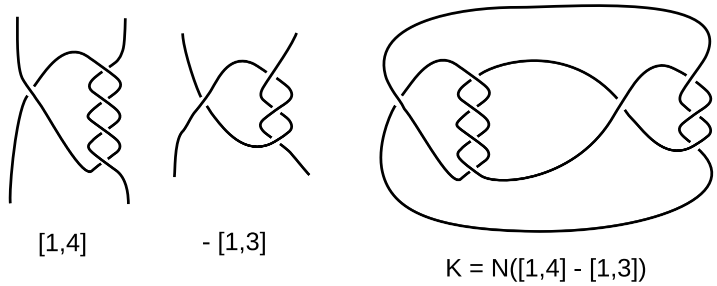


Figure 20: $K = N([1,4] - [1,3])$

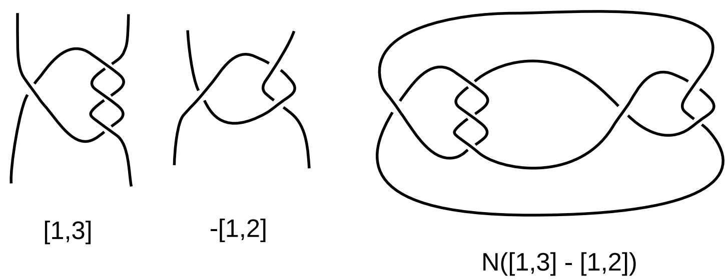


Figure 21: $N([1,3] - [1,2])$

7 Constructing Hard Unknots

In this section we indicate how to construct hard unknots by using positive alternating tangles A and B such that $N(A - B)$ is unknotted. By our main results we know how to construct infinitely many such pairs of tangles by taking a continued fraction and its convergent, with the corresponding tangles in reduced (alternating) form.

Let's begin with the case of $5/4 = [1,4] = [1,3,1]$ and $4/3 = [1,3]$. In Figure 20 we show the standard representations of $[1,4]$ and $-[1,3]$ as tangles, and $K = N([1,4] - [1,3])$. The reader will note that this diagram is a hard unknot with 9 crossings, one less than our original Culprit of Figure 16. In Figure 24 we give a diagram H that is isotopic to the mirror image of K. In Theorem 8 we show that H is one of a small collection of minimal hard unknot diagrams having the form $N(A-B)$ for reduced positive rational tangle diagrams A and B. In most cases, if one takes the standard representations of the tangles A and B, and forms the diagram for $N(A - B)$, the resulting unknot diagram will be hard. There are some exceptions however, and the next example illustrates this phenomenon.

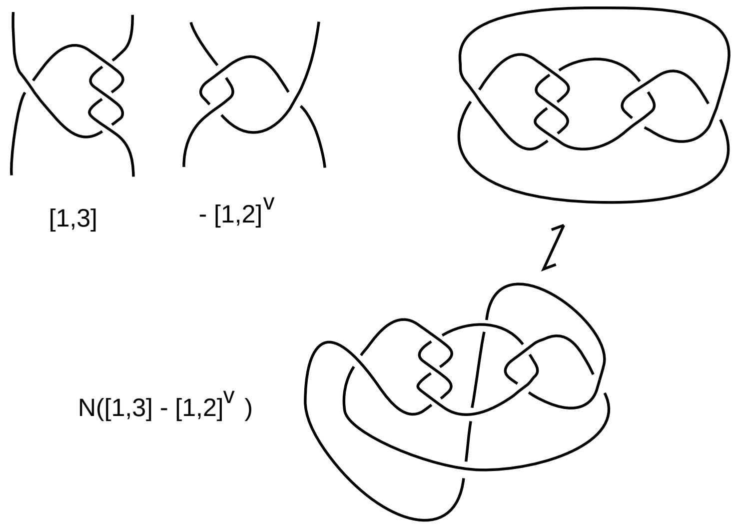


Figure 22: $N([1,3] - [1,2]^v)$

In Figure 21 we show the standard representations of $[1,3]$ and $[1,2]$ as tangles, and the diagram of $N([1,3] - [1,2])$. This diagram, while unknotted, is not a hard unknot diagram due to the three-sided outer region. This outer region allows a type III Reidemeister move on the surface of the 2-sphere.

The Tucking Construct. In Figure 22 we have replaced $[1,2]$ by $[1,2]^v$, the 180^0 turn of the tangle $[1,2]$ about the vertical direction in the page. Now we see that the literal diagram of $N([1,3] - [1,2]^v)$ is of course still unknotted and is also not a hard unknot diagram. However this diagram can be converted to a hard unknot diagram by tucking an arc as shown in the Figure. The resulting hard unknot is the same diagram of 10 crossings that we had in Figure 16 as our initial Culprit. Note that a direct tuck on Figure 21 does not work (there would be a type III move made available). Note also that the other possibilities of flipping both tangles in Figure 21 or flipping the first tangle do not lead to hard unknots. We call this strategem the *tucking construct.* The tucking prevents the appearance of a type II move in the outermost region of the diagram.

The Culprit Revisited. Let's consider the example in Figure 16 again. Here we have $P/Q = -3/4$ and $R/S = 2/3$. We have $P/Q + R/S = -3/4 + 2/3 = -1/12$. Thus $N([-3/4] + [2/3])$ is an unknot by Theorem 6. This is exactly the unknot C' illustrated in Figure 16. Note that C is the tucking construct applied to C'.

We can make infinitely many examples of this type. View Figure 23. The pattern is as follows. Suppose that $T = [P/Q]$ and $T' = [R/S]$ are rational

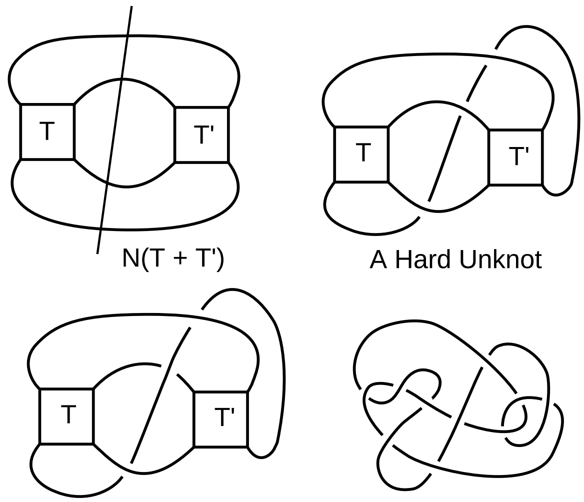


Figure 23: The Tucking Construct

tangles such that $PS + QR = \pm 1$. Then we know that $N(T + (T')^v)$ is an unknot. Furthermore we can assume that each of the tangles T and T' are in alternating form. The two tangle fractions have opposite sign and hence the alternation of the weaves in each tangle will be of opposite type. We create a new diagram for $N(T + (T')^v)$ by putting an arc from the bottom of the closure entirely underneath the diagram as shown in Figure 23. This is the general form of the tucking construct.

Remark 7. View the bottom of Figure 23 and note that we have indicated that the hard unknot becomes an alternating and hence knotted diagram (this can be proved by applying the results in [25]) by switching one crossing in the tucking construction. Thus the tucking construction is also an infinite source of knots of unknotting number one.

Hard Unknots Recipe. In a tangle $[a_1, \ldots, a_n]$, call a_1 the *integer part* of the tangle. Here we use tangles $T = [a_1, \ldots, a_n]$ and $T' = [-a_1, \ldots, -a_{n-1}]$ with all a_i positive. Except for very few cases with small number of total crossings, the following hold:

1. If the integer parts both face in the same direction, then $N(T + T')$ is a hard unknot diagram.
2. If the integer parts both face the outside region, thus making a type II move available on the 2-sphere, then the tucking construct will create a hard unknot.

