KNOT INVARIANTS AND THE REPRESENTATION THEORY OF QUANTUM GROUPS

BRYAN WANG PENG JUN

Contents

1. Introduction

1.1. Knot theory is, roughly speaking, the study of 'knotted' curves in space, up to continuous deformation. One of the fundamental questions in knot theory is therefore to distinguish between knots that cannot be obtained from each other via continuous deformation. The key tool used to answer such questions is that of a *knot invariant*.

Perhaps one of the most influential and consequential knot invariants, which spurred much of the development of modern knot theory, is the Jones polynomial. It was discovered by Reshitikhin and Turaev ([RT]) that the Jones polynomial has connections to and naturally arises from the representation theory of the *quantum group* $\mathbf{U}_q(\mathfrak{sl}_2)$, a deformation of the classical enveloping algebra $\mathbf{U}(\mathfrak{sl}_2)$. Later, Khovanov constructed a *categorification* of the Jones polynomial ([Kh]), thereby obtaining a strictly stronger knot invariant which is now known as Khovanov homology. It was then shown by Stroppel ([St]) that Khovanov homology itself arises from the *categorified* representation theory of $\mathbf{U}_q(\mathfrak{sl}_2)$. The situation can be summarised as follows:

$$\begin{array}{ccc} \text{Jones polynomial} & \longleftarrow & \text{Representation theory of } \mathbf{U}_q(\mathfrak{sl}_2) \\ \Big\downarrow{\scriptstyle \text{Categorification}} & & \Big\downarrow{\scriptstyle \text{Categorification}} \\ \text{Khovanov homology} & \longleftarrow & \text{Categorified representation theory of } \mathbf{U}_q(\mathfrak{sl}_2) \end{array}$$

1.2. In this report, we understand and make precise the key approaches leading to the Reshitikhin-Turaev invariants and Khovanov homology, thereby relating the representation theory of quantum groups to the theory of knot invariants. The report is organised as follows. We recall the essentials of knot theory and knot invariants in Section 2. We consider the Khovanov homology in Section 3. We study the quantum group $\mathbf{U}_q(\mathfrak{sl}_2)$ and its representations in Section 4. We construct the Reshetikhin-Turaev invariants arising from quantum groups in Section 5.

Acknowledgement: This project was conducted under the Undergraduate Research Opportunities Programme in Mathematics at the National University of Singapore, and supervised by Prof Bao Huanchen. I would like to thank Prof Bao for his guidance, helpful suggestions and strong support throughout the entire project.

2. Preliminaries on knot invariants

2.1. Links and tangles.

2.1.1. The central objects of knot theory that are of importance in our study of knot invariants are links and tangles. Informally, a knot is the embedding of a single closed loop in 3D space; a link is the embedding of a disjoint union of n closed loops in 3D space; and a tangle is a 'part' of a link in 3D space. Each closed loop may be given an orientation, from which we obtain *oriented* links and tangles.

The key point is that we consider (oriented) knots, links and tangles up to *isotopy*, that is, we consider two links or tangles to be the same if there is a continuous deformation in 3D space from one link/tangle to the other (so that the link/tangle does not pass through itself).

2.1.2. Since we are studying links/tangles modulo isotopy, and in particular are only interested in the combinatorial and algebraic structure arising from links/tangles modulo isotopy, it is conceptually more convenient (and indeed equivalent) to define them as *piecewise linear* embeddings, and to define their equivalence in more discrete terms. Accordingly, we have

Definition 2.1. (Links) A *link* is a disjoint, piecewise-linear, embedding of n copies of S^1 into $\mathbb{R}^3$. Each copy of S^1 is called a *component* of the link. Equivalently, it is a disjoint union of n non-self-intersecting closed polygonal arcs in $\mathbb{R}^3$, with each polygonal arc a component of the link. A link is *oriented* if each component is given a fixed orientation.

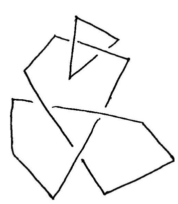


Figure 1. Example of link.

Definition 2.2. (Δ-move) Given a link L, two consecutive vertices A, B of a component (which is a polygonal arc) of L, and a point C in $\mathbb{R}^3$ such that the triangle ABC intersects L only along the segment AB, a Δ*-move* on L replaces L with a new link L' defined by removing segment AB and replacing it with segments AC, CB.

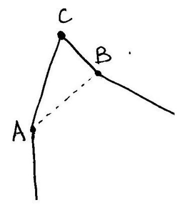


FIGURE 2. Example of Δ-move.

Definition 2.3. (Equivalence) Two links L, L' are (combinatorially) *equivalent* if there is a sequence of Δ-moves bringing L to L' in $\mathbb{R}^3$.

2.1.3. The informal definition of a tangle as a 'part' of a link gives rise to two similar but distinct definitions of a tangle that are often used in knot theory. To avoid confusion, we have chosen to call tangles of Definition 2.4 as *boxed* tangles, and those of Definition 2.5 as *circled* tangles. This is not standard terminology, and in practice, it will usually be clear from the context which type of tangle we are referring to, so we will not use these terms often.

Definition 2.4. (Boxed tangles) A (boxed) *tangle* is a disjoint, piecewise-linear, embedding of some copies of S^1 and the unit interval $[0,1]$ into $\mathbb{R}^2 \times [0,1]$, so that the boundary of the copies of $[0,1]$ lies within $\mathbb{R}^2 \times \{0,1\}$. Each point of this boundary is then called an *endpoint* of the tangle. Each copy of S^1 or $[0,1]$ is called a *component* of the link. Equivalently, it is a disjoint union of n non-self-intersecting, not necessarily closed polygonal arcs in $\mathbb{R}^2 \times [0,1]$, with each polygonal arc a component of the tangle.

Definition 2.5. (Circled tangles) A (circled) *tangle* is a disjoint, piecewise-linear, embedding of some copies of S^1 and the unit interval $[0,1]$ into the unit ball, so that the boundary of the copies of $[0,1]$ lies within the unit sphere. Each point of this boundary is then called an *endpoint* of the tangle. Each copy of S^1 or $[0,1]$ is called a *component* of the link. Equivalently, it is a disjoint union of n non-self-intersecting, not necessarily closed polygonal arcs in $\mathbb{R}^2 \times [0,1]$, with each polygonal arc a component of the tangle.

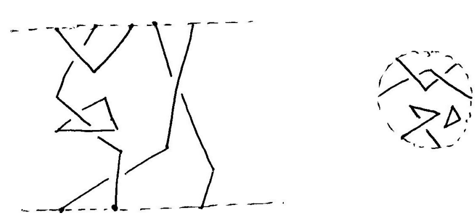


FIGURE 3. Example of boxed (left) and circled (right) tangles.

A tangle is oriented if each component is given a fixed orientation. The Δ-moves and equivalence of tangles are defined similarly as that for links (Definitions 2.2, 2.3).

2.1.4. Observe that a tangle with no endpoints may be viewed simply as a link and vice versa; we will often do so later. Furthermore, we will usually consider *oriented* links and tangles.

2.2. **Link and tangle diagrams; Reidemeister's theorem.**

2.2.1. While we study links and tangles up to equivalence in $\mathbb{R}^3$, in practice, it is usually not feasible to work with the equivalence of these objects in 3D space. For a start, we would like to represent links and tangles using 2D diagrams. We would then like to consider the equivalence of these 2D diagrams as representing equivalent links or tangles, and our study of 3D links and tangles modulo equivalence is reduced to the study of these 2D diagrams modulo equivalence of diagrams. This motivates the following:

Definition 2.6. (Regular projection) A *regular projection* of a link or tangle L is a projection of L from $\mathbb{R}^3$ onto $\mathbb{R}^2$ such that: each point of $\mathbb{R}^2$ which is mapped to by more than one point of L, is mapped to by only two points of L; there are only finitely such points; and no such point is mapped to by a vertex of a polygonal arc (component) of L.

Informally, a regular projection is simply a representation of the link/tangle as a 2D diagram so that the link/tangle can be unambiguously recovered from the diagram (up to equivalence, of course). It is easy to convince oneself that a regular projection exists for any link/tangle; the idea is that we may make suitable (small) Δ-moves so that no more than two points map to the same point in $\mathbb{R}^2$.

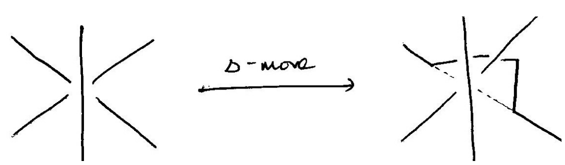


FIGURE 4. Making a Δ-move to obtain a regular projection.

2.2.2. Given any link or tangle L, a regular projection allows us to obtain a 2D link or tangle diagram if, in addition, the points which are mapped to by two points of L are represented suitably as over- or under-crossings. Note a slight 'abuse of notation' in that although we define links and tangles as polygonal arcs, the diagrams we draw often involve smooth curves. This is merely a matter of convenience and ease of conceptual understanding.

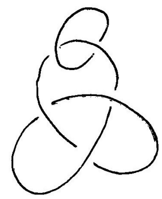


FIGURE 5. The same link as in Figure 1, drawn as smooth curves. Note that this also shows that we have all along been implicitly using 2D diagrams to represent our links and tangles!

2.2.3. As mentioned, we would like to reduce our study of 3D links and tangles modulo equivalence, to the study of these 2D link and tangle diagrams modulo equivalence of diagrams. The following Theorem 2.7 does precisely this and is therefore of central importance in knot theory and in our subsequent discussion.

Theorem 2.7. *(Reidemeister's Theorem) Two links/tangles are equivalent (recall (Definition 2.3) that this is iff they are related by a sequence of* Δ*-moves), if and only if their corresponding diagrams (under regular projection) are related by a sequence of so-called Reidemeister moves (or R-moves), which are defined as in Figure 6. If the links/tangles are oriented, then the set of moves is the same as in Figure 6, taken over all possible orientations of each component involved in the move.*

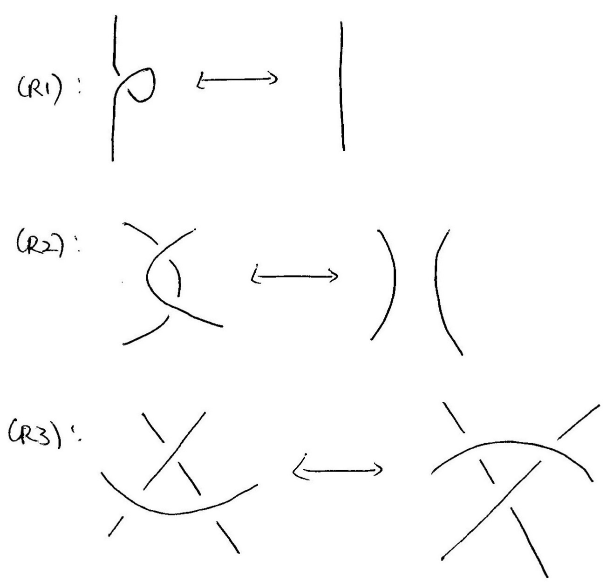


FIGURE 6. Reidemeister moves.

Proof. (Sketch). The backward direction is intuitively straightforward, so we focus on the forward direction. It suffices to show that each Δ-move can be realised by a sequence of Reidemeister moves.

The idea is to consider the triangle ABC in any such Δ-move (Definition 2.2) and the crossings contained within ABC (under the regular projection). ABC admits a triangulation so that each piece of the triangulation contains at most one such crossing. Since the original Δ-move can be realised by a sequence of Δ-moves corresponding to each piece of the triangulation, it remains to verify that each Δ-move where there is at most one crossing within the triangle ABC can be obtained by a sequence of Reidemeister moves, which is a straightforward case verification. □

Corollary 2.8. *The central consequence of Reidemeister's theorem is the following:*

$$\frac{\textit{links or tangles}}{\textit{equivalence}} \cong \frac{\textit{link or tangle diagrams}}{\textit{Reidemeister moves}}.$$

In other words, our study of (3D) links and tangles is reduced to the study of (2D) link and tangle diagrams, modulo the R-moves.

Remark 2.9. We remark here that of the three Reidemeister moves, (R2) and (R3) appear to be more 'tautological' as they involve only the 'planar' movement of one component across another or across a crossing, whereas (R1) appears to contain some 'non-trivial' content in that it removes a 'twist' from the diagram. This is not merely a cosmetic difference and will play a role in our subsequent discussion, giving rise to the notion of *framed* links and tangles (also known as *ribbons*), which are links and tangles modulo only the relations (R2) and (R3).

2.3. **Link and tangle invariants; Jones polynomial.**

2.3.1. Our definition of link and tangle equivalence and the consequential Theorem 2.7 provides us a means to tell if two links or tangles are indeed equivalent. However, the problem of greater significance in knot theory is to distinguish links and tangles; that is, to tell if two links or tangles are *not* equivalent. The key tool used to answer questions in this regard are *knot invariants*, or to be more precise, link and tangle invariants.

A *link or tangle invariant* is, roughly speaking, an assignment

$$f : \frac{\text{links or tangles}}{\text{equivalence}} \to S$$

for some set S, or equivalently, an assignment

$$f : \frac{\text{link or tangle diagrams}}{\text{Reidemeister moves}} \to S.$$

While it is easy enough to define f on link or tangle diagrams (e.g. number of crossings), the key point is that of well-definedness: f must remain invariant under equivalence, or under Reidemeister moves. This is precisely where the term 'invariant' comes from, even though f is expected to take a wide range of values for different links/tangles.

2.3.2. Perhaps the knot invariant of greatest significance in modern knot theory is the *Jones polynomial*. Here, we define the Jones polynomial via a *state-sum formulation* of the Kauffman bracket. Defining the Jones polynomial in this way also allows us to illustrate the process of *deframing*, which we will employ again in the subsequent sections.

Definition 2.10. (Smoothing) Given a crossing in an (unoriented) link/tangle diagram D, a *smoothing* is a local change of the diagram by replacing the crossing as follows:

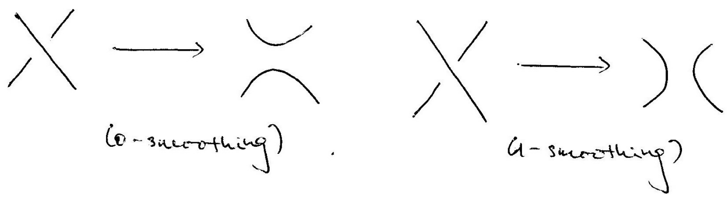


FIGURE 7. 0-smoothing (left) and 1-smoothing (right).

A change of the first kind is called a *0-smoothing*, while that of the second kind is called a *1-smoothing*. (The choice of which is 0 and which is 1 is merely a matter of convention.)

A *complete* smoothing of a link/tangle diagram D with n crossings is a diagram in which all n crossings have been smoothed. The resulting diagram has no crossings; in particular if D is a link diagram, the resulting diagram is a collection of closed loops in the plane.

Definition 2.11. (States) Given an (unoriented) link/tangle diagram D with n crossings labelled 1 to n, a *state* of D is a complete smoothing of D corresponding to one of the 2^n elements of $\{0,1\}^n$, which indicates how each of the n crossings is to be smoothed (0-smoothed or 1-smoothed). Given an element $s \in \{0,1\}^n$, denote by $s(D)$ the corresponding completely smoothed diagram.

Definition 2.12. (Kauffman bracket, state-sum formulation) Given an (unoriented) link diagram D with n crossings and its 2^n states corresponding to the elements of $\{0,1\}^n$, the *Kauffman bracket* $\langle \cdot \rangle$ is defined by:

$$\langle D \rangle = \sum_{s \in \{0,1\}^n} q^{i(s)-n/2}(-q-q^{-1})^{c(s(D))}$$

where q is an indeterminate, $i(s)$ denotes the number of 1s in s, and $c(s(D))$ denotes the number of closed loops in the completely smoothed diagram $s(D)$.

Example 2.13. Figure 8 illustrates an example in the case of the link consisting of two interlocking loops. The diagram has $n = 2$ crossings and so the sum is taken over $2^2 = 4$ states.

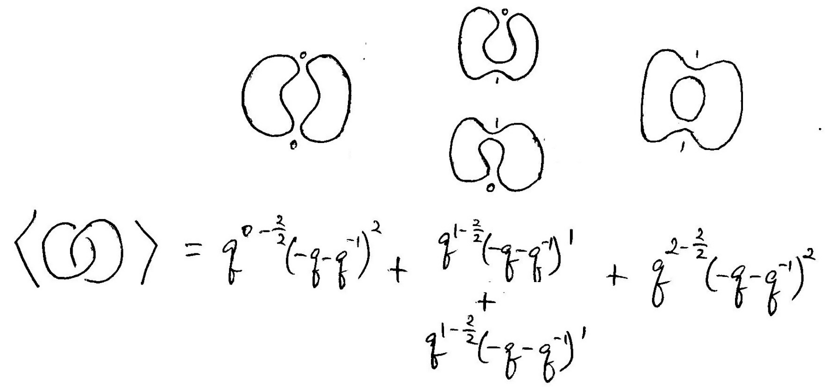


FIGURE 8. Example of Kauffman bracket.

Remark 2.14. If D has $n = 0$ crossings, we adopt the convention of taking $s = \emptyset$ as the empty smoothing and set

$$\langle D \rangle = (-q - q^{-1})^{c(D)}$$

Remark 2.15. It is easy to verify (by considering the states s with equal $i(s)$) that although the states $s(D)$ of a link diagram D depend on a choice of ordering of the n crossings of D, the Kauffman bracket $\langle D \rangle$ does not.

The Kauffman bracket is *a priori* well-defined for all (unoriented) link diagrams D. However, the point is that we need to study how the Kauffman bracket behaves under the R-moves.

Lemma 2.16. *Given an (unoriented) link diagram D, let D' be the link diagram obtained by the disjoint addition of a single closed loop to D. Then*

$$\langle D' \rangle = (-q - q^{-1})\langle D \rangle$$

Proof. Straightforward verification. □

Lemma 2.17. *(Skein relation for Kauffman bracket) Given an (unoriented) link diagram D and one of its crossings C, let D_0, D_1 be the link diagrams obtained by the 0-smoothing and 1-smoothing respectively of C. The Kauffman bracket satisfies the following skein relation:*

$$\langle D \rangle = q^{-1/2}\langle D_0 \rangle + q^{1/2}\langle D_1 \rangle$$

Proof. Without loss of generality (Remark 2.15), we may suppose C is last in the order of the n crossings of D, and the remaining $n - 1$ crossings, which will be precisely the crossings of D_0 and D_1, are ordered the same way in D, D_0, D_1. For a given $s \in \{0, 1\}^n$, let s_n denote its last element (0 or 1), and $s' \in \{0, 1\}^{n-1}$ denote the resultant when the last element s_n is removed. We have

$$\begin{aligned}\langle D\rangle &= \sum_{s;s_n=0} q^{i(s)-n/2}(-q-q^{-1})^{c(s(D))} \\ &\quad + \sum_{s;s_n=1} q^{i(s)-n/2}(-q-q^{-1})^{c(s(D))} \\ &= \sum_{s;s_n=0} q^{-1/2}q^{i(s')-(n-1)/2}(-q-q^{-1})^{c(s'(D_0))} \\ &\quad + \sum_{s;s_n=1} q^{1/2}q^{i(s')-(n-1)/2}(-q-q^{-1})^{c(s'(D_1))} \\ &= q^{-1/2}\langle D_0\rangle + q^{1/2}\langle D_1\rangle.\end{aligned}$$

□

Remark 2.18. Lemma 2.17 allows us to essentially translate our *global* Definition 2.12 of the Kauffman bracket into a *local* relation. This is essential as the R-moves are really local relations on link diagrams. While it is possible (and perhaps more standard) to take Lemma 2.17 as the definition of the Kauffman bracket, some care is needed to show well-definedness.

Proposition 2.19. *(Kauffman bracket under R-moves) The Kauffman bracket is invariant under (R2) and (R3). As for (R1), observe that any 'twist' in an (unoriented) link diagram, as in (R1), may be uniquely positioned (under rotations of the diagram only) so that the twist points toward the right. Then we have to consider whether the crossing in the twist is an over-crossing or an under-crossing (going from the bottom up). We have:*

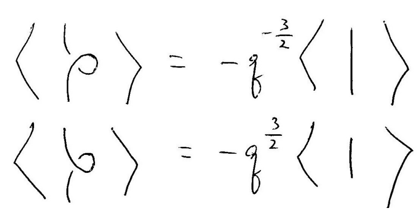


FIGURE 9. Kauffman bracket under (R1).

Proof. We have:

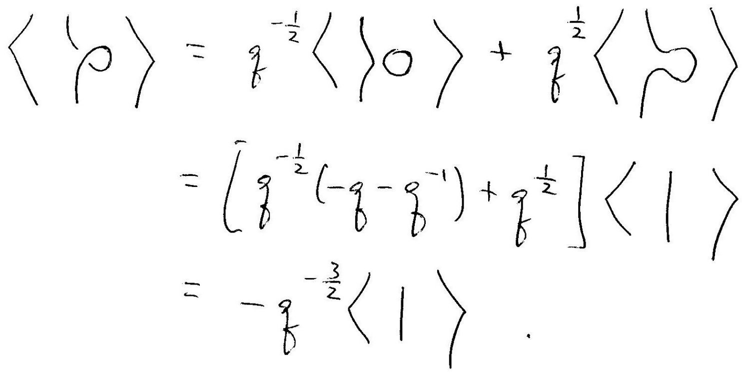


FIGURE 10. Kauffman bracket under (R1) (over-crossing).

The other case for (R1) (under-crossing) is similar, with only a difference in signs. As for (R2) and (R3), we have:

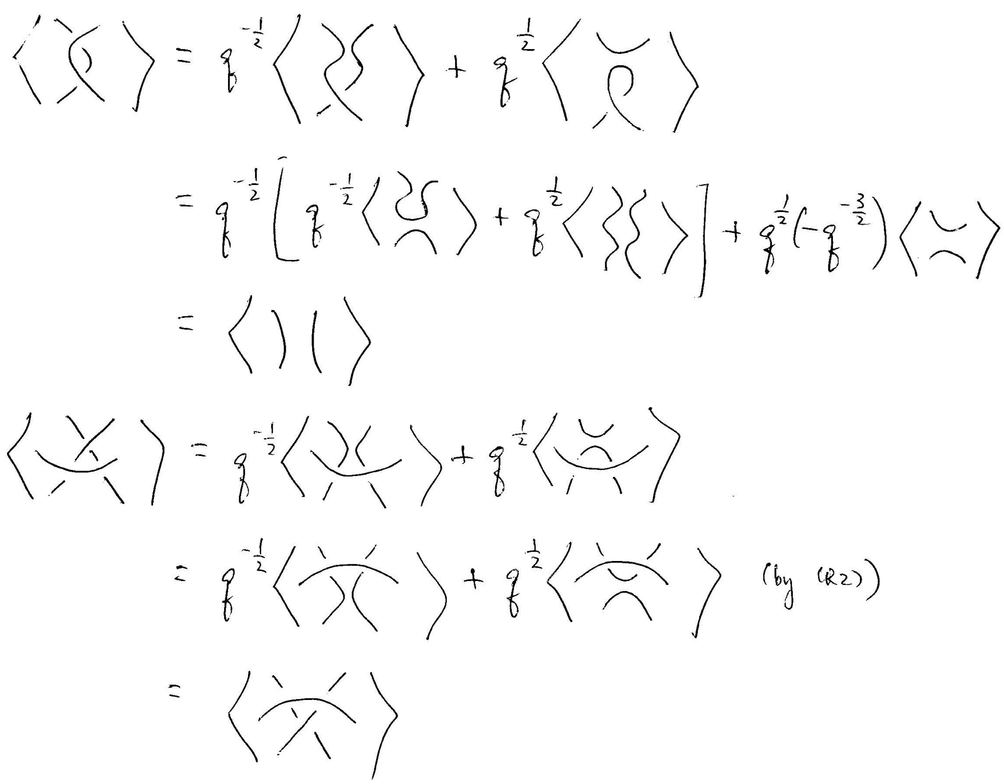


FIGURE 11. Kauffman bracket under (R2), (R3).

□

2.3.3. Proposition 2.19 tells us that the Kauffman bracket is an *almost-invariant* of links: it is invariant under (R2) and (R3), but under (R1) it is changed by scalar multiplication. (It is, however, therefore an invariant of framed links, cf. Remark 2.9.)

Since that the Kauffman bracket is only changed by scalar multiplication under (R1), one may consider adding a multiplicative factor to the Kauffman bracket which precisely 'cancels' the scalar multiplication of (R1) in order to obtain a true invariant of links.

This is not possible for unoriented links, because there is essentially only 'one' type of crossing in an unoriented diagram, so there is no way to distinguish between the two versions of (R1) which give rise to different scalars.

However, the situation is different if we consider *oriented* links and obtain an invariant of *oriented* links. This is known in the knot theory literature as *deframing*.

Definition 2.20. (Right-handed and left-handed crossings) Given an oriented link or tangle diagram D, and an (oriented) crossing C, the component l_u which passes *under* in this crossing C may be uniquely positioned, under rotations of D only, so that it points upward in the diagram. Then the component l_o which passes *over* in C may point either toward the right, whence we call C a *right-handed (RH) crossing*, or toward the left, whence we call C a *left-handed (LH) crossing*. In a diagram D, let $n_+(D)$ and $n_-(D)$ denote the number of RH and LH crossings respectively.

Now, keeping the terminology of Proposition 2.19, we observe that if the crossing in the twist is an over-crossing, then removing this twist under (R1) always decreases the number of RH crossings by 1, and keeps the number of LH crossings the same, regardless of whether the component is oriented upward or downward.

Similarly, if the crossing is an under-crossing, then removing this twist under (R1) always decreases the number of LH crossings by 1, and keeps the number of RH crossings the same, regardless of whether the component is oriented upward or downward.

(R2) decreases the number of crossings of each type by 1 each, while (R3) does not change the number of crossings of either type. We therefore obtain the following:

Theorem 2.21. *(Jones polynomial) Given an oriented link diagram D with underlying unoriented link diagram D' (obtained by 'forgetting' the orientation of D), the quantity*

$$J(D) := (-q^{3/2})^{n_+(D)-n_-(D)}\langle D'\rangle$$

is an invariant of oriented links.

Proof. Follows from Proposition 2.19 and the preceding discussions. □

We conclude with a result (and a preliminary lemma which is itself interesting in its own right) that will be useful later on when working with the Jones polynomial.

Lemma 2.22. *Given any oriented link diagram D with n crossings, it is always possible to change D into a diagram of an unlink (several disjoint closed loops) by changing some of its over-crossings to under-crossings and vice versa (while keeping the rest of the diagram unchanged).*

Proof. First identify the components of D; it is certainly possible to change the crossings so that the link diagram corresponds to a link in 3D space for which no two components ever possess the same z-coordinate (z-axis perpendicular to the 2D diagram).

Now for each component, we choose a point on it (which is not a crossing) and start 'walking' along it in the direction of its orientation. Every time we encounter a crossing (with itself) *which we have not encountered before*, change it (if need be) so that the strand we are currently on lies *under* the other strand. Otherwise, if we encounter a crossing (with itself) *which we have encountered before*, this means that the strand we are currently on must lie *over* the other strand (by the preceding sentence), and we leave the crossing as is. It is clear then that each component is now an unlink. ☐

Proposition 2.23. *('Local' formulation of Jones polynomial) The Jones polynomial is uniquely defined by the following relations:*

- *For any diagram D of an unlink with n components (n disjoint closed loops), we have $J(D) = (-q - q^{-1})^n$;*
- *Given any oriented link diagram D and one of its crossing C, we have the following local (at C) skein relation:*

Figure 12. Skein relation for Jones polynomial.

Proof. There are two parts to this statement. First we need to show that the Jones polynomial we have already defined satisfies the two stated properties. The first is immediate; for the second, we have:

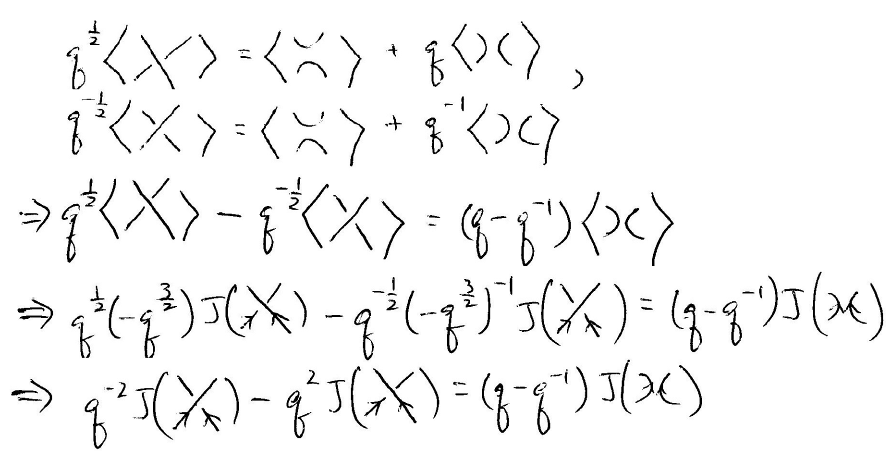


Figure 13. Proof of skein relation for Jones polynomial.

as desired.