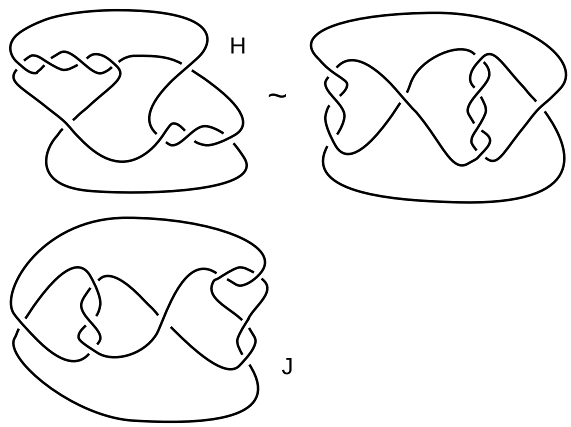


Figure 24: Smallest hard unknots

3. If the integer parts both face the inside region, then we do not get a hard unknot and we cannot remedy this fact using tucking.

The reader should compare these remarks to the contents of Figure 19.

8 The Smallest Hard Unknots

Figure 24 illustrates two hard unknot diagrams H and J with 9 crossings ($K = -H$ appears earlier in Figure 20).

Two equivalent versions of the diagram H appear in Figure 24. The right-hand version of H in this figure is of the form:

$$H = N([1,3] - [1,4]) = N([4/3] - [5/4]).$$

Note that $[1,3]$ and $[1,4] = [1,3,1]$ are convergents. Thus H and K are mirror images of each other. The diagram J in Figure 24 is of the form:

$$J = N([1,3] - [1,2,2]) = N([4/3] - [7/5]).$$

Note that $[1,3] = [1,2,1]$ and $[1,2,2] = [1,2,1,1]$ are convergents. Note also that the crossings in J corresponding to 1 in $[1,3]$ and -1 in $[-1,-2,-2]$ can be switched and we will obtain another diagram J', arising as sum of two alternating rational tangles, that is also a hard unknot. This diagram can be obtained from the diagram J without switching crossings, by performing flypes (Figure 12) on the subtangles $[1,3]$ and $[-1,-2,-2]$ of J, and then doing an isotopy of this new diagram on the 2-sphere. (We leave the verification of this statement to the reader.) Thus the diagram J' can be obtained from J by flyping. A similar remark applies to the diagram H, giving a corresponding diagram H', but in this case H' is easily seen to be equivalent to H by a planar isotopy that does not involve any Reidemeister moves. Thus, up to these sorts of modifications, we have produced essentially two hard unknot diagrams with 9 crossings.

We have the following result.

Theorem 8. *The diagrams H and J shown in Figure 24 are, up to flyping subtangle diagrams and taking mirror images, the smallest hard unknot diagrams in the form $N(A-B)$ where A and B are rational tangles in reduced positive alternating form.*

Proof. It is easy to see that we can assume that $A = [P/Q]$ where P and Q are positive, relatively prime and P is greater than Q. We leave the proof that one can choose P greater than Q to the reader.

We then know from Theorem 7 that $B = [-R/S]$ where one of P/Q and R/S is a convergent of the other. We can now enumerate small continued fractions. We know the total of all terms in A and B must be less than or equal to 9 since H and J each have nine crossings.

In order to make a 9-crossing unknot example of the form $N(A-B)$ where A and B are rational tangles in reduced positive alternating form, we must partition the number 9 into two parts corresponding to the number of crossings in each tangle. It is not hard to see that we need to use the partition $9 = 4+5$ in order to make a hard unknot of this form. Furthermore, 4 must correspond to the continued fraction $[1,3]$, as $[2,2]$ will not produce a hard unknot when combined with another tangle. Thus, for producing 9 crossing examples we must take $A = [1,3]$. Then, in order that A and B be convergents, and B have 5 crossings, the only possibilities for B are $B = [1,4]$ and $B = [1,2,2]$. These choices produce the diagrams H, H', J, J'. It is easy to see that no diagrams with less than 9 crossings will suffice to produce hard unknots, due to the appearance of Reidemeister moves related to the smaller partitions. This completes the proof. $\square$

We made the following conjecture:

Conjecture 1. *Up to mirror images and flyping tangles in the diagrams, the hard unknot diagrams H and J of 9 crossings, shown in Figure 24, have the least number of crossings among all hard unknot diagrams.*

This conjecture has been verified for us by Slavik Jablan, using his program "Linknot" [23].

9 The Goeritz Unknot

The earliest appearance of a hard unknot is a 1934 paper of Goeritz [18]. In this paper Goeritz gives the hard unknot G shown in Figure 25. As the reader

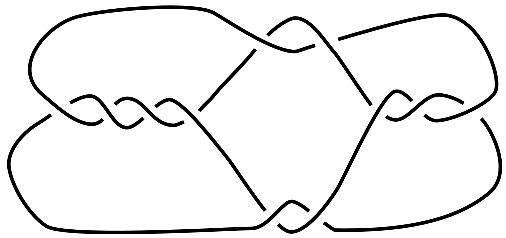


Figure 25: The Goeritz Hard Unknot

can see (for example by flyping vertically the tangle [-3] twice), this example is a variant on $N([4]+[-3])$ which is certainly unknotted. The Goeritz example has 11 crossings, due to the extra 2-twists that make it a hard unknot. It is part of an infinite family based on the pattern $N([n]+[-n+1])$.

In Figure 26 we illustrate the steps in another unknotting process for G. A move labeled *swing* consists in swinging the corresponding marked arc in the diagram. A swing move when realized by Reidemeister moves will complicate the diagram before taking it to the indicated simplification. After two swing moves G has been transformed into the diagram H of the previous section that is again a hard unknot diagram, but with 9 crossings rather than 11!

10 Recalcitrance Revisited

Suppose that $K[a]$ denotes the knot diagram shown in Figure 27, so that

$$K[a]=N(-[a]+S+[a]+T]).$$

We shall assume that $K[0]=N(S+T)$ is unknotted. We also assume that the diagram $K[a]$ is hard, a generalization of the form of the hard diagram H in Figure 24. Let's also assume that it takes N Reidemeister moves to transform $[-1]+S+[1]$ to S. This transformation is easily accomplished in three dimensions by one full rotation, but may require many Reidemeister moves in the plane (keeping the ends of the tangle fixed). It is certainly possible to produce alternating rational tangles S and T with this property that no less than N Reidemeister moves can accomplish an isotopy from $[-1]+S+[+1]+T$ to $S+T$. Let $C=C(S)+C(T)$ denote the total number of crossings in the tangles S and T. Then the recalcitrance (Definition 2) of $K[0]$ is $R(0)=k/C$ for some k and the recalcitrance of $K[a]$ is generically given by the formula

$$R(a)=\frac{aN+k}{2a+C}$$

since each extra turn of S will add N Reidemeister moves to the untying, and the number of crossings of $K[a]$ is equal to $2a+C$. We conclude from this that

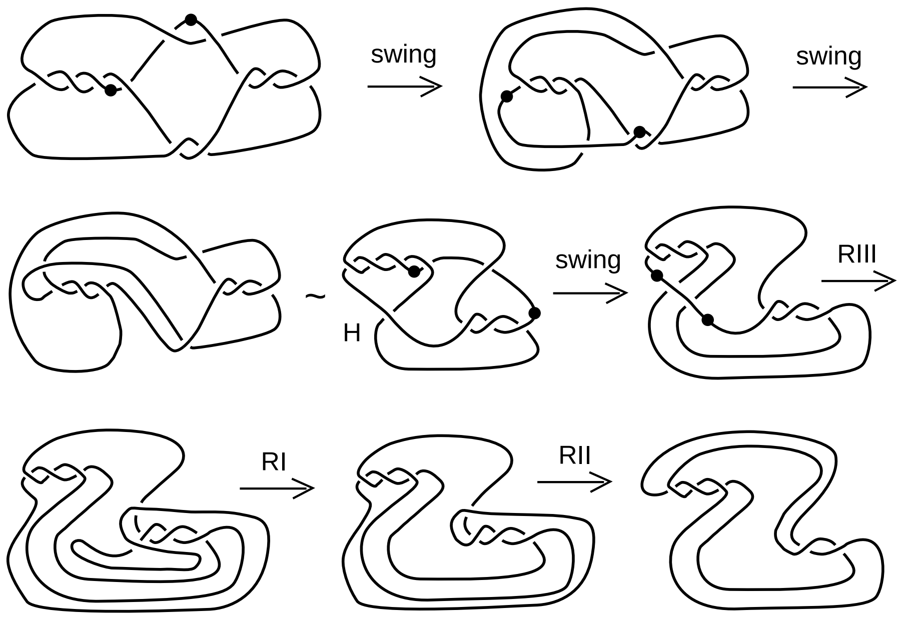
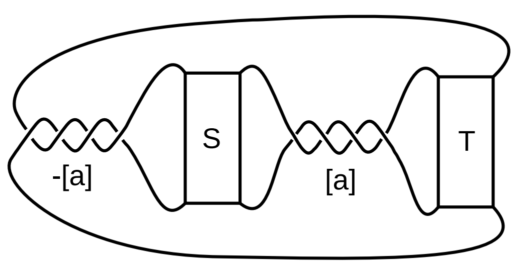