Now, for the uniqueness, all we need to show is that J is computable for any diagram D just from the two stated properties. Everything then follows from the well-definedness of the Jones polynomial. For this we simply induct on n, the number of crossings in D.

For any diagram D we may, by Lemma 2.22, change it into a diagram of the unlink by changing some number of its crossings. It is clear then that so long as we can compute J for any diagram with fewer than n crossings, and we can compute J for any diagram of the unlink with some number of components (this is just the first property), that we can compute J for the diagram D by repeated application of the skein relation (second property), which is the desired conclusion. □

Example 2.24. As an example to illustrate the proof of the preceding proposition, suppose we want to calculate the Jones polynomial of the left most link (two interlocking loops) of Figure 14 below. Changing the top crossing will give us a diagram of two disjoint closed loops, i.e. the middle link of Figure 14, which we can certainly compute. The last link involved in the skein relation has only one crossing, that is, one fewer crossing than our original link, which by inductive hypothesis we can also compute. Therefore we are able to compute the Jones polynomial of our original link (two interlocking loops) via the skein relation.

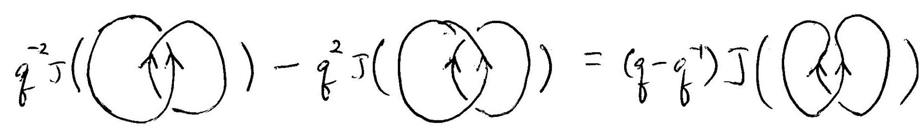


FIGURE 14. Computing Jones polynomial via skein relation.

3. Khovanov homology

In our state-sum formulation of the Kauffman bracket (and hence the Jones polynomial), we considered 2^n states of a n-crossing link diagram and defined the Kauffman bracket as a sum over all 2^n states. With this 'global' approach, we did not use much information about the states themselves.

Khovanov [Kh] obtained a link invariant, known as the *Khovanov homology*, which assigns to each link diagram a chain complex of graded $\mathbb{Z}$-modules or vector spaces, so that this chain complex is homotopy invariant under equivalence of link diagrams.

In this section, we obtain the Khovanov homology via an approach involving *dotted cobordisms*, first suggested by Bar-Natan in [BN]. This approach differs slightly from Khovanov's original approach, but we will see that we eventually recover Khovanov's original formulation.

The plan is as follows. In Sections 3.1, 3.2, 3.3, 3.4, we incrementally build the 'category of cobordisms' that we will require to define the Khovanov homology, along the way highlighting the key considerations of this set-up. In Section 3.5 we define the Khovanov homology, and the subsequent Sections 3.6, 3.7 are devoted to understanding the Khovanov homology in greater detail, in particular how it behaves under the R-moves. In Section 3.8 we do a deframing, exactly as we did with the Kauffman bracket. Finally, Sections 3.9, 3.10 investigate the relation between the Khovanov homology and the Jones polynomial.

In this section, all tangles are assumed to be *circled* tangles (Definition 2.5).

3.1. **Category $\mathscr{C}(P)$.**

3.1.1. We would like to consider the information contained in each state, and the possible relations between states. To that end, we view each state, which is really a tangle diagram with no crossings, as an object in a category.

Definition 3.1. (Category $\mathscr{C}$; objects) Given a finite set of fixed endpoints P on the unit circle, category $\mathscr{C}(P)$ has as its objects tangle diagrams with no crossings, and set of endpoints precisely P, modulo equivalence (isotopy) of diagrams. In particular, if $P = \emptyset$, then $\mathscr{C}(\emptyset)$ has as objects link diagrams with no crossings, modulo equivalence of diagrams.

Remark 3.2. We will incrementally build our definition of $\mathscr{C}(P)$ via a sequence of definitions in Sections 3.1, 3.2, 3.3 and 3.4. The upshot is we would eventually like $\mathscr{C}(P)$ to be a graded, additive category of *dotted* cobordisms, over which we can define chain complexes.

We would also like to represent relations between states.

Definition 3.3. (Category $\mathscr{C}$; morphisms) Given two objects (diagrams) $D_1, D_2 \in \mathscr{C}(P)$, the set of morphisms $\mathrm{Hom}(D_1, D_2)$ is the set of *cobordisms* with 'bottom' boundary D_1 and 'top' boundary D_2, modulo isotopy (continuous deformation) in 3-dimensional space. That is, denoting the unit disk by B, it is the set of 2-dimensional surfaces embedded in the 3-dimensional space (cylinder) $B \times [0,1]$ with bottom boundary precisely $D_1 \subset B \times \{0\}$, top boundary precisely $D_2 \subset B \times \{1\}$, and side boundary $P \times [0,1]$, modulo isotopy in 3-dimensional space.

Composition is defined by vertical stacking in the obvious manner, and the identity morphism for each object is clear (for example, given a diagram D we may take the cobordism $D \times [0,1]$).

3.1.2. Since we allow *all* cobordisms with the stipulated boundaries as morphisms, not all morphisms will be of interest to us. We are particularly interested in the relation or morphisms between *adjacent* states.

Definition 3.4. (Adjacent states) Given two states represented by $s_1, s_2 \in \{0,1\}^n$, we say s_1, s_2 form a pair of *adjacent states* iff s_1, s_2 differ in one and only one position. The state which contains 0 in this position is called the tail of this pair, while the other state (which contains 1) is called the head.

Given a diagram D and smoothings D_1, D_2 corresponding to adjacent states s_1 (tail) and s_2 (head), so that crossing C is 0-smoothed in s_1 and 1-smoothed in s_2, there is always a morphism from D_1 to D_2 obtained by placing a *saddle cobordism* at the crossing C (and the 'identity' surface everywhere else). Essentially, the relation between diagrams D_1 and D_2 of adjacent state is that of merging two components into one, or vice versa, with all other components unchanged.

Example 3.5. Consider the figure-eight link with one crossing represented in Figure 15, and its two states corresponding to the 0-smoothing and 1-smoothing respectively. The morphism in $\mathscr{C}(P)$ between these adjacent states is also depicted in Figure 15; it is an upside-down 'pair of pants' cobordism.

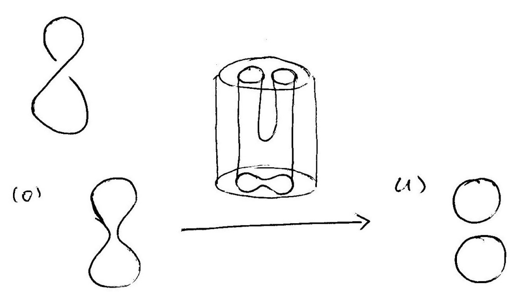


FIGURE 15. Adjacent states of the figure-eight link.

3.2. Chain complexes in $\mathscr{C}(P)$.

3.2.1. We would like to form chain complexes of objects and morphisms in the category $\mathscr{C}(P)$. To do so, we require $\mathscr{C}(P)$ to be an *additive* category, in order for us to have a notion of chain complexes, and of homotopy.

Definition 3.6. (Making $\mathscr{C}(P)$ pre-additive) Henceforth, we make $\mathscr{C}(P)$ a pre-additive category by replacing each hom-set $\mathrm{Hom}(D_1, D_2)$ with the free $\mathbb{Z}$-module generated by its elements; in other words, we allow formal $\mathbb{Z}$-linear combinations of morphisms in $\mathscr{C}(P)$. Composition is defined in the obvious bilinear manner, and identity remains the same.

(Making $\mathscr{C}(P)$ additive) Now that $\mathscr{C}(P)$ is pre-additive, we further make it an additive category by allowing objects to be formal (finite) direct sums of original objects in $\mathscr{C}(P)$, and morphisms therefore as matrices of original morphisms (those of Definition 3.6) of the suitable dimension. Each hom-set is still a $\mathbb{Z}$-module via matrix addition, and composition is defined exactly in the obvious way via matrix multiplication. Identity morphisms are identity matrices.

Remark 3.7. Definition 3.6 provides a universal construction for making any given pre-additive category $\mathscr{C}$ additive (known as the *additive closure* of $\mathscr{C}$). Therefore, even as we modify the objects/morphisms of $\mathscr{C}(P)$ later on, we always understand the final category $\mathscr{C}(P)$ as being additive via this process.

3.2.2. With $\mathscr{C}(P)$ now additive - in particular, there is the notion of the zero morphism in $\mathscr{C}(P)$, we may now define chain complexes with objects and morphisms in $\mathscr{C}(P)$, as well as of (degree-zero) morphisms between chain complexes, in exactly the usual way. The composition of any two successive morphisms in a chain complex is the zero morphism in $\mathscr{C}(P)$.

Definition 3.8. (Indexing of chain complexes) We adopt the convention that our chain complexes are indexed in ascending index order; that is, our chain complex C contains objects $\cdots, C^{-1}, C^0, C^1, \cdots$ and maps $d^i : C^i \to C^{i+1}$ so that $d^{i+1} \circ d^i = 0$.

Given a chain complex C, let $C[s]$ ($s \in \mathbb{Z}$) denote the complex with same differentials and objects $C[s]^{r+s} = C^r$ for all r.

We also have the notion of homotopy which is also defined in exactly the same way as usual. For the sake of completeness, we state the definition here:

Definition 3.9. (Homotopy) Given two chain complexes C, D over $\mathscr{C}(P)$ and two chain complex morphisms $f, g : C \to D$, we say f, g are *homotopic* if there exist morphisms (in $\mathscr{C}(P)$) $s^i : C^i \to D^{i-1}$ so that $f^i - g^i = s^{i+1}d^i + d^{i-1}s^i$ for all i.

We say that two complexes C, D are *homotopically equivalent* if there exist chain complex morphisms $f, g : C \to D$ so that fg and gf are each homotopic to the respective identity morphisms on C and D.

3.3. **Grading.**

3.3.1. Since each hom-set in $\mathscr{C}(P)$ is a $\mathbb{Z}$-module, we may make $\mathscr{C}(P)$ a *graded* category in the sense that each hom-set is a graded $\mathbb{Z}$-module.

Definition 3.10. (Grading on morphisms) For each cobordism(morphism) S in $\mathscr{C}(P)$ (i.e. the original morphisms defined in Definition 3.3), we assign a grading on S by $\deg S := \chi(S) - |P|/2$, where $\chi(S)$ is the usual Euler characteristic of the surface S.

In this way, since each hom-set (as of Definition 3.6) is generated by all such S, each hom-set may be viewed as a graded $\mathbb{Z}$-module.

First, although we did not mention this earlier, in order for $\mathscr{C}(P)$ to contain non-trivial objects (tangle diagrams), the number of endpoints $|P|$ must be even, so that this is still a $\mathbb{Z}$-grading.

Second, the quantity $-|P|/2$ is chosen such that all identity morphisms are therefore of degree zero (each identity morphism contains $|P|/2$ disjoint surfaces which are each simply '2D sheets' of Euler characteristic 1).

Finally, we would like the grading of morphisms to be additive under composition. This is easily verified using the inclusion-exclusion principle for χ and considering the fact that each object of $\mathscr{C}(P)$ (and therefore the boundary along which two cobordisms 'meet' in their composition) is a collection of $|P|/2$ embeddings of $[0,1]$ and some number of closed loops (S^1), which therefore has Euler characteristic $|P|/2$.

Definition 3.11. (Grading on objects) Henceforth, we introduce a grading shift on the objects of $\mathscr{C}(P)$, by replacing the objects D of $\mathscr{C}(P)$ by (formal) objects $D\{d\}$ for all $d \in \mathbb{Z}$, so that D is identified with $D\{0\}$ and there is a $\mathbb{Z}$-action $(d, D) \mapsto D\{d\}$. Each hom-set $\text{Hom}(D_1\{d_1\}, D_2\{d_2\})$ is exactly equal (as a $\mathbb{Z}$-module) to the original hom-set $\text{Hom}(D_1, D_2)$, but with all gradings increased by $(d_2 - d_1)$. It is easily verified that gradings are still additive under composition, and identities are still of degree zero.

Example 3.12. Observe that the saddle or 'pair of pants' cobordism between any two adjacent states, as given in Example 3.5, has grading -1. It is easily verified that this is always true for the saddle cobordism between adjacent states, even in tangle diagrams.

Remark 3.13. Now that we have a grading, we will want all the meaningful morphisms that we talk about (i.e. differentials, chain morphisms, homotopies) to all be of degree zero. In the subsequent discussion we will need to verify or put in special effort to make this happen wherever possible.

Remark 3.14. The philosophical reason for introducing gradings will become clear in Section 3.10; specifically, Lemma 3.37. Roughly speaking, multiplication by the polynomial indeterminate q corresponds to grading shift.

3.4. **Dotted cobordisms.**

3.4.1. Now $\mathscr{C}(P)$ is a graded, additive category over which we can form chain complexes. However, $\mathscr{C}(P)$ contains as morphisms *all* cobordisms without discrimination (cf. Section 3.1.2). In other words, $\mathscr{C}(P)$ is still too general to produce any meaningful theory.

One of the key properties of the Kauffman bracket which makes it meaningful is the ability to 'factor out' single disjoint closed loops as scalars, cf. Lemma 2.16. We essentially have the (local) relation

$$\langle \bigcirc \rangle = (-q)\langle \emptyset \rangle + (-q)^{-1}\langle \emptyset \rangle$$

where $\emptyset$ represents the (locally) empty diagram.

In the same spirit, we would also like to have a local relation in $\mathscr{C}(P)$

$$\bigcirc \cong \emptyset\{1\} \oplus \emptyset\{-1\}$$

where the isomorphism is in the category $\mathscr{C}(P)$. We would also like the isomorphisms involved to be of degree zero, since $\mathscr{C}(P)$ is graded.

Now, the cap morphism $\bigcirc \to \emptyset$ is of degree 1, so that the cap morphism $\bigcirc \to \emptyset\{-1\}$ is of degree 0. However, it is impossible to obtain in this way a degree 0 cap morphism $\bigcirc \to \emptyset\{1\}$. This motivates the following:

Definition 3.15. (Dotted cobordisms) Henceforth, we introduce a *dotting* on the morphisms(cobordisms) of $\mathscr{C}(P)$, as follows. Each cobordism may now carry dots on each of its connected components, modulo free movement of each dot along each connected component. Every dot lowers the grading of the cobordism by 2.

Now, we naturally obtain a singly-dotted cap morphism $\bigcirc \to \emptyset\{1\}$ of degree 0. Similarly, we have a singly-dotted cup morphism $\emptyset\{-1\} \to \bigcirc$ and an undotted cup morphism $\emptyset\{1\} \to \bigcirc$, all of degree 0. The situation is depicted in Figure 16. Note here we do not draw in the boundary cylinder as we have done before, to emphasise the fact that these are local relations.

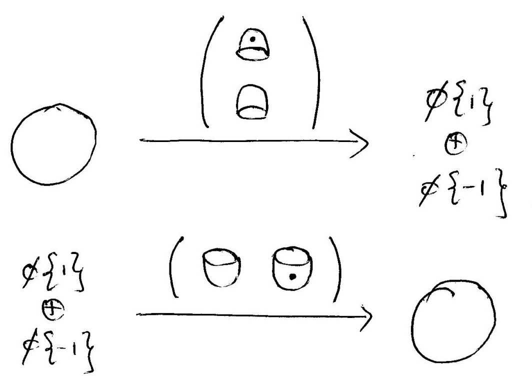


FIGURE 16. Dotted cobordisms(morphisms) of a single closed loop.

Now, we would like these morphisms to be precisely the isomorphisms (in either direction) that we are looking for. To that end, we require:

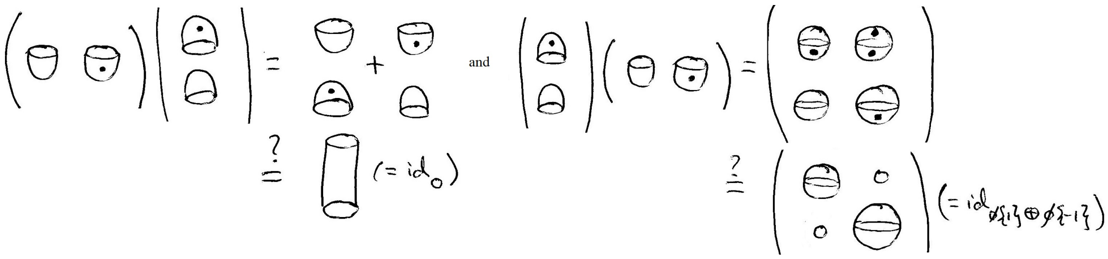


FIGURE 17. Dotted isomorphisms.

This motivates the following.

Definition 3.16. (Relations on dotted cobordisms) Recall that each hom-set of $\mathscr{C}(P)$ forms a (free) $\mathbb{Z}$-module. We now replace each hom-set by its quotient (as $\mathbb{Z}$-module), modulo the following local relations:

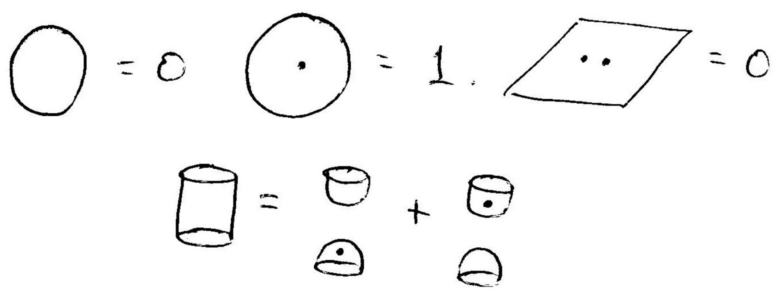


FIGURE 18. Local relations for $\mathscr{C}(P)$.

The last of these relations is often called the *neck-cutting* relation as it allows us to 'cut' a cylindrical component open into two separate components.

First, the notation of the above local relations means precisely the following: we take quotient by the $\mathbb{Z}$-(sub)module R generated by: all cobordisms with a single sphere as one of its components; all cobordisms with a component with at least two dots; the difference of all cobordisms differing by precisely the disjoint addition of a single dotted sphere; the difference between all cobordisms with a (locally) cylindrical component and the sum of the two cobordisms obtained from the local change of the neck-cutting relation.

Second, we need to check the well-definedness of composition. But this follows directly because we mod out by *local* relations only: more precisely, if f, g are morphisms with $f - g \in R$, then for any morphism h, $fh - gh = (f - g)h \in R$ and $hf - hg = h(f - g) \in R$.

Third, we need to check that the gradings are preserved so that each hom-set is still a graded $\mathbb{Z}$-module. But this directly follows because each local relation is degree-homogeneous: for example, a singly-dotted sphere is of degree zero, and similarly for the neck-cutting relation.

Finally, we make $\mathscr{C}(P)$ additive as in Remark 3.7.

Lemma 3.17. *(cf. Lemma 2.16) In $\mathscr{C}(P)$, we have*

$$\bigcirc \cong \emptyset\{1\} \oplus \emptyset\{-1\},$$

with explicit (degree-zero) isomorphisms given in Figure 16.

Proof. Follows from the preceding discussion. □

3.5. Khovanov's homology for links and tangles.

3.5.1. Our definition of $\mathscr{C}(P)$ and of chain complexes over $\mathscr{C}(P)$ is now complete. We are now in a position to define our invariant on (unoriented) link and tangle diagrams, similarly to Definition 2.12.

Definition 3.18. (Khovanov homology, unoriented case) Given an unoriented tangle (or link) diagram D with n crossings and set of endpoints P, let $[\![D]\!]$ be the (bounded) chain complex defined as follows.

We keep the notations of Definition 2.12 and note in addition that for each state $s \in \{0,1\}^n$, $s(D)$ is an object in $\mathscr{C}(P)$.

For all $0 \leq r \leq n$, define

$$[\![D]\!]^r := \bigoplus_{s;i(s)=r} s(D)\{r\}$$

as an object in the (additive) category $\mathscr{C}(P)$ (with grading shift by r). (For $r < 0$ and $r > n$, set $[\![D]\!]^r$ to be the empty object (empty direct sum) in $\mathscr{C}(P)$.)

Let us now define the differential, $d^r : [\![D]\!]^r \to [\![D]\!]^{r+1}$ for all $0 \leq r < n$. (For other r, we certainly take the zero morphism.) Recall that this is essentially a matrix of morphisms(cobordisms) between the direct summands of $[\![D]\!]^r$ and $[\![D]\!]^{r+1}$, and so we want to specify the morphism for each pair of direct summands:

For every two direct summands $s_1(D)$ and $s_2(D)$ ($i(s_1) = r, i(s_2) = r+1$), if s_1, s_2 are not adjacent states, then take the zero morphism. Otherwise, if they are adjacent states (then necessarily s_1 is the tail and s_2 the head), and take as morphism the saddle cobordism described in Example 3.5. In addition, suppose s_1, s_2 differ in the jth position. Then if there are an odd number of 1s in both s_1 and s_2 before the jth position, we take instead the negative of the saddle cobordism (recall our morphisms are formal $\mathbb{Z}$-linear combinations).

First, by the remark in Example 3.12, all our morphisms thus defined are (homogeneous) of degree zero. Second, we need $d^{r+1} \circ d^r = 0$, and it is straightforward to see that the choice of signs for the saddle cobordisms ensures this. Essentially all we need to consider are states s_1, s_2, s_3, s_4 with $i(s_1) = r, i(s_2) = i(s_3) = r+1, i(s_4) = r+2$, such that s_2 differs from s_1 only in the ith position, s_3 differs from s_1 only in the jth position, and s_4 differs from s_1 only in the ith and jth positions. The composition cobordism $s_1 \to s_2 \to s_4$ is equal (up to isotopy) to the composition $s_1 \to s_3 \to s_4$, but with opposite signs.

Example 3.19. Figure 19 illustrates an example in the case of the link consisting of two interlocking loops. Note the grading shifts to ensure that each saddle cobordism is of degree zero, and the signs on the saddle cobordisms to ensure that their composition is zero.

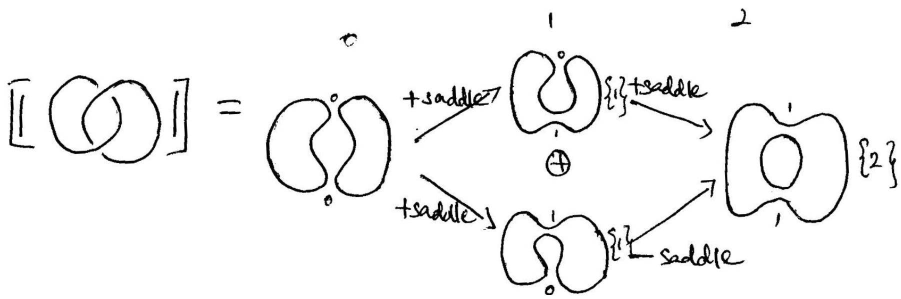


FIGURE 19. Example of Khovanov homology.

3.6. Tangles as part of links; Tensor product of complexes.

3.6.1. We now have a well-defined chain complex $[\![D]\!]$ for each (unoriented) tangle diagram D.

However, as in Remark 2.18, we would like to translate our global definition of the Khovanov homology into something *locally tractable*, since we want to study how $[\![D]\!]$ behaves under the R-moves, and the R-moves are really local relations.

Definition 3.20. (Local part of link diagram) Given a link diagram D, let Γ be a simple closed loop (which we often view as just a circle, up to isotopy), with interior I (and exterior I^c), whose intersection with D is a finite subset of points P, none of which are crossings of D. Further suppose D does not intersect Γ tangentially. Then $D \cap I$ is called a *local part* of D (with respect to Γ) and may be viewed as a tangle diagram with endpoints P, in particular, an object of $\mathscr{C}(P)$.

Furthermore, $D \cap I^c$ may *also* be viewed as a tangle diagram and thus an object $\mathscr{C}(Q)$ for some Q. While this may be intuitively clear from our informal notion of tangles, we make this more precise as follows. First, we may always consider a (crossing-preserving) isotopy of the 2D diagram D (not isotopy of links or tangles), so that all crossings of $D \cap I^c$ are brought into another (disjoint from Γ) disk Γ' with interior I', and D intersects Γ' (again, non-tangentially) at some finite set of points Q. Then $D \cap I'$ is the object of $\mathscr{C}(Q)$ under consideration, after fixing the portion of D outside I and I'. An example is given in Figure 20.

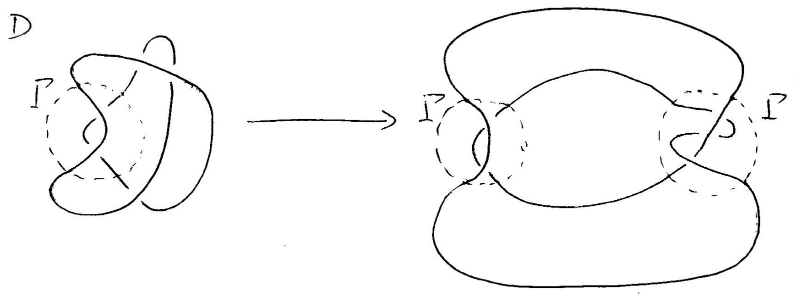


FIGURE 20. Viewing $D \cap I^c$ as an object of $\mathscr{C}(Q)$.

Having fixed the portion of D outside Γ and Γ', as well as fixing Γ and Γ', we now observe that D can be recovered uniquely from $D \cap I$ and $D \cap I'$. In other words, every choice of two tangle diagrams $D_1 \in \mathscr{C}(P)$ and $D_2 \in \mathscr{C}(Q)$ gives rise to a unique link diagram by embedding D_1 in Γ and D_2 in Γ'.

Definition 3.21. ('Tensor' product of diagrams) As above, fix the portion of D outside Γ and Γ', as well as fix Γ and Γ'. For any two tangle diagrams $D_1 \in \mathscr{C}(P)$ and $D_2 \in \mathscr{C}(Q)$, denote by $D_1 \otimes D_2$ the link diagram obtained by embedding D_1 in Γ and D_2 in Γ'. We also extend $\otimes$ to direct sums of diagrams (objects) in $\mathscr{C}(P)$ and $\mathscr{C}(Q)$ in the obvious manner.

Observe also that for any two morphisms(cobordisms) $f \in \mathscr{C}(P)$ and $g \in \mathscr{C}(Q)$, we may form $f \otimes g$ by embedding the cobordisms f, g within $\Gamma \times [0,1]$ and $\Gamma' \times [0,1]$ respectively, and taking the identity surface everywhere else. There is also the obvious extension of $\otimes$ to matrices of morphisms which is compatible with the extension of $\otimes$ to direct sums of objects.

The other key point is that all crossings of D are contained within Γ and Γ'. In particular, a (complete) smoothing of D is in correspondence with (complete) smoothings of $D \cap I$ and $D \cap I'$.

Lemma 3.22. *Suppose $D_1 := D \cap I$ has n_1 crossings and $D_2 := D \cap I'$ has n_2 crossings, and the ordering of crossings in D is such that the crossings in $D \cap I$ are ordered first. For each state $s \in \{0,1\}^n$, let $s_1 \in \{0,1\}^{n_1}, s_2 \in \{0,1\}^{n_2}$ correspond to the first n_1 and last n_2 positions of s respectively. Then*

$$s(D) = s_1(D_1) \otimes s_2(D_2).$$

Proof. Obvious. □

The last point that we have yet to consider is the grading. We would certainly like to define $\otimes$ so that the grading and grading shifts are additive with respect to it. But since we have already explicitly defined the grading and $\otimes$ on cobordisms, this is something we need to check.

Lemma 3.23. *(Additivity of grading with respect to $\otimes$) Suppose $f \in \mathscr{C}(P)$ and $g \in \mathscr{C}(Q)$ are cobordisms of respective degree d_1, d_2. Then $f \otimes g \in \mathscr{C}(\emptyset)$ has degree $d_1 + d_2$.*

Proof. Observe that the portion of D outside Γ and Γ' that we previously fixed is really a collection of $(|P|+|Q|)/2$ embeddings of $[0,1]$, and so their identity surface is really a collection of $(|P|+|Q|)/2$ '2D sheets' which will be attached to f, g along $|P|+|Q|$ boundaries of the form $[0,1]$. Therefore

$$\chi(f \otimes g) = \chi(f) + \chi(g) + (|P|+|Q|)/2 - (|P|+|Q|) = (\chi(f) - |P|/2) + (\chi(g) - |Q|/2),$$

as desired. □

With Lemma 3.23, we are assured that $\otimes$ is well-defined with respect to the grading.