Figure 26: Unknotting the Goeritz Unknot



K[a] = N(-[a] + S + [a] + T) = N(S + T)

Figure 27: Twisting Up the Recalcitrance

for large a the recalcitrance $R(a)$ is as close as we like to $N/2$. This shows that *the recalcitrance of an unknotted diagram can be arbitrarily large.* There is no upper limit for the ratio of the number of moves needed to undo the knot in relation to the number of crossings in the original knot diagram.

11 Collapsing to Knots and Links

In this section we point out that our results about unknotted closures of sums of rational tangles can be generalized to collapses to arbitrary knots and links. This leads to the concept of a *hard knot diagram*, generalizing our notion of a hard unknot diagram. We shall say that a knot or link diagram K is *hard* if it admits no type III moves and has no simplifying type I or type II moves. Any sequence of Reidemeister moves that takes this diagram to a smaller diagram for that knot or link must first increase the number of crossings of K. The concepts of this paper, such as recalcitrance, can be generalized to this domain. We can ask, given any diagram K for the least number of Reidemeister moves that will transform K to some minimal diagram for the knot type of K. Many questions can be asked at this juncture. In this section we formulate some collapse results and give examples. We encourage the reader to explore further!

It is worth putting these results in perspective. In this paper we are exploring the combinatorial space of all knot and link diagrams by using the Reidemeister moves. Hard unknots and hard knots are local minima in this combinatorial space. The problem with a local minimum is that it may not be a global minimum. Hard knots can often be made smaller (we want a minimal diagram for the knot), but again must be made larger before becoming smaller. Each time we find a local minimum we may have to face the complex possibilities involved in climbing up out of it before things can be simplified. These features are fundamental to the study of knots and links via diagrams.

Notation. Let R be *any* 2-tangle. We shall write

$$S = [a_1, \ldots, a_n, R]$$

to denote the tangle

$$[a_1] + \frac{1}{[a_2] + \frac{1}{\cdots + \frac{1}{[a_n] + \frac{1}{R}}}}.$$

Note that we will occasionally replace $[a]$ by a in the notation for continued fraction forms of tangles. Note that if R is an arbitrary tangle, then $(R^{rot})^{rot} = R^{\pi}$ where R^{π} denotes the result of turning R by 180^0 around an axis perpendicular to the page. In the case of rational tangles R, we know that $R^{\pi} = R$, but in the general case the tangle and its 180^0 rotate may be distinct topologically. This means that $1/(1/R) = R^{\pi}$ in the general case. Similar remarks apply to R^v, the result of rotating R about the vertical axis and R^h, the result of rotating R about the horizontal axis. Many of the isotopies that we use for rational tangles may replace R by rotations of R by 180^0 about horizontal or vertical axes or both. Note that $R^\pi = R^{hv}$.

Definition 5. Letting $P/Q = [a_1, a_2, \ldots, a_n]$ and $V/W = [b_1, b_2, \ldots, b_m]$, it is convenient to make the following definitions for any tangles T and S :

1. $[P/Q]\sharp T = [a_n, a_{n-1}, \ldots, a_1 + T]$.
2. $[a_1, a_2, \ldots, a_n, T]\sharp[b_1, b_2, \ldots, b_n, S] = [T, a_n, a_{n-1}, \ldots, a_1 + b_1, b_2, \ldots, b_m, S]$

Note that this generalizes our previous definitions of the $\sharp$ operation. In fact, we have a generalization of Theorem 4, that the numerator closure of the sum of two rational tangles is a rational knot, to the following statement:

Theorem 9. *For any tangles T and S (not necessarily rational) the following equation expresses the numerator of the sum of the tangles indicated below. In this statement we have used the symbol $\hat{=}$ to mean that T and S on the right may differ from their counterparts on the left-hand side of the equation by 180^0 rotations (as discussed above). In the case of S and T rational or invariant under such rotations, the sign $\hat{=}$ may be replaced by $=$.*

$$N([a_1, a_2, \ldots, a_n, T] + [b_1, b_2, \ldots, b_n, S]) \hat{=} N([T, a_n, a_{n-1}, \ldots, a_1 + b_1, b_2, \ldots, b_m, S]).$$

In particular,

$$N([a_1, a_2, \ldots, a_n, T] + [-a_1, -a_2, \ldots, -a_n, S]) \hat{=} N(T + S).$$

In the special case where $T = [\infty]$ and $S = [0]$ this last equation becomes

$$N([a_1, a_2, \ldots, a_n] + [-a_1, -a_2, \ldots, -a_{n-1}]) = U$$

where U is unknotted, recovering Proposition 2. Similarly we have

$$N([a_1, \ldots, a_n + T] + [-a_1, \ldots, -a_n]) \hat{=} N(T)$$

and

$$N([a_1, \ldots, a_n, T] + [-a_1, \ldots, -a_n]) \hat{=} D(T).$$

Note that since T and S are arbitrary tangles, we can also regard this Theorem as a statement about black boxes called T and S that are connected as in Figure 28 to form rationally extended tangles and closures. This is the same as considering tangles that contain 4-valent rigid graphical vertices. Such graphs can be modelled as embeddings in three-dimensional space by attaching strings to rigid bodies (such as geometrical balls) in place of the vertices.

Proof. The proof follows from the isotopies indicated in Figure 28 and the identities $[a_1, \ldots, a_n, 0] = [a_1, \ldots, a_{n-1}]$ and $[a_1, \ldots, a_n, \infty] = [a_1, \ldots, a_n]$. $\square$

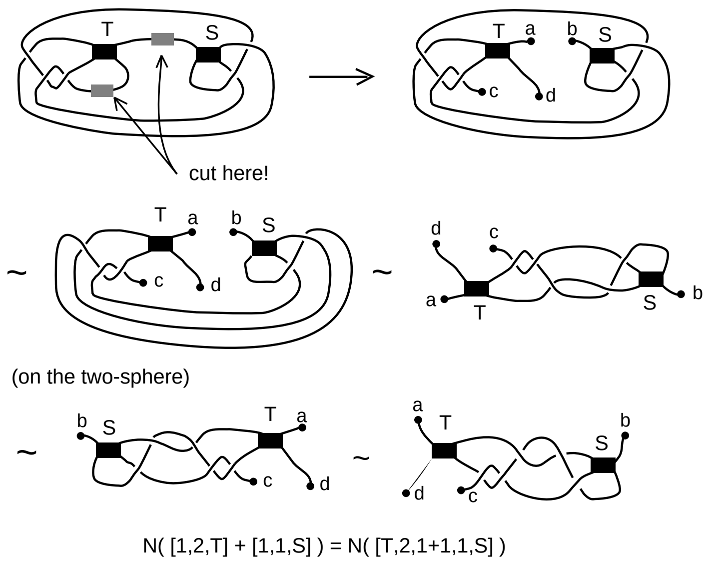


Figure 28: The Black Boxes

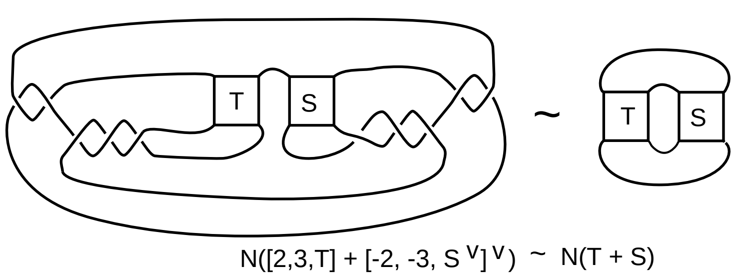


Figure 29: Black Boxes and Braid Cancellation

Remark 8. For T and S any 2-tangles, the cancellation to $N(T+S)$ in Theorem 9 can be stated more precisely as follows:

$$N([a_1, a_2, \ldots, a_n, T] + [-a_1, -a_2, \ldots, -a_n, S^v]^v) = N(T + S).$$

See Figure 29 for a direct illustration of this result. Note that this formula tells us that we have rational collapses to any knot or link, since we can always cut such a knot or link into one tangle or a sum of two tangles and embed these in cancelling rational patterns. Just as we have constructed hard unknot diagrams, we can construct hard *knot* diagrams.