Definition 3.24. We extend $\otimes$ to grade-shifted objects in the obvious additive manner, so that the grading of morphisms of grade-shifted objects is also additive under $\otimes$.

3.6.2. Given D and a local part $D \cap I$, we would like to study how $\llbracket D \rrbracket$ changes when $D \cap I$ (locally) changes. Then, rather than considering relations between entire $\llbracket D \rrbracket$, we can consider relations between local $\llbracket D \cap I \rrbracket$. We hence want to consider first the relation between $\llbracket D \rrbracket$ and $\llbracket D \cap I \rrbracket$ (and therefore by the symmetry of our preceding discussion, also the relation between $\llbracket D \cap I' \rrbracket$).

The preceding discussion (in particular Lemma 3.22) is strong indication that we should view $\llbracket D \rrbracket$ as the usual tensor product of complexes $\llbracket D \cap I \rrbracket$ and $\llbracket D \cap I' \rrbracket$. We keep the notations of Lemma 3.22.

Proposition 3.25. *$\llbracket D \rrbracket$ is the "tensor product" of complexes $\llbracket D_1 \rrbracket$ and $\llbracket D_2 \rrbracket$, in the following (usual) sense: $\llbracket D \rrbracket$ is the "total complex" of the "double complex" formed by $\llbracket D_1 \rrbracket$ and $\llbracket D_2 \rrbracket$, as in the following picture, where d_1, d_2 denote the respective differentials in $\llbracket D_1 \rrbracket$ and $\llbracket D_2 \rrbracket$:*

$$\begin{array}{ccccccccc}
 & & \cdots & & \cdots & & \cdots & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
\cdots & \longrightarrow & \llbracket D_1 \rrbracket^{p-1} \otimes \llbracket D_2 \rrbracket^{q-1} & \xrightarrow{id\otimes(-1)^{p-1}d_2^{q-1}} & \llbracket D_1 \rrbracket^{p-1} \otimes \llbracket D_2 \rrbracket^{q} & \xrightarrow{id\otimes(-1)^{p-1}d_2^{q}} & \llbracket D_1 \rrbracket^{p-1} \otimes \llbracket D_2 \rrbracket^{q+1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle d_1^{p-1}\otimes id} & & \downarrow{\scriptstyle d_1^{p-1}\otimes id} & & \downarrow{\scriptstyle d_1^{p-1}\otimes id} & & \\
\cdots & \longrightarrow & \llbracket D_1 \rrbracket^{p} \otimes \llbracket D_2 \rrbracket^{q-1} & \xrightarrow{id\otimes(-1)^{p}d_2^{q-1}} & \llbracket D_1 \rrbracket^{p} \otimes \llbracket D_2 \rrbracket^{q} & \xrightarrow{id\otimes(-1)^{p}d_2^{q}} & \llbracket D_1 \rrbracket^{p} \otimes \llbracket D_2 \rrbracket^{q+1} & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle d_1^{p}\otimes id} & & \downarrow{\scriptstyle d_1^{p}\otimes id} & & \downarrow{\scriptstyle d_1^{p}\otimes id} & & \\
\cdots & \longrightarrow & \llbracket D_1 \rrbracket^{p+1} \otimes \llbracket D_2 \rrbracket^{q-1} & \xrightarrow{id\otimes(-1)^{p+1}d_2^{q-1}} & \llbracket D_1 \rrbracket^{p+1} \otimes \llbracket D_2 \rrbracket^{q} & \xrightarrow{id\otimes(-1)^{p+1}d_2^{q}} & \llbracket D_1 \rrbracket^{p+1} \otimes \llbracket D_2 \rrbracket^{q+1} & \longrightarrow & \cdots \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & \cdots & & \cdots & & \cdots & &
\end{array}$$

Then $\llbracket D \rrbracket$ has objects formed by the usual 'diagonal direct sums' of the total complex

$$\llbracket D \rrbracket^r = \bigoplus_{p+q=r} \llbracket D_1 \rrbracket^p \otimes \llbracket D_2 \rrbracket^q$$

and differentials obtained accordingly as in the picture.

We denote this chain complex by $[\![D_1]\!] \otimes [\![D_2]\!]$. In other words, $[\![D]\!] = [\![D_1]\!] \otimes [\![D_2]\!]$.

Proof. That the objects coincide follows from Lemma 3.22, that $\otimes$ commutes with $\oplus$, and the additivity of grading shift over $\otimes$.

As for the differentials, the key point is this: for each state $s \in \{0,1\}^n$ with $i(s) = r$, $i(s_1) = p, i(s_2) = q$, consider all adjacent states s' with $i(s') = r + 1$ and similarly define $s'_1 \in \{0,1\}^{n_1}, s'_2 \in \{0,1\}^{n_2}$. Then we must have either $i(s'_1) = p+1, i(s'_2) = q$, or $i(s'_1) = p, i(s'_2) = q+1$. Conversely, any two states $s'_1 \in \{0,1\}^{n_1}, s'_2 \in \{0,1\}^{n_2}$ with $i(s'_1) = p+1, i(s'_2) = q$, or $i(s'_1) = p, i(s'_2) = q+1$, define an adjacent state $s' = s'_1 + s'_2$ of s.

In the former case, then the morphism $s \to s'$ in $[\![D]\!]$ is precisely the tensor product $(s_1 \to s'_1) \otimes \mathrm{id}$ of morphisms in $[\![D_1]\!], [\![D_2]\!]$. If the latter, then the morphism is $\mathrm{id} \otimes (-1)^p(s_2 \to s'_2)$. (Recall how signs are placed on our cobordisms.) It is straightforward then to see that the differentials coincide. □

To be complete, we record the following easily-verified fact.

Lemma 3.26. *(Additivity of grading shift and height shift) Grading shifts on complexes (all objects grade shifted by the same amount) and height shifts on complexes are additive with respect to $\otimes$.*

Proof. The additivity of grading shift follows from that for objects, while the additivity for height shift is obvious. □

3.6.3. The preceding discussion tells us that we are dealing with nothing but the usual tensor product of complexes; in particular, we can apply almost exactly all the usual theory of homological algebra unchanged.

Definition 3.27. (Tensor product of chain morphisms) Suppose D_1, D'_1, D_2, D'_2 are tangle diagrams (with same set of endpoints P) and $[\![D_1]\!], [\![D'_1]\!], [\![D_2]\!], [\![D'_2]\!]$ their corresponding complexes, and link diagrams $D = D_1 \otimes D_2, D' = D'_1 \otimes D'_2$ with $[\![D]\!] = [\![D_1]\!] \otimes [\![D_2]\!], [\![D']\!] = [\![D'_1]\!] \otimes [\![D'_2]\!]$. Suppose also $f : [\![D_1]\!] \to [\![D'_1]\!]$, $g : [\![D_2]\!] \to [\![D'_2]\!]$ are chain complex morphisms (cf. Section 3.2.2). Then the tensor product $f \otimes g : [\![D]\!] \to [\![D']\!]$ is defined so that $(f \otimes g)^r$ sends $[\![D_1]\!]^p \otimes [\![D_2]\!]^q$ to $[\![D'_1]\!]^p \otimes [\![D'_2]\!]^q$ via the morphism $f^p \otimes g^q$ (for all $p + q = r$).

Lemma 3.28. *(Additivity of grading with respect to $\otimes$) Suppose f and g are chain morphisms of respective graded degree d_1, d_2. Then $f \otimes g$ has degree $d_1 + d_2$.*

Proof. Follows from Lemma 3.23. □

Lemma 3.29. *(Homotopy equivalence of tensor product) Keeping the notations of Definition 3.27 with $D_2 = D'_2$, suppose $f_1, f_2 : [\![D_1]\!] \to [\![D'_1]\!]$ are homotopic chain morphisms. Then $f_1 \otimes id$, $f_2 \otimes id : [\![D]\!] \to [\![D']\!]$ are homotopic. (cf. Definition 3.9)*

Proof. Define morphisms $s'^r : [\![D]\!]^r \to [\![D']\!]^{r-1}$ by mapping $[\![D_1]\!]^p \otimes [\![D_2]\!]^q$ to $[\![D_1']\!]^{p-1} \otimes [\![D_2]\!]^q$ via the morphism $s^p \otimes \mathrm{id}$ for all $p+q=r$. Then $(f_1 \otimes \mathrm{id})^r - (f_2 \otimes \mathrm{id})^r$ maps $[\![D_1]\!]^p \otimes [\![D_2]\!]^q$ to $[\![D_1']\!]^p \otimes [\![D_2]\!]^q$ via

$$\begin{aligned}
&(f_1^p - f_2^p) \otimes \mathrm{id} \\
&= (s^{p+1} d_1^p + d_1'^{p-1} s^p) \otimes \mathrm{id} \\
&= \Big((s^{p+1} \otimes \mathrm{id})(d_1^p \otimes \mathrm{id}) + (s^p \otimes \mathrm{id})(\mathrm{id} \otimes (-1)^p d_2^q) \Big) \\
&\quad + (d_1'^{p-1} \otimes \mathrm{id} + \mathrm{id} \otimes (-1)^{p-1} d_2^q)(s^p \otimes \mathrm{id}) \\
&\quad - s^p \otimes (-1)^p d_2^q - s^p \otimes (-1)^{p-1} d_2^q \\
&= s'^{r+1} d^r + d^{r-1} s'^r.
\end{aligned}$$

□

Corollary 3.30. *(Homotopy equivalence of tensor product) Keeping the notations of Definition 3.27 with $D_2 = D_2'$, suppose $f : [\![D_1]\!] \to [\![D_1']\!]$ is a homotopy equivalence between $[\![D_1]\!], [\![D_1']\!]$ with inverse $g : [\![D_1']\!] \to [\![D_1]\!]$ (such that fg and gf are each homotopic to the respective identity morphisms). Then $f \otimes id$ is a homotopy equivalence between $[\![D]\!], [\![D']\!]$ (with inverse $g \otimes id$). Observe also that if f is of degree zero, then $f \otimes id$ is also of degree zero.*

Proof. Immediate from Lemma 3.29. □

3.7. **Invariance of $[\![\cdot]\!]$ under R-moves.**

Corollary 3.30 is the key tool which therefore allows us to reduce our study of $[\![\cdot]\!]$ modulo the R-moves to the study of $[\![\cdot]\!]$ on the local parts (tangles) involved in each R-move. In this section we aim to do just that.

3.7.1. There is one point that we should first address and that is the order of crossings in the definition of $[\![\cdot]\!]$, cf. Remark 2.15.

Lemma 3.31. $[\![\cdot]\!]$ *is homotopy invariant under a change in the order of crossings.*

Proof. Because permutations are generated by transpositions, it suffices to consider the case where two adjacent crossings are transposed. Essentially the only change that results is a sign flip on all morphisms from states of the form $\cdots 01 \cdots \to \cdots 11 \cdots$ or $\cdots 10 \cdots \to \cdots 11 \cdots$. But then we have an isomorphism between the two complexes by mapping all states of the form $\cdots 11 \cdots$ to themselves via $-\mathrm{id}$, and all other states via id. □

3.7.2. Let us now study how $[\![\cdot]\!]$ behaves under each R-move.

We begin first with a (somewhat standard) lemma that we will use to simplify complexes up to homotopy, which we will henceforth call the ‘cancellation lemma’. The key idea is to ‘cancel’ isomorphisms which appear within the complex up to homotopy.

Lemma 3.32. *('Cancellation lemma') Given the following complex* Ω *over* $\mathscr{C}(P)$*:*

$$\cdots \longrightarrow A \xrightarrow{\begin{pmatrix} a \\ b \end{pmatrix}} B \oplus C \xrightarrow{\begin{pmatrix} \phi & d \\ c & e \end{pmatrix}} D \oplus E \xrightarrow{\begin{pmatrix} f & g \end{pmatrix}} F \longrightarrow \cdots$$

where $\phi : B \to D$ *is an isomorphism, and each of* A, B, C, D, E, F *are (finite) direct sums of objects in* $\mathscr{C}(P)$*, so that the morphisms are each block matrices of the appropriate dimension. Then we have the following isomorphism of complexes*

$$\begin{array}{ccccccccc}
\cdots & \longrightarrow & A & \xrightarrow{\begin{pmatrix} a \\ b \end{pmatrix}} & B \oplus C & \xrightarrow{\begin{pmatrix} \phi & d \\ c & e \end{pmatrix}} & D \oplus E & \xrightarrow{\begin{pmatrix} f & g \end{pmatrix}} & F & \longrightarrow & \cdots \\
 & & \downarrow{\scriptstyle id} & & \downarrow{\scriptstyle \begin{pmatrix} id & \phi^{-1}d \\ 0 & id \end{pmatrix}} & & \downarrow{\scriptstyle \begin{pmatrix} id & 0 \\ -c\phi^{-1} & id \end{pmatrix}} & & \downarrow{\scriptstyle id} & & \\
\cdots & \longrightarrow & A & \xrightarrow{\begin{pmatrix} 0 \\ b \end{pmatrix}} & B \oplus C & \xrightarrow{\begin{pmatrix} \phi & 0 \\ 0 & e-c\phi^{-1}d \end{pmatrix}} & D \oplus E & \xrightarrow{\begin{pmatrix} 0 & g \end{pmatrix}} & F & \longrightarrow & \cdots
\end{array}$$

This second complex has a direct summand

$$0 \longrightarrow B \xrightarrow{\phi} D \longrightarrow 0$$

which is homotopic to the zero complex, and so the original complex Ω *is homotopic to the complex*

$$\cdots \longrightarrow A \xrightarrow{b} C \xrightarrow{e-c\phi^{-1}d} E \xrightarrow{g} F \longrightarrow \cdots$$

That is, we have cancelled the isomorphism $\phi : B \to D$*, up to a change in morphism* $C \to E$*. If all morphisms are degree zero then all isomorphisms and homotopies are also degree zero.*

Proof. Each part of the lemma is a straightforward verification. $\square$

3.7.3. Under (R1), as in Proposition 2.19, we consider two cases. In the over-crossing case, we consider the complex and degree-zero isomorphism (by the key Lemma 3.17).