The next Theorem is a direct consequence of Theorem 9. We leave its proof to the reader.

Theorem 10. *(i) Letting*

$$M = M(\vec{a}) = M(a_1)M(a_2)\ldots M(a_n) = \begin{pmatrix} P & Q' \\ Q & U \end{pmatrix},$$

and given F/G, a fraction in reduced form, then there exists a unique reduced fraction R/S such that

$$N([F/G]) = N([P/Q] + [R/S]),$$

and

$$\frac{F}{G} = \frac{Num(P/Q + R/S)}{Num(Q'/U + R/S)}$$

where $Num(a/b + c/d) = ad + bc$. Note that Q'/U is a convergent of P/Q, so we see that a combination of a fraction and its convergent appears in the decomposition of an arbitrary rational knot, fully generalizing our results about the unknot.

(ii) For any 2-tangle T,

$$N([P/Q] + [-P/Q']\sharp T) = N(T).$$

This is another generic form of the collapse result, and this one specializes directly to our unknot theorem when we take $T = [1/N]$ so that $N(T)$ is an unknot. In that case the fractions P/Q and $(P/Q')\sharp(1/N)$ are convergents.

(iii) If T and S are arbitrary 2-tangles and $[P/Q]$ denotes a specific rational tangle with a given continued fraction expansion, as above, then

$$N([P/Q]\sharp T + [-P/Q']\sharp S) = N(T + S).$$

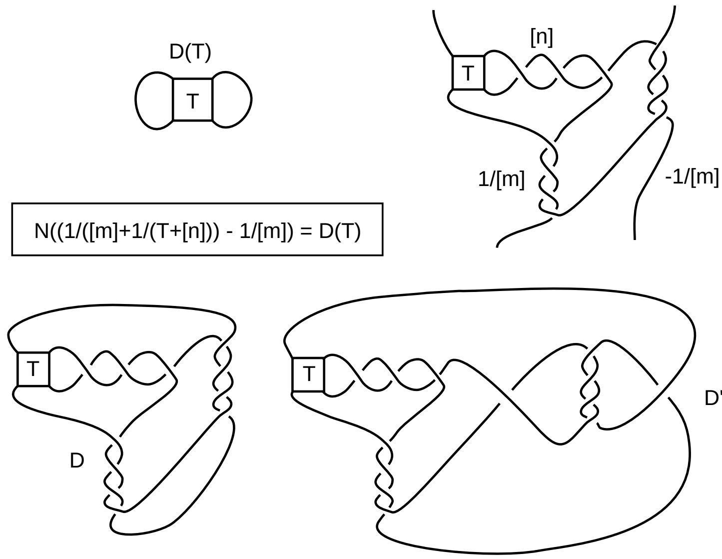


Figure 30: Sumners' Example

Example 1. View Figure 30. This figure illustrates the identity

$$N(\frac{1}{[m]+\frac{1}{[n]+T}} - \frac{1}{[m]}) = D(T)$$

where $D(T)$ denotes the denominator of T, and T is any tangle. The example is due to DeWitt Sumners and is fashioned to illustrate that an equation

$$N(A+B) = K$$

for fixed K and variable A and B may have infinitely many solutions, as we have proved above. Let's analyse the class of examples in Figure 30 in our formalism. Note that for any tangle T,

$N([T,0]) = N(T + 1/[0]) = N(T + [\infty]) = D(T)$.

Then we have

$N(\frac{1}{[m]+\frac{1}{[n]+T}} - \frac{1}{[m]}) = N([0,m,[n]+T] + [0,-m]) = N([[n]+T,m,0+0,-m])$

$= N([n+T,m,0,-m]) = N([[n]+T,0]) = D([n]+T) = D(T)$,

making this class a special case of the first part of the Proposition.

Example 2. View Figure 31. Here we have a knot $K = N(T+S)$ and have constructed three diagrams associated with K. The first diagram is $N([2,3,T]+[-2,-3,S])$, which by the results of this section is isotopic to K, and it is a hard diagram. The next diagram is made by applying a 180^0 turn and replacement to one of the subtangles of the previous diagram. It is hard in the plane but

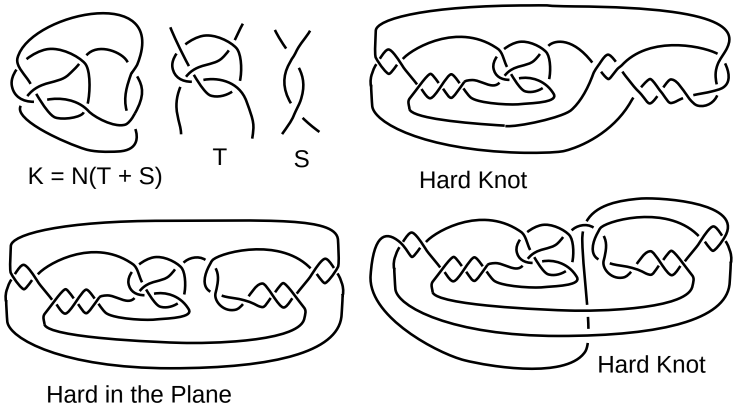


Figure 31: A Hard Knot

not on the 2-sphere. In the third diagram we remedy this situation by using the tucking construct, obtaining another hard diagram related to K.

Remark 9. Given a knot or link one can apply our methods of collapse in reverse, producing complex configurations that collapse back to that knot or link. This configuration can be allowed to undergo a crossing switch that locks it into a new knot or link that does not collapse. Thus we can create towers of knots and links such that each stage of the tower will collapse to the previous stage under a single crossing switch.

12 Stability in Processive DNA Recombination

In this section we use the techniques of this paper to study properties of the topology of processive DNA recombination. We follow the tangle model of DNA recombination [12, 13, 43] developed by C. Ernst and D.W. Sumners. In this model the DNA is divided into two regions corresponding to two tangles O and I and a recombination site that is associated with I. This division of the DNA is a model of how the enzyme that performs the recombination traps a part of the DNA, thereby effectively dividing it into the tangles O and I. The recombination site is represented by another tangle R. The entire arrangement is then a knot or link $K[R] = N(O + I + R)$. We then consider a single recombination in the form of starting with $R = [0]$, the zero tangle, and replacing R with a new tangle; for our purposes this will be the tangle $[1]$ or the tangle $[-1]$. Processive recombination consists in consecutively replacing again and again by $[1]$ or by $[-1]$ at the same site. Thus, in processive recombination we obtain the knots and links

$$K[n] = N(O + I + [n]).$$

The knot or link $K[0] = N(O + I)$ is called the *DNA substrate*, and the tangle $O + I$ is called the *substrate tangle*. It is of interest to obtain a uniform formula for knots and links $K[n]$ that result from the processive recombination.

In some cases the substrate tangle is quite simple and is represented as a single tangle $S = O + I$. For example, we illustrate processive recombination in Figure 32 with $R = [+1]$, $S = [-1/3] = [0, -3]$ and $I = [0]$ with $n = 0, 1, 2, 3, 4$. Note that by Proposition 1,

$$K[n] = N(S + [n]) = N([0, -3] + [n]) = N([-3, 0 + n]) = N([-3, n])$$

$$= N(-[3, -n]) = N(-[2, 1, n - 1]).$$

This formula gives the abstract form of all the knots and links that arise from this recombination process. This formula goes infinitely many steps further than Figure 32. Looking at Figure 32 we see that the sequence is: *Unknot, Hopf Link, Figure Eight Knot, Whitehead Link, Funny Twisty Diagram, and so on.* A single neat formula for the whole sequence is welcome.