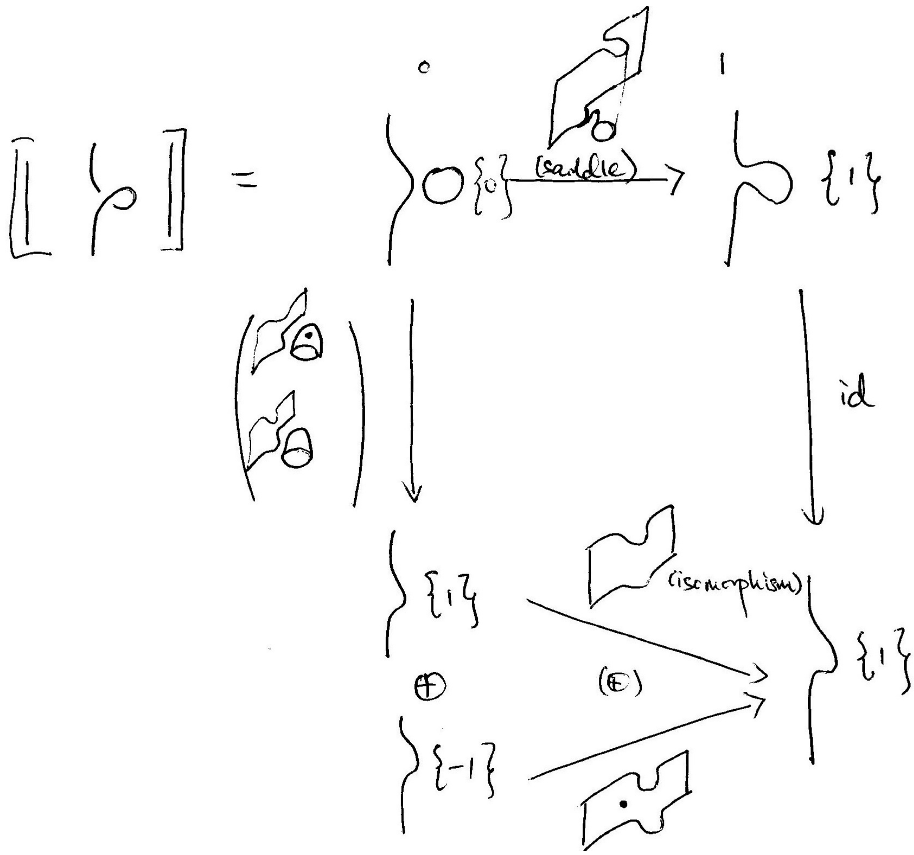


FIGURE 21. Complexes for (R1). Here we include all the morphisms in full for illustration.

Now, cancelling the morphism marked 'isomorphism' by Lemma 3.32, we see that the second complex is degree-zero homotopic to the single term complex but with a grading shift of −1. We obtain similarly the relation for the under-crossing case. In other words, we have:

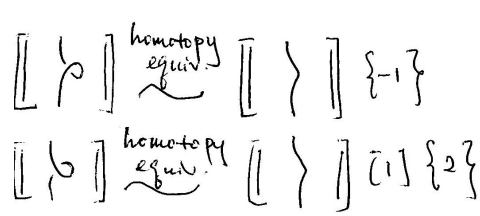


FIGURE 22. Complexes under (R1).

3.7.4. For (R2), exactly the same techniques apply as in (R1). For brevity, we will omit the cobordisms which can be inferred from our pictures (one may just bear in mind that all cobordisms, isomorphisms and homotopies are degree zero).

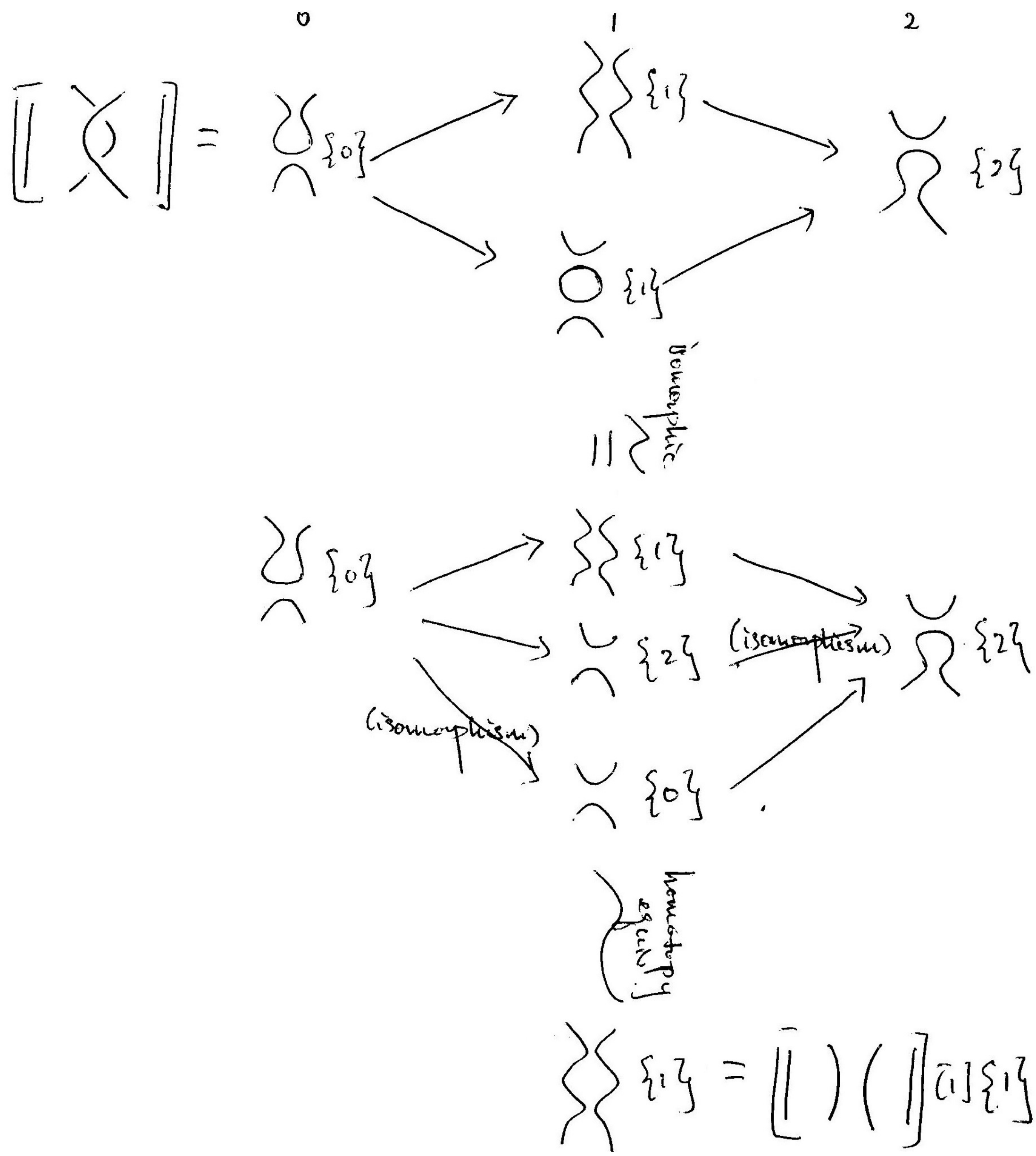


FIGURE 23. Complexes under (R2).

3.7.5. Under (R3), the situation is significantly more complex. While exactly the same techniques apply, the key issue we have to deal with is that the isomorphisms we want to cancel by Lemma 3.32 appear in the 'middle' of our complexes. Whereas in (R1) and (R2) we did not have to worry about the change in morphism $C \to E$ that occurs in Lemma 3.32, because either $C = 0$ or $E = 0$ (and thus we could 'cancel' the isomorphism without further comment), this is not the case in (R3), and we will have to do some computations regarding $C \to E$. However, the essential idea is still exactly the same as that for (R1) and (R2).

Since we have to deal with the actual (saddle) cobordisms that occur, we introduce some notation. First, observe that all the tangle diagrams we will ever deal with under (R3) have exactly $|P| = 6$ endpoints, so we fix these 6 points on a circle and represent all tangle diagrams in this circle. Now each saddle cobordism will involve 4 of these 6 points, with two components in the source and two components in the target. Our notation for tangle diagrams and saddle cobordisms is then largely self-explanatory from the following example:

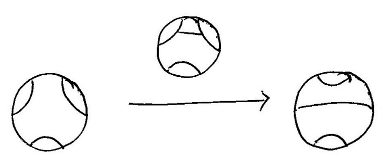


FIGURE 24. Example of notation.

Now we deal with each of the two complexes that arise from either side of (R3) separately. We have

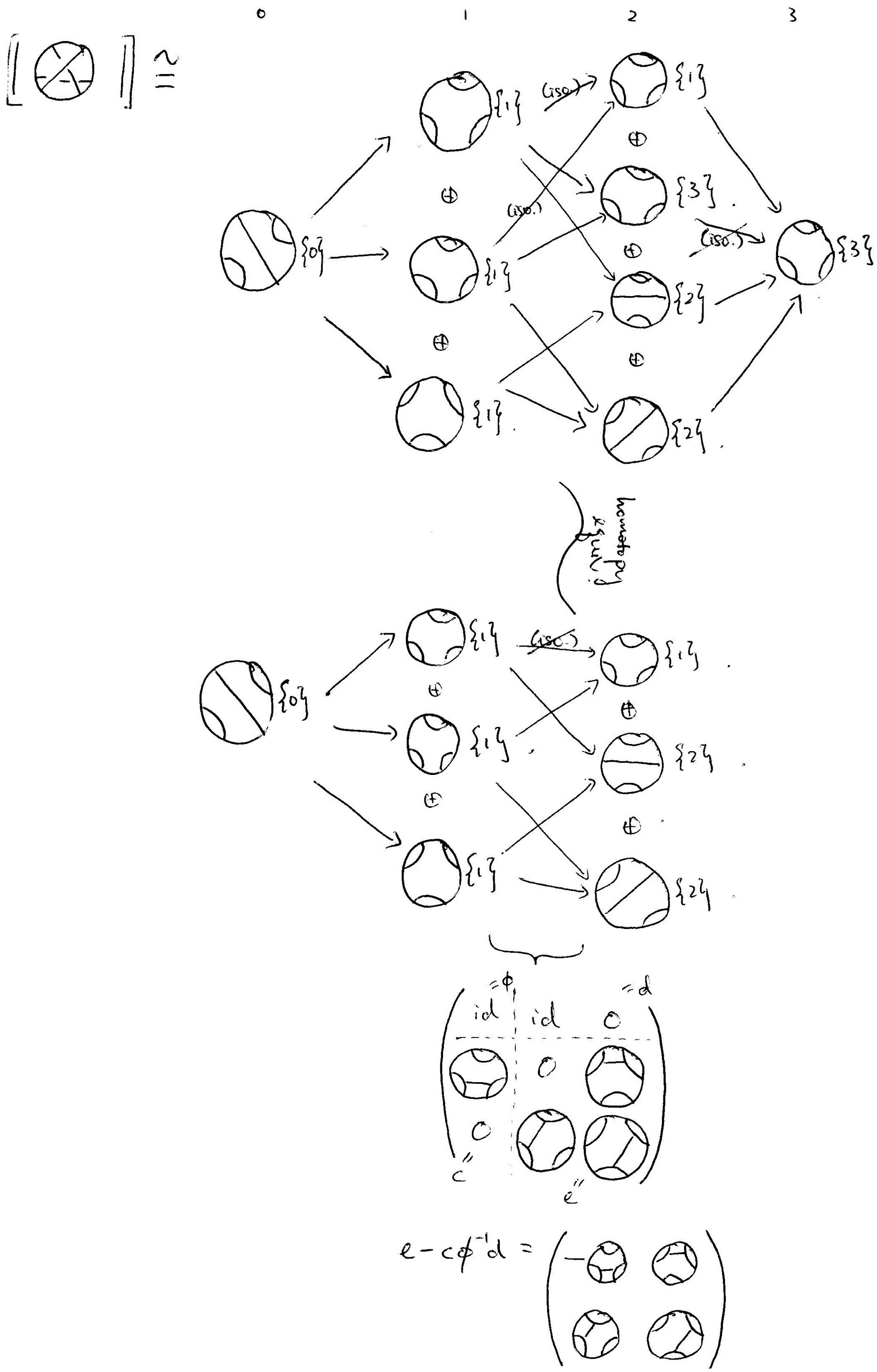


Figure 25. One side of (R3).

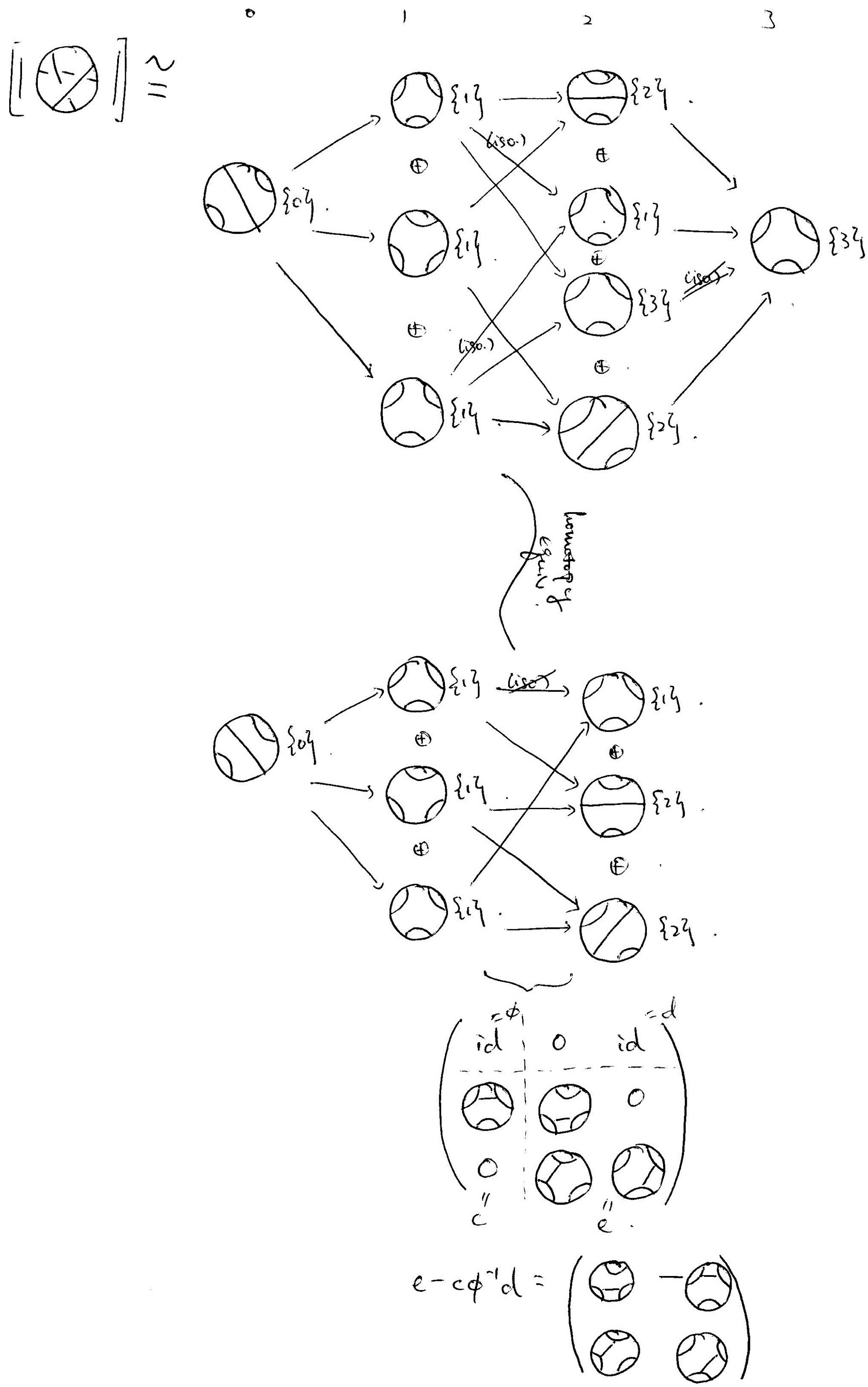


FIGURE 26. Other side of (R3). Note the rearrangement of direct summands in the height 2 object of the second complex still gives rise to an isomorphic complex (via a permutation matrix).

so that the two complexes are each homotopic to isomorphic complexes (differing only in sign on the first row of the 2×2 morphisms $e - c\phi^{-1}d$) and thus to each other (with no height or grading shifts needed).

3.8. **Deframing.**

3.8.1. The situation now is as follows. Under (R1-3), $[\![\cdot]\!]$ is not homotopy invariant, but is homotopy invariant up to height and grading shifts. This tells us that we may be able to obtain a true invariant of *oriented* links via the process of deframing, exactly as in Section 2.3.3.

We can summarise the situation as follows: we have degree-zero homotopy equivalences

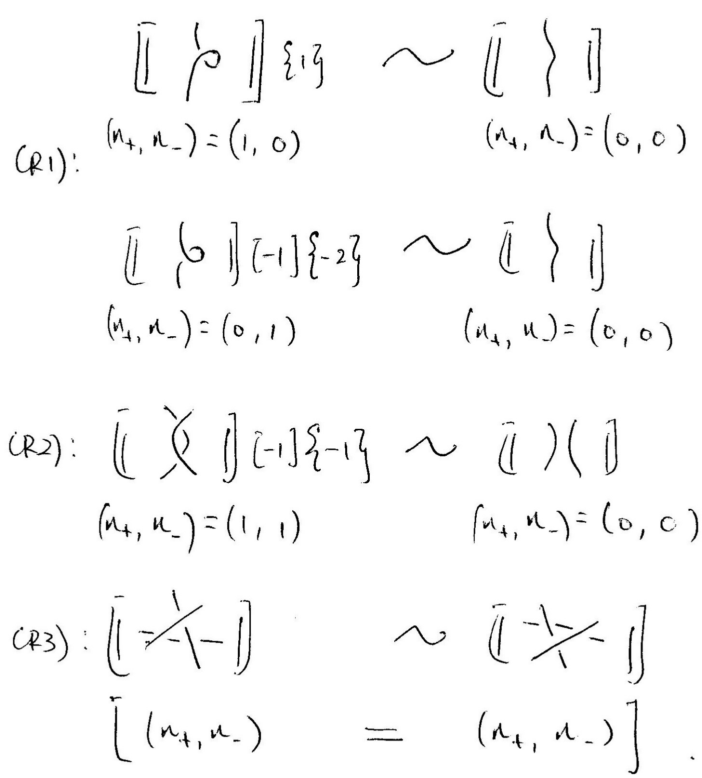


FIGURE 27. Summary of homotopy equivalences under (R1-3).

Furthermore, n_+, n_-, grading shifts and height shifts are all additive under $\otimes$ (Lemma 3.26). Consequently, we finally obtain the following.

Theorem 3.33. *(Khovanov homology) Given an oriented link diagram D with underlying unoriented link diagram D' (obtained by 'forgetting' the orientation of D), the quantity*

$$Kh(D) := [\![D']\!][-n_-(D)]\{n_+(D) - 2n_-(D)\}$$

is an invariant of oriented links.

Proof. Follows from the preceding discussions and Figure 27. □

3.9. **Obtaining a computable invariant.**

3.9.1. We would ideally like to have *computable* link invariants. Our current formulation of $Kh(D)$ suffers from two problems in this regard: first, it is defined over the abstract category $\mathscr{C}(P)$; second, it is in general not so easy to compute homotopy.

Ideally, one would like to work over an *abelian* category of, say, (graded) modules, so that we may take homology, whose invariance would follow from that of homotopy.

To do so, we will need to have a functor from our category $\mathscr{C}$ to our desired abelian category. Our invariant then becomes a chain complex over this abelian category, and homotopy invariance is preserved, from which invariance of homology groups follows.

Furthermore, since we are interested only in obtaining a link invariant (the Jones polynomial is defined only for links), we need only define our functor on $\mathscr{C}(\emptyset)$. Homotopy invariance is still preserved since our previous results imply homotopy invariance within $\mathscr{C}(\emptyset)$ for links.

Now, observe that the hom-sets in our category $\mathscr{C}$ are all, by construction, graded $\mathbb{Z}$-modules. The most natural choice for our functor is therefore

Definition 3.34. We define our functor F from $\mathscr{C}(\emptyset)$ to the category of graded $\mathbb{Z}$-modules as follows. For each object $D \in \mathscr{C}(\emptyset)$, we set $F(D) = \mathrm{Hom}(\emptyset, D)$ (where $\emptyset$ is the empty object, or empty diagram). Then there is only one natural choice for the morphisms in $\mathscr{C}(\emptyset)$ and that is composition (on the left): for each morphism $\phi : D_1 \to D_2$ in $\mathscr{C}(\emptyset)$ we set $F(\phi) = (f \mapsto \phi \circ f)$ for all $f \in \mathrm{Hom}(\emptyset, D_1)$.

We would now like to look at the structure of the graded $\mathbb{Z}$-modules thus obtained. Observe that each object $D \in \mathscr{C}(\emptyset)$ is nothing but a collection of disjoint closed loops (since it is a smoothing of a link diagram). Therefore, by application of the neck-cutting relation (Definition 3.16), we can reduce any morphism (cobordism) in $\mathrm{Hom}(\emptyset, D)$ to nothing but (a sum of) cobordisms where to each closed loop in D corresponds a disk (or a cup, if you like, since we usually present our morphisms as going from the bottom up) which is either undotted or dotted.

Now, an undotted disk is of degree 1 and a dotted disk is of degree -1. We therefore have

Lemma 3.35. *If $D \in \mathscr{C}(\emptyset)$ is a collection of n disjoint closed loops, then $F(D)$ is as a graded $\mathbb{Z}$-module isomorphic to the n-fold tensor product of the free graded $\mathbb{Z}$-module V_0 with two generators v_+, v_- of homogeneous degree 1 and -1 respectively.*

Proof. The idea is as in the preceding discussion. □

One can then compute what the saddle cobordisms (Example 3.5) should map to under F. We will not go into the details here, as we are not so much interested in the details of this formulation as much as we would like to understand its relation to the

Jones polynomial. The upshot is that we can recover Khovanov's original formulation of his homology invariant defined over graded $\mathbb{Z}$-modules in this way.

3.10. **Decategorification; Jones polynomial.**

In what sense is the Khovanov homology a *categorification* of the Jones polynomial? Very roughly speaking, categorification is the process of realising a 'basic' structure (sets, groups, etc.) as being derived from the higher structure of a *category*.

For example, the finite-dimensional vector spaces over some field forms a category. However, since we are really only interested in objects up to isomorphism when working within a category, we know also that up to isomorphism the finite-dimensional vector spaces are completely 'parameterised' by the non-negative integers, i.e. by their dimension. Therefore, in this rough sense, the category of finite-dimensional vector spaces *categorifies* the set, or group of integers, and conversely, *decategorifying* the category of finite-dimensional vector spaces 'recovers' the set or group of integers.

Of course, we would like to have at least some precise notion of what it means to 'decategorify' a category. Because, as mentioned, we are looking at objects up to isomorphism, the right notion to consider is that of the Grothendieck group of an (abelian) category.

Definition 3.36. (Grothendieck group of an abelian category) Let C be any abelian category. The Grothendieck group $K(C)$ is the free abelian group generated by the *isomorphism classes* of the objects of C, denoted $[A]$ for each object $A \in C$, modulo the following relations: for each short exact sequence $0 \to A' \to A \to A'' \to 0$, we have the relation $[A] = [A'] + [A'']$.

Lemma 3.37. *The Grothendieck group of the category of finite-dimensional graded vector spaces (over some field, say $\mathbb{Q}$) is isomorphic to $\mathbb{Z}[q, q^{-1}]$.*

Proof. Just as the finite-dimensional vector spaces are completely 'parameterised' by the non-negative integers, i.e. by their dimension, the finite-dimensional *graded* vector spaces are completely 'parameterised' by their *graded dimension*. □

Combining Definition 2.12, Lemmas 3.35 and 3.37, it is clear now how to recover the Jones polynomial from the Khovanov homology: after tensoring with a suitable field (say $\mathbb{Q}$), the alternating sum of the graded dimensions of the objects in our chain complex recovers the Kauffman bracket, which combined with our deframing process for both invariants, recovers the Jones polynomial. But the alternating sum of graded dimensions of objects is also the alternating sum of graded dimensions of homology groups, or equivalently it is the *Euler characteristic* of our chain complex (with respect to the Euler-Poincaré mapping of taking graded dimensions). Consequently, the Khovanov homology is, in the sense we have defined, a categorification of the Jones polynomial.

Remark 3.38. Why should the Euler characteristic be the *right* notion of decategorification, given a chain complex? Roughly speaking, the Grothendieck group we have constructed is in fact the *universal* Euler-Poincaré mapping with respect to isomorphism classes of objects in C.

4. The quantum group $\mathbf{U}_q(\mathfrak{sl}_2)$

In the preceding sections, we have shown how an approach to the Kauffman bracket (Jones polynomial) can extend to the Khovanov homology, and how the Khovanov homology can reproduce the Jones polynomial. However, we have not shown how the Jones polynomial may possibly arise beyond knowing beforehand its formulation (commonly in terms of skein relation).

The key idea is that we want to produce knot invariants in a more motivated manner, by starting with the three Reidemeister moves and formulating them as (algebraic) relations that are to be satisfied, rather than formulating the knot invariants first and then checking them against the three R-moves.

In this section and the next, we will consider how the Jones polynomial can naturally arise from representations of the so-called *quantum group* $\mathbf{U}_q(\mathfrak{sl}_2)$. We will do so via an approach of Reshitikhin-Turaev ([RT]) (and in fact this approach extends more generally for any (complex) semisimple Lie algebra $\mathfrak{g}$), and in the process show also how this leads naturally to the setting of the *coloured* tangles which they originally considered.

4.1. **Preliminaries.**

In order to have a fuller picture of the setting, let us begin with some preliminaries from abstract algebra. Fix a ground field k.

4.1.1. There are several equivalent ways to view the notion of an *algebra* over the field k. Here, we fix a point of view that will be most useful to us, for reasons which will come clear.

Definition 4.1. An *algebra* (A, μ, η) over k is a vector space over k equipped with a bilinear map $\mu : A \otimes A \to A$ (the 'multiplication') and a linear map $\eta : k \to A$ (the 'unit'), so that the multiplication is associative (i.e. the following diagram commutes):

$$\begin{array}{ccc} A \otimes A \otimes A & \xrightarrow{\mu \otimes \mathrm{id}} & A \otimes A \\ \downarrow{\scriptstyle \mathrm{id} \otimes \mu} & & \downarrow{\scriptstyle \mu} \\ A \otimes A & \xrightarrow{\quad \mu \quad} & A \end{array}$$

and the unit is compatible with the vector space structure (i.e. the following diagrams commute):

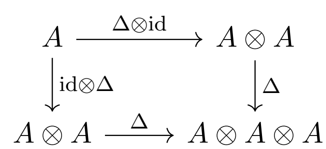
$$\begin{array}{ccc} k \otimes A & \xrightarrow{\eta \otimes \mathrm{id}} & A \otimes A \\ & \searrow^{can} & \downarrow{\scriptstyle \mu} \\ & & A \end{array} \qquad \begin{array}{ccc} A \otimes k & \xrightarrow{\mathrm{id} \otimes \eta} & A \otimes A \\ & \searrow^{can} & \downarrow{\scriptstyle \mu} \\ & & A \end{array}$$

where *can* is the canonical identification of $k \otimes A$ with A (as vector spaces).

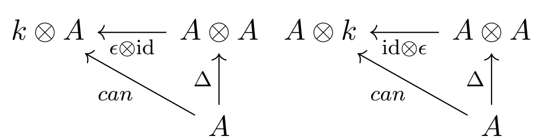
Given two A-modules M_1, M_2, we would like to form their tensor product $M_1 \otimes M_2$ which is also an A-module. However, with no other conditions on A, $M_1 \otimes M_2$ can only be made an $(A \otimes A)$-module. In order to make $M_1 \otimes M_2$ an A-module, we therefore need an algebra morphism $\Delta : A \to A \otimes A$ so that $a \in A$ acts on $m_1 \otimes m_2 \in M_1 \otimes M_2$ as $\Delta(a)$. Furthermore, we naturally want also for $\otimes$ to satisfy associativity; that is, $(M_1 \otimes M_2) \otimes M_3$ should be naturally isomorphic to $M_1 \otimes (M_2 \otimes M_3)$ (via $(m_1 \otimes m_2) \otimes m_3 \mapsto m_1 \otimes (m_2 \otimes m_3)$. Therefore Δ should satisfy some sort of 'associativity' condition.