We say that the formula

$$K[n] = N(-[2, 1, n - 1])$$

for $n > 1$ is *stabilized* in the sense that all the terms in the continued fraction have the same sign and the n is in *one single place* in the fraction. In general, a *stabilized fraction* will have the form

$$\pm[a_1, a_2, \ldots a_{k-1}, a_k + n, a_{k+1}, \ldots, a_n]$$

where all the terms a_i are positive for $i \neq k$ and a_k is non-negative. It is not quite obvious that the output of a recombination process will be stabilized in this sense. We shall prove below that this is always the case. For another example view Figure 33, where we show a process that stabilizes after the second term, with the transition from first to second term having its own idiosyncracy.

Let's see what the form of the processive recombination is for an arbitrary sequence of recombinations. We start with

$$O = [a_1, a_2, \ldots, a_r]$$

$$I = [b_1, b_2, \ldots, b_s].$$

Then

$$K[n] = N(O + (I + [n])) = N([a_1, a_2, \ldots, a_r] + [n + b_1, b_2, \ldots, b_s])$$

$$= N([a_r, a_{r-1}, \ldots, a_2, a_1 + n + b_1, b_2, \ldots, b_s]).$$

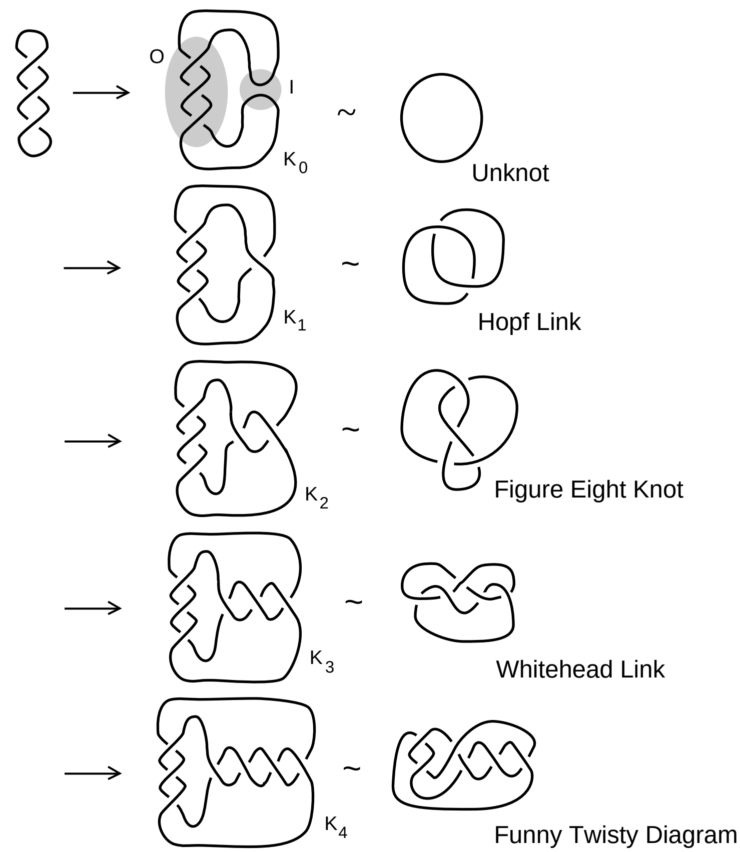


Figure 32: Processive Recombination with $S = [-1/3]$

Proposition 3. *The formula*

$$K[n] = N([a_r, a_{r-1}, \ldots, a_2, a_1 + n + b_1, b_2, \ldots, b_{s-1}, b_s])$$

can be simplified to yield a stable formula (in the sense of the discussion above) for the processive recombination when n is sufficiently large.

Proof. Apply Proposition 1. □

Here is an example. Suppose we take $O = [1, 1, 1, 1]$ and $I = [-1, -1, -1]$ so that the DNA substrate is an (Fibonacci) unknot. Then by the above calculation:

$$K[n] = N([1, 1, 1, 1 + n + (-1), -1, -1]) = N([1, 1, 1, n, -1, -1]).$$

Suppose that n is positive. Applying Proposition 1, we get:

$$K[n] = N([1, 1, 1, n, -1, -1]) = N([1, 1, 1, n - 1, 1, 0, 1]) = N([1, 1, 1, n - 1, 2]),$$

and this is a stabilized form for the processive recombination.

More generally, suppose that $O = [a_1, \ldots, a_n]$ where all of the a_i are positive. Let $I = [-a_1, \ldots, -a_{n-1}]$. Then $K[0] = N(O + I)$ is an unknotted substrate by our result about convergents (Proposition 2 and Theorem 7). Consider $K[n]$ for positive n. We have

$$K[n] = N([a_n, a_{n-1}, \ldots, a_2, a_1 + n - a_1, -a_2, \ldots, -a_{n-1}])$$

$$= N([a_n, a_{n-1}, \ldots, a_2, n, -a_2, \ldots, -a_{n-1}])$$

$$= N([a_n, a_{n-1}, \ldots, a_2, (n - 1), 1, (a_2 - 1), a_3, \ldots, a_{n-1}]).$$

If $a_2 - 1$ is not zero, the process terminates immediately. Otherwise there is one more step. In this way the knots and links resulting from the recombination process all have a uniform stabilized form. Further successive recombination just adds more twist in one entry in the continued fraction diagram whose closure is $K[n]$.

The reader may be interested in watching a visual demonstration of these properties of DNA recombination. For this we recommend the program *Ginterface* (TangleSolver) [44] of Mariel Vasquez. Her program can be downloaded from the internet as a Java applet, and it performs and displays DNA recombination. Figure 33 illustrates the form of display for this program. The reader should be warned that the program uses the reverse order from our convention when listing the terms in a continued fraction. Thus we say $[1, 1, 2]$ while the program uses $(2, 1, 1)$ for the same structure.

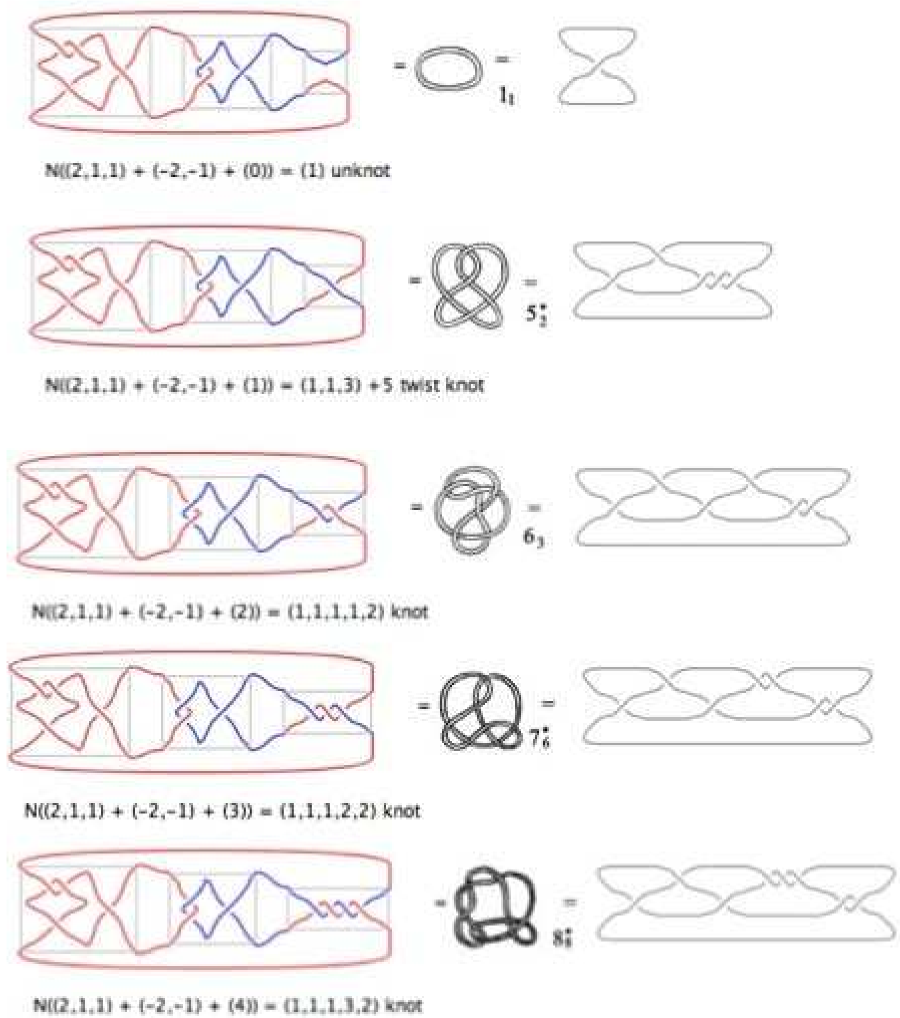


Figure 33: Processive Recombination with $S = O + I$ and $O = [1, 1, 1, 1]$, $I = [-1, -1, -1]$.

Remark 10. Along with this stabilization process, there is the general form of the recombination process itself that is related to our study of collapse. Our results apply to the process of successive recombination in DNA where a molecular configuration may undergo a crossing switch of DNA strands that locks it into a new topological configuration. Thus, in successive recombination towers of knots and links are created, each one a single crossing switch away from the next. The patterns of long sequences of such recombinations are beyond the reach of experiment, but can be understood mathematically. We hope that the results of this paper will apply to future experiments with DNA recombination.

13 Afterthoughts - Farey Series, Continued Fractions, Pick's Theorem and Ford Circles

The main theme of this paper has been our result (Theorem 6) that, given two fractions $\frac{p}{q}$ and $\frac{r}{s}$ such that $|ps - qr| = 1$, we can construct an unknot diagram $N([\frac{p}{q}] - [\frac{r}{s}])$ from the rational tangles $[\frac{p}{q}]$ and $[\frac{r}{s}]$. The arithmetic condition $|ps - qr| = 1$ has so many beautiful and surprising connections to other mathematics, that we must mention them in this last section of the paper! We shall touch on Farey series [14] and continued fractions, the Riemann Hypothesis, Pick's Theorem and Ford circles [6]. There is more, but we hope that this will give the reader a taste, and perhaps there will arise new connections with knots and unknots as well through this discussion.

To remind us to think of knots, we shall call fractions $\frac{p}{q}$ and $\frac{r}{s}$ such that $|ps - qr| = 1$ an *unknot pair*. Note that if $\frac{p}{q}, \frac{r}{s}$ is an unknot pair, then $\frac{q}{p}, \frac{s}{r}$ is also an unknot pair (see Remark 6 in Section 4). When $\frac{p}{q} < \frac{r}{s}$, then we have $ps - qr = -1$, the plus sign appearing when the fractions are in the other order. It is convenient to include the formal fractions $0/1$ and $1/0$ in these discussions just as we have done earlier in the paper with the tangles $[0]$ and $[\infty]$.

13.1 Farey Series and Continued Fractions

Given two fractions $\frac{a}{b}$ and $\frac{c}{d}$ such that $ad - bc = -1$, one can form the *mediant* fraction $\frac{e}{f} = \frac{a+c}{b+d}$. If $\frac{a}{b} < \frac{c}{d}$, then it is easy to see that $\frac{a}{b} < \frac{e}{f} < \frac{c}{d}$ and that $af - be = -1 = ed - fc$. This means that one can iterate the mediant construction, producing infinitely many unknot pairs from the given pair $(\frac{a}{b}, \frac{c}{d})$.

This iteration via the mediant fraction was discovered by John Farey [14] in 1816. It follows from Farey's construction that if one starts with the fractions $(\frac{0}{1}, \frac{1}{1})$ and iterates the mediant construction forever, then all rational numbers in the interval $[0, 1]$ will appear uniquely in their reduced forms.

In order to create any positive rational number greater than 1 we also consider the formal fraction $\frac{1}{0}$ and apply the mediant construction. Note that, then

$\frac{1}{1} = \frac{0+1}{1+0}$ (one is the mediant of zero and infinity) so $\frac{0}{1}$ and $\frac{1}{0}$ can be seen as the "primal ancestors" of the rational numbers.

Given a real number $x > 1$, one can consider the rational numbers which, when expressed in lowest terms, have denominators less than x. The *Farey series* corresponding to x is the (ordered) set of positive rational numbers less than or equal to 1 which, in reduced form, have denominators less than x. For example, the Farey series corresponding to 6 is (in order):

$$\{\frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}\}.$$

The reader will notice that each adjacent pair of fractions in the Farey series above is an unknot pair. Each Farey series can be obtained by iterating the mediant construction starting with $\{\frac{0}{1}, \frac{1}{1}\}$, and omitting at each step the fractions whose denominators exceed x, until we reach a set for which every mediant is no longer permitted. For example, here is the generation of the Farey series corresponding to 6 (with $\frac{0}{1}$ and $\frac{1}{1}$ retained at the left and the right).

$$\{\frac{0}{1}, \frac{1}{1}\}$$
$$\{\frac{0}{1}, \frac{1}{2}, \frac{1}{1}\}$$
$$\{\frac{0}{1}, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, \frac{1}{1}\}$$
$$\{\frac{0}{1}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{1}{1}\}$$
$$\{\frac{0}{1}, \frac{1}{5}, \frac{1}{4}, \frac{1}{3}, \frac{2}{5}, \frac{1}{2}, \frac{3}{5}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{1}{1}\}$$

Adjacent pairs are unknot pairs because each fraction is the mediant of its neighbors.

Remark 11. Recall that the *Riemann Hypothesis* says that all the non-trival zeroes of the Riemann zeta function

$$\zeta(s) = \Sigma_{n=1}^{\infty} \frac{1}{n^s}$$

lie on the half-line $s = \frac{1}{2} + it$ where i is the square root of -1 and t is a real number. This famous unsolved problem is equivalent to a statement about the Farey series: Let $A(x)$ denote the number of terms in the Farey series corresponding to x. Let $\delta_j(x)$ denote the amount by which the j-th term of the Farey series for x differs from $j/A(x)$. The following conjecture is equivalent to the Riemann Hypothesis.

Conjecture 2. (Franel and Landau). *For each $\epsilon > 0$ there exists a constant K, depending upon ϵ, such that $\Sigma_{j=1}^{A(x)} |\delta_j(x)| < Kx^{\frac{1}{2}+\epsilon}$ as $x \longrightarrow \infty$.*

The equivalence is due to Franel and Landau [16]. See also [10]. Thus the Riemann Hypothesis is equivalent to a statement about how certain collections of unknots jostle one another as pairs of rational numbers in the real line.

In Figure 34 (top) we give an illustration of the Farey generating process for positive rational numbers. Here the tiers of numbers are connected in a graph, whose top nodes are the "ancestors" $\frac{0}{1}$ and $\frac{1}{0}$. A given fraction is the mediant of its two ancestors. In the graph each fraction has two lines upward to its immediate ancestors. The graph contains a natural binary tree starting from $\frac{1}{1}$ (see bottom illustration of Figure 34). We have indicated the genesis of this binary tree as a subgraph of this graph that shows all ancestors of the Farey fractions.

A given positive continued fraction $[a_1, a_2, ..., a_n]$ can be found (say n is odd) by starting at $\frac{1}{1}$ in the tree and heading downward by a_1 edges to the right, a_2 edges to the left, a_3 edges to the right, $\ldots$, a_{n-1} edges to the left, and finally $a_n - 1$ edges to the right. If n is even, the sequence will start with a_1 edges to the right and end with $a_n - 1$ edges to the left.

For example $\frac{7}{5} = [1, 2, 2]$ is related to the sequence of instructions $RLLR$ where R denotes "right" and L denotes "left". The reader will note that these instructions take one from $\frac{1}{1}$ to $\frac{7}{5}$ in the tree of Figure 34. For another example, take the instruction $RRLL$ and note that it takes one from $\frac{1}{1}$ to $\frac{7}{3}$. We have $\frac{7}{3} = [2, 3]$ which indeed is related to $RRLL$. On the other hand, we also have $\frac{7}{3} = [2, 2, 1]$ and this also is related to $RRLL$. Thus the truncation to $a_n - 1$ left or right steps in the last part of the path corresponds to our earlier discussion of the ambiguity of the last term in the sequence for the continued fraction.

The upshot of this structure is that one can use a Farey tree (also called the *Stern-Brocot tree*) to enumerate the possible unknots and their corresponding continued fraction pairs. It is easy to see from this prescription that a mediant and one of its ancestors are convergents. Compare these comments with our Theorem 8. Note also that since the rationals are dense in the real numbers, this tree provides a way to obtain for each positive real number, a continued fraction that converges to it. Each real number is the limit of a path going down the binary tree. For example, the golden ratio $\phi = \frac{1+\sqrt{5}}{2} = [1, 1, 1, 1, \ldots]$ is the limit of the sequence $\{R, RL, RLR, RLRL, RLRLR, \ldots\}$. The inverse of the golden ratio, $\frac{1}{\phi} = \frac{\sqrt{5}-1}{2} = [0, 1, 1, 1, 1, \ldots]$, is the limit of the sequence $\{L, LR, LRL, LRLR, LRLRL, \ldots\}$. Each path down the infinite binary tree corresponds to the continued fraction expansion of a unique real number.

13.2 Pick's Theorem

Pick's Theorem [1], [2] states that the area of a polygon in the standard integral lattice in the Euclidean plane is given by the formula

$$A = I + B/2 - 1$$

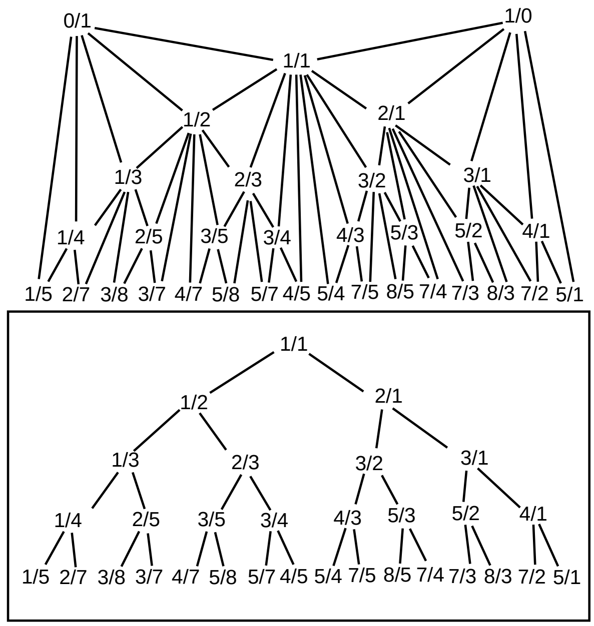


Figure 34: The graph of Farey fractions and binary tree of ancestors

where I denotes the number of lattice points in the interior of the polygon and B denotes the number of lattice points on the boundary. See Figure 35 for an example.

The simplest case of Pick's Theorem is when there are no interior lattice points and the figure is a triangle with just three lattice points on the boundary. Call such a figure a *small triangle.* The area of a small triangle is, by Pick's Theorem, equal to $1/2$. If one node of the small triangle is at the origin, then the other two nodes can be viewed as integral vectors (a, b) and (c, d) in the plane. One can show that if a small triangle is formed by the origin and the tips of the two vectors, then the determinant $ad - bc$ has absolute value equal to 1. Hence the two vectors form an alternative basis for the integer lattice in the plane. They satisfy the formula $|ad - bc| = 1$ and yield unknot fraction pairs $\frac{a}{b}, \frac{c}{d}$ or $\frac{b}{a}, \frac{d}{c}$. This means that the fundamental case of Pick's Theorm is directly related to the structure of unknot pairs and to consecutive fractions in the Farey construction.

We can illustrate this relationship with the integer lattice by plotting vectors in a finite lattice. We will say that two vectors are *adjacent* in the integer lattice if together with the origin they form a triangle with no interior lattice points. Starting with a finite lattice, one can plot all lattice vectors with non-zero slopes less than or equal to one whose coordinates are relatively prime. This gives a finite collection of vectors and it is easy to see that two such vectors are adjacent, in the sense given above, if and only if one can be rotated about

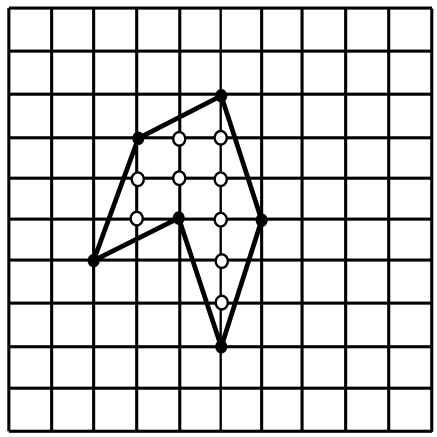


Figure 35: Area of a Lattice Polygon

the origin into the other without encountering another vector in the collection. Thus adjacency becomes rotational adjacency in such a plot. In Figure 36 we illustrate this pattern by plotting all such vectors in the 5×5 lattice. This reproduces the Farey series for denominators no larger than 5 if we associate the vector (a, b) with the fraction b/a. The reader will note that the collection of fractions corresponding to the points in the figure are

$$\{1/5,\ 1/4,\ 1/3,\ 2/5,\ 1/2,\ 3/5,\ 3/4,\ 1/1\}.$$

The topic of Pick's Theorem brings us back to topology in that the general case of the Theorem follows from the special case of the triangle with no interior lattice points by the use of Euler's formula for plane graphs:

$$V - E + F = 2$$

when the connected graph in the plane has V nodes, E edges and F faces (including the outer, unbounded, face). One triangulates the polygon using the lattice points. All the faces are small triangles except for the outer face. From this it follows that $3(F-1) + B = 2E$, and we have that $V = I + B$. These two equations and the Euler formula give $(F-1)/2 = I + B/2 - 1$. The area of the polygon is $(F-1)/2$ since it is composed of $F-1$ triangles, each of area one-half. This gives Pick's Theorem. It is curious to think of the triangulation of the polygon giving us $F-1$ unknots to examine.

13.3 Ford Circles

Finally, there is an amazing geometric interpretation of pairs of fractions a/b and c/d such that $|ad - bc| = 1$. Associate to each reduced fraction a/b on the

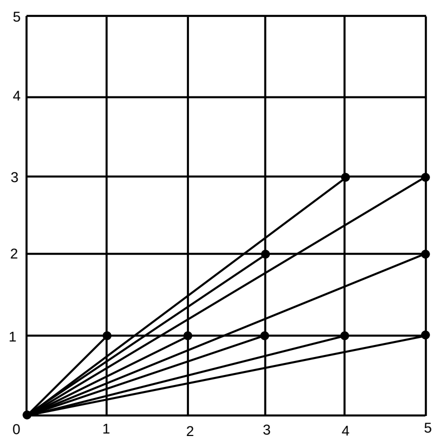


Figure 36: Lattice Vectors and Farey Fractions

real line a circle tangent to the real line of diameter $1/b^2$. This is the *Ford circle* associated with the fraction (see, for example, [15],[6]). One can easily show that *two fractions form an unknot pair if and only if their Ford circles are tangent to one another*. This means that the mediant of an unknot pair produces a new circle that is tangent to both of its ancestral Ford circles. See Figure 37 for an illustration of this geometry.

Ford circles can be regarded as curves in the complex plane. The set of Ford circles is invariant under the action of the modular group of transformations of the complex plane. By interpreting the upper half of the complex plane as a model of the hyperbolic plane (the Poincaré half-plane model) the set of Ford circles can be interpreted as a tiling of the hyperbolic plane. Thus, any two Ford circles are congruent in hyperbolic geometry. If C and C' are two tangent Ford circles, then the half-circle joining their respective fractions (on the real line) that is perpendicular to the real line is a hyperbolic line that also passes through the tangent point of the two circles. This half-circle is indicated with a dashed curve in Figure 37.

We have seen many aspects of the topology, number theory and geometry related to unknot pairs in the course of this paper. We hope that this last section stimulates the curiosity of the reader to make new explorations in this domain.

References

[1] M. Aigner, G.M. Ziegler, "Proofs from The Book", Springer-Verlag (1991).

[2] M. Bruckheimer, A. Arcavi, Farey series and Pick's area theorem, *The Mathematical Intelligencer*, **17**, No. 2 (1995), 64-67.

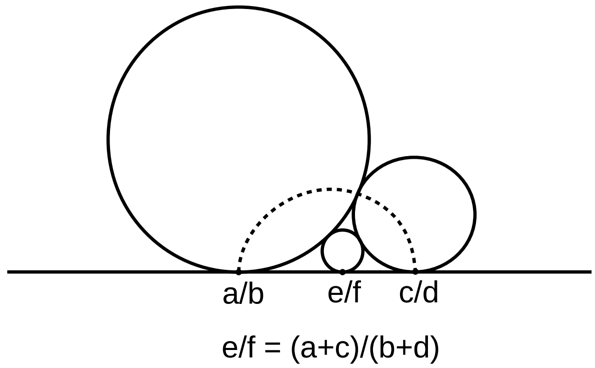


Figure 37: Ford Circles

[3] G. BURDE, Verschlingungsinvarianten von Knoten und Verkettungen mit zwei Brücken, *Math. Zeitschrift*, **145** (1975), 235-242.

[4] G. BURDE, H. ZIESCHANG, "Knots", de Gruyter Studies in Mathematics **5** (1985).

[5] J.H. CONWAY, An enumeration of knots and links and some of their algebraic properties, *Proceedings of the conference on Computational problems in Abstract Algebra held at Oxford in 1967*, J. Leech ed., (First edition 1970), Pergamon Press, 329-358.

[6] J.H. CONWAY, R.K. GUY, "The Book of Numbers", Springer Verlag (1996).

[7] I. DARCY, Solving unoriented tangle equations involving 4-plats, *J. Knot Theory Ramifications*, **14** (2005), no. 8, 993-1005.

[8] I. DARCY, Solving oriented tangle equations involving 4-plats, *J. Knot Theory Ramifications*, **14** (2005), no. 8, 1007-1027.

[9] I.A. DYNNIKOV, Arc presentations of links: Monotonic simplification, *Fund. Math.* **190** (2006), 29-76.

[10] H.M. EDWARDS, "Riemann's Zeta Function", Academic Press (1974), Dover Publications (2001).

[11] D. EPSTEIN, C. GUNN, "Supplement to Not Knot", (booklet accompanying the video "Not Knot"),<http://www.geom.uiuc.edu/video/NotKnot/>, A. K. Peters, Ltd., Natick, MA 01760-4626, USA.

[12] C. ERNST, D.W. SUMNERS, A calculus for rational tangles: Applications to DNA Recombination, *Math. Proc. Camb. Phil. Soc.*, **108** (1990), 489-515.

[13] C. ERNST, D.W. SUMNERS, Solving tangle equations arising in a DNA recombination model. *Math. Proc. Cambridge Philos. Soc.*, **126**, No. 1 (1999), 23-36.

[14] J. FAREY, On a curious property of vulgar fractions, *Philosophical Magazine*, **47**, (1816), 385-386.

[15] L.R. FORD, Fractions, *American Mathematical Monthly*, **45**, No. 9 (Nov., 1938), 586-601.

[16] J. Franel, E. Landau, Les suites de Farey et le problème des nombres premiers, *Gottinger Nachr.*, (1924), 198-206.

[17] J.S. Frame, Continued fractions and matrices, Classroom notes, C.B. Allendoerfer ed., *The Amer. Math. Monthly*, **56** (1949), 98-103.

[18] L. Goeritz, Bemerkungen zur Knotentheorie, *Abh. Math. Sem. Univ. Hamburg*, **10** (1934), 201-210.

[19] J.R. Goldman, L.H. Kauffman, Rational Tangles, *Advances in Applied Math.*, **18** (1997), 300-332.

[20] J.H. Hass, C. Lagarias, The number of Reidemeister moves needed for unknotting, *J. Amer. Math. Soc.*, **14**, No. 2 (2001), pp. 399-428.

[21] G. Hemion, On the classification of the homeomorphisms of 2-manifolds and the classification of three-manifolds, *Acta Math.*, **142**, no. 1-2, (1979), 123-155.

[22] A. Henrich, L. H. Kauffman, On upper bounds for an unknotting sequence of Reidemeister moves. arXiv:1006.4176 (to appear).

[23] S. Jablan (private conversation), $< http://math.ict.edu.yu/ >$.

[24] V.F.R. Jones, A polynomial invariant for links via von Neumann algebras, *Bull. Amer. Math. Soc.*, **129** (1985), 103-112.

[25] L.H. Kauffman, State Models and the Jones Polynomial, *Topology*, 26 (1987), pp. 395-407.

[26] L.H. Kauffman, S. Lambropoulou, On the classification of rational tangles, *Advances in Applied Math*, **33**, No. 2 (2004), 199237.

[27] L.H. Kauffman, S. Lambropoulou, On the classification of rational knots. *L'Enseignement Mathématiques*, **49**, (2003), 357-410.

[28] L. H. Kauffman, S. Lambropoulou, Unknots and molecular biology, *Milan. J. Math.*, **74** (2006), 227-263.

[29] A. Kawauchi, "A Survey of Knot Theory", Birkhäuser Verlag (1996).

[30] A.Ya. Khinchin, "Continued Fractions", Dover (1997) (republication of the 1964 edition of Chicago Univ. Press).

[31] T. Kanenobu, H. Murakami, Two-bridge knots with unknotting number one, *Proc. Amer. Math. Soc.* ,**98**, No. 3, Nov. 1986, 499-502.

[32] P. Kohn, Two-bridge links with unlinking number one, *Proc. Amer. Math. Soc.*, **113**, No. 4, Dec. 1991, 1135-1147.

[33] K. Kolden, Continued fractions and linear substitutions, *Archiv for Math. og Naturvidenskab*, **6** (1949), 141-196.

[34] K. Millett (private conversation at Institutes Hautes Etudes Scientifiques, Burres Sur Yvette, France, circa 1988).

[35] H. Morton, An irreducible 4-string braid with unknotted closure, *Proc. Camb. Phil. Soc.*, **93**, (1983), no. 2, pp. 259-261.

[36] J.M. Montesinos, Revetêments ramifiés des noeuds, Espaces fibres de Seifert et scindements de Heegaard, *Publicaciones del Seminario Mathematico Garcia de Galdeano, Serie II, Seccion 3* (1984).

[37] C.D. Olds, "Continued Fractions", New Mathematical Library, Math. Assoc. of Amerika, **9** (1963).

[38] K. Reidemeister, “Knotentheorie” (Reprint), Chelsea, New York (1948).

[39] K. Reidemeister, Knoten und Verkettungen, *Math. Zeitschrift*, **29** (1929), 713-729.

[40] R. Scharein, KnotPlot (program available from the web), <http://wren.pims.math.ca/knotplot>.

[41] H. Schubert, Knoten mit zwei Brücken, *Math. Zeitschrift*, **65** (1956), 133-170.

[42] L. Siebenmann, Lecture Notes on Rational Tangles, Orsay (1972) (unpublished).

[43] D.W. Sumners, Untangling DNA, *Math.Intelligencer*, **12** (1990), 71-80.

[44] M. Vasquez, TangleSolver (program available from the web),<http://math.berkeley.edu/ mariel/>.

[45] H.S. Wall, “Analytic Theory of Continued Fractions”, D. Van Nostrand Company, Inc. (1948).

L.H. Kauffman: Department of Mathematics, Statistics and Computer Science, University of Illinois at Chicago, 851 South Morgan St., Chicago IL 60607-7045, U.S.A.

S. Lambropoulou: Department of Mathematics, National Technical University of Athens, Zografou Campus, GR-157 80 Athens, Greece.

E-mails: `kauffman@math.uic.edu` `sofia@math.ntua.gr`
`http://www.math.uic.edu/~kauffman/` `http://www.math.ntua.gr/~sofia`