We also want the ground field k to act as a 'unit' A-module; so k should have an A-module structure and $k \otimes M$ (resp. $M \otimes k$) should be identified with M. We therefore need an algebra morphism $\epsilon : A \to k$ so that $a \in A$ acts on k as (multiplication by) $\epsilon(a)$. This leads to the following:

Definition 4.2. A *bialgebra* $(A, \mu, \eta, \Delta, \epsilon)$ over k is an algebra over k equipped with an algebra morphism $\mu : A \otimes A \to A$ (the 'comultiplication') and an algebra morphism $\epsilon : A \to k$ (the 'counit'), so that the comultiplication is coassociative (i.e. the following diagram commutes):



and the counit satisfies (i.e. the following diagrams commute):



where *can* is the canonical identification of $k \otimes A$ with A (as vector spaces).

Remark 4.3. Observe that the conditions satisfied by Δ, ϵ are *dual* to that of μ, η in that the arrows in the respective commutative diagrams are systematically reversed. There is accordingly a notion of a *coalgebra* A which has only linear maps Δ, ϵ (satisfying the same conditions). A bialgebra is then both an algebra and a coalgebra such that their structures are compatible, i.e. Δ and ϵ are algebra morphisms.

Remark 4.4. In the case of vector spaces V, W, we know that $V \otimes W, W \otimes V$ are canonically isomorphic via the flip $\tau : v \otimes w \mapsto w \otimes v$. In general, as long as the bialgebra A is *cocommutative*, i.e. the following diagram commutes (cf. the corresponding diagram for commutative multiplication μ)

$$\begin{array}{ccc} A & \xrightarrow{\Delta} & A \otimes A \\ & \searrow^{\Delta} & \downarrow \tau \\ & & A \otimes A \end{array},$$

then $V \otimes W, W \otimes V$ are canonically isomorphic *as A-modules* via the flip. We will see later that, while counter-intuitive at first glance (as we are forgoing a rather natural condition), this is in fact *undesirable* for producing (non-trivial) knot invariants, because the flip satisfies $\tau^2 = \text{id}$. One of the crucial ideas, however, is that we will still want to have (natural) isomorphisms between $V \otimes W, W \otimes V$ for any A-modules V, W, just not via the flip. In fact, this is one of the motivating reasons for the introduction of quantum groups - to produce (interesting) non-commutative, non-cocommutative bialgebras.

Working with Δ may in general be slightly tricky in practice because $\Delta(a)$ is an element of $A \otimes A$. To ease things, we fix here the following notation, which is attributed to Sweedler:

Definition 4.5. (Sweedler's notation) For A a bialgebra and $a \in A$, we denote $\Delta(a) = \sum_{(a)} a' \otimes a''$. Where there is no ambiguity, we will usually drop the subscript $_{(a)}$.

We would like to go one step further and put, for any A-module M, an A-module structure on M^* (its dual as vector spaces). We can consider, for a linear endomorphism S on A, an action of $a \in A$ on M^* by sending $f \mapsto (m \mapsto f(S(a)m))$. In order for this to be an A-action (satisfy associativity) we need for S to be an anti-homomorphism, that is, $S(a_1 a_2) = S(a_2) S(a_1)$ for $a_1, a_2 \in A$.

However, this alone is not sufficient. As a minimal first step, with the introduction of M^* as an A-module, we would also like the simplest canonical maps involving M^*, that is, the *evaluation* and the *coevaluation*, to also be A-linear.

In order for the evaluation $\text{ev}_M : M^* \otimes M \to k$ to be A-linear, we require for all $a \in A$

$$\text{ev}_M(a(f \otimes m)) = \sum \text{ev}_M(a' f \otimes a'' m) = \sum f(S(a') a'' m) = f\Big(\big(\sum S(a') a''\big) m\Big)$$
$$= a\text{ev}_M(f \otimes m) = \epsilon(a) f(m) = f(\eta \circ \epsilon(a) m)$$

This suggests we need $\sum S(a') a'' = \eta \circ \epsilon(a)$, or that the following diagram commutes:

$$\begin{array}{ccc} A & \xrightarrow{\Delta} & A \otimes A \\ \downarrow{\scriptstyle \eta\circ\epsilon} & & \downarrow{\scriptstyle S\otimes \text{id}} \\ A & \xleftarrow[\mu]{} & A \otimes A \end{array}$$

As for the coevaluation $\delta_M : k \to M \otimes M^*$ to be A-linear (of course assuming M finite dimensional as vector space), the computation is more intricate, but we require for all $a \in A$

$$\begin{aligned}
\delta_M(a(1)) &= \epsilon(a) \sum_i m_i \otimes m_i^* = \sum_i \eta \circ \epsilon(a) m_i \otimes m_i^* \\
= a\delta_M(1) &= \sum_i \sum_{(a)} a' m_i \otimes a'' m_i^* \\
&= \sum_i \sum_{(a)} a' m_i \otimes \big(\sum_j m_i^*(S(a'')m_j) m_j^* \big) \\
&= \sum_i \sum_{(a)} \sum_j a' m_i^*(S(a'')m_j) m_i \otimes m_j^* \\
&= \sum_{(a)} \sum_j a' \sum_i (m_i^*(S(a'')m_j) m_i) \otimes m_j^* \\
&= \sum_{(a)} \sum_j a' S(a'') m_j \otimes m_j^*
\end{aligned}$$

and so we need $\sum a' S(a'') = \eta \circ \epsilon(a)$, or that the following diagram commutes:

$$\begin{array}{ccc}
A & \xrightarrow{\Delta} & A \otimes A \\
\Big\downarrow{\scriptstyle \eta\circ\epsilon} & & \Big\downarrow{\scriptstyle \mathrm{id}\otimes S} \\
A & \xleftarrow[\mu]{} & A \otimes A
\end{array}$$

Definition 4.6. A *Hopf algebra* $(A, \mu, \eta, \Delta, \epsilon, S)$ over k is a bialgebra over k equipped with a linear endomorphism S on A which is also an anti-homomorphism, so that the following diagrams commute:

$$\begin{array}{ccc}
A & \xrightarrow{\Delta} & A \otimes A \\
\Big\downarrow{\scriptstyle \eta\circ\epsilon} & & \Big\downarrow{\scriptstyle S\otimes \mathrm{id}} \\
A & \xleftarrow[\mu]{} & A \otimes A
\end{array}
\qquad\qquad
\begin{array}{ccc}
A & \xrightarrow{\Delta} & A \otimes A \\
\Big\downarrow{\scriptstyle \eta\circ\epsilon} & & \Big\downarrow{\scriptstyle \mathrm{id}\otimes S} \\
A & \xleftarrow[\mu]{} & A \otimes A
\end{array}$$

S is called the *antipode* of the Hopf algebra A.

Lemma 4.7. *Given a Hopf algebra A and a (finite-dimensional) A-module M, the evaluation and coevaluation maps $ev_M : M^* \otimes M \to k$ and $\delta_M : k \to M \otimes M^*$ are each A-linear.*

Proof. As above. ∎

4.2. The classical case of $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$.

Fix the ground field $k = \mathbb{C}$, its key properties being that it has characteristic zero and is algebraically closed. We focus first on the $\mathfrak{sl}_2$ case as it is a 'building block' for subsequent general (semisimple) Lie algebras and in fact certain key results in the general case will come to rely on the corresponding result for $\mathfrak{sl}_2$.

4.2.1. Recall the Lie algebra $\mathfrak{sl}_2(\mathbb{C})$ has basis $e = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $f = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $h = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and its universal enveloping algebra $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$ is the algebra over $\mathbb{C}$ generated by E, F, H modulo relations

$$\begin{aligned} EF - FE &= H \\ EH - HE &= -2E \iff EH = (H-2)E \\ FH - HF &= 2F \iff FH = (H+2)F \end{aligned}$$

Remark 4.8. There is a standard Hopf algebra structure on $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$ (or more generally, $\mathbf{U}(\mathfrak{g})$ for any Lie algebra $\mathfrak{g}$), given by $\Delta(x) = 1 \otimes x + x \otimes 1$, $\epsilon(x) = 0$ and $S(x) = -x$ for all $x \in \mathfrak{g}$. The issue is, cf. Remark 4.4, that Δ is cocommutative.

We would like to consider (finite-dimensional) representations of $\mathfrak{sl}_2(\mathbb{C})$, or equivalently modules M over the algebra $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$ which are finite-dimensional as $\mathbb{C}$-vector spaces. Equivalently, we specify how E, F, H act on M as linear endomorphisms, satisfying the above relations.

We recall first some standard results from the representation theory of $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$, to set the fuller picture.

Definition 4.9. (Weights, weight spaces) A (non-zero) eigenvector $v \in M$ of H with eigenvalue λ is said to be of *weight* λ, and the eigenspace of H corresponding to eigenvalue λ is called a *weight space* of M (of weight λ) and denoted M_λ.

Lemma 4.10. *For all λ, we have $EM_\lambda \subset M_{\lambda+2}$ and $FM_\lambda \subset M_{\lambda-2}$.*

Proof. Straightforward from the latter two of the three defining relations satisfied by E, F, H. □

Lemma 4.10 motivates the following definition and lemma:

Definition 4.11. (Highest weight) If $v \in M$ is of weight λ and $Ev = 0$, then we say in addition that v is of *highest weight* λ.

Lemma 4.12. *There is a highest weight vector in every finite-dimensional module M.*

Proof. The finite-dimensionality of M guarantees that it has only finitely many weights, whence the result follows from Lemma 4.10. □

If M is finite-dimensional irreducible, therefore, it must be generated (as $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$-module) by a highest weight vector $v \in M$.

Lemma 4.13. *Let $v \in M$ be of highest weight λ. For $i \geq 0$, denote $v_i = \frac{1}{i!}F^i v$ (in particular $v_0 = v$). Then we have $Hv_i = (\lambda - 2i)v_i$, $Ev_i = (\lambda - i + 1)v_{i-1}$ and $Fv_i = (i+1)v_{i+1}$.*

Proof. Straightforward computations from the defining relations of E, F, H. □

Lemma 4.14. *Let $v \in M$ be of highest weight λ, suppose v generates M and M is finite-dimensional. Then $\lambda = \dim(M) - 1$ (in particular, λ is a non-negative integer), and M has weight spaces $M_\lambda, M_{\lambda-2}, \cdots, M_{-\lambda}$ each of dimension 1.*

Proof. Keeping notations of Lemma 4.13, since the v_i have distinct weights, there must exist some j with $v_j \neq 0$ and $v_{j+1} = 0$. Then $0 = Ev_{j+1} = (\lambda - j)v_j$ gives $\lambda = j$. Now $v_0, \cdots, v_j$ are linearly independent (since they have distinct weights) and it is clear from Lemma 4.13 that they span M, so they form a basis for M, from which everything follows. □

Lemma 4.15. *Keeping assumptions and notations of Lemma 4.14, we have also that M is irreducible.*

Proof. Suppose M' is a proper submodule of M. Then it has a highest weight vector v' which must also be a highest weight vector for M, and must therefore be a scalar multiple of v, but then this forces $M' = M$. □

The point here is that the finite-dimensional irreducible modules over $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$ are completely parameterised by the non-negative integer λ.

4.3. The quantum case of $\mathbf{U}_q(\mathfrak{sl}_2(\mathbb{C}))$.

Again, fix the ground field $\mathbb{C}$ and an indeterminate $q \in \mathbb{C}$. Here we will also assume q is not a root of unity, for reasons which will soon become clear. We often also view the situation as being over the ground field $\mathbb{C}(q)$.

The algebra $\mathbf{U}_q(\mathfrak{sl}_2(\mathbb{C}))$ is now defined as the algebra generated by E, F, K, K^{-1}, modulo relations

$$KK^{-1} = K^{-1}K = 1$$
$$EF - FE = \frac{K - K^{-1}}{q - q^{-1}}$$
$$EK = q^{-2}KE$$
$$FK = q^2 KF$$

Throughout this subsection we will denote $U = \mathbf{U}_q(\mathfrak{sl}_2(\mathbb{C}))$.

Lemma 4.16. *U may be ($\mathbb{Z}$-)graded by assigning $\deg(E) = 1$, $\deg(F) = -1$ and $\deg(K) = \deg(K^{-1}) = 0$. If $u \in U$ is homogeneous of degree i, then $KuK^{-1} = q^{2i}u$.*

Proof. For the first statement, one simply has to verify that the defining relations are all degree-homogeneous. The second statement follows directly from the first, third and fourth relations. □

Lemma 4.17. *There is a non-cocommutative Hopf algebra structure on U defined by*

$$\Delta(E) = E \otimes 1 + K \otimes E, \quad \Delta(F) = F \otimes K^{-1} + 1 \otimes F, \quad \Delta(K) = K \otimes K$$
$$\epsilon(E) = \epsilon(F) = 0, \quad \epsilon(K) = 1$$
$$S(E) = -K^{-1}E, \quad S(F) = -FK, \quad S(K) = K^{-1}.$$

Proof. To check well-definedness, one simply has to verify that the images of each of the generators (E, F, K, K^{-1}) satisfy the same four defining relations (bearing in mind that S is an anti-homomorphism). Checking coassociativity of Δ and that S is an antipode essentially need also be done only on the generators. These are all straightforward computations. □

Lemma 4.18. *We have $S^2(u) = K^{-1}uK$ for all $u \in U$.*

Proof. Again, we need only check this on the generators. □

As in the classical case, we would again like to consider (finite-dimensional) representations of U, or equivalently modules M over the algebra U which are finite-dimensional as vector spaces. Equivalently, we specify how E, F, K act on M as linear endomorphisms, satisfying the above relations. The key point is that everything proceeds more or less as in the classical case.

Definition 4.19. (Weights, weight spaces) A (non-zero) eigenvector $v \in M$ of K with eigenvalue λ is said to be of *weight* λ, and the eigenspace of K corresponding to eigenvalue λ is called a *weight space* of M (of weight λ) and denoted M_λ.

Lemma 4.20. *For all λ, we have $EM_\lambda \subset M_{q^2\lambda}$ and $FM_\lambda \subset M_{q^{-2}\lambda}$.*

Proof. Straightforward from the latter two of the four defining relations satisfied by E, F, K. □

Lemma 4.20 motivates the following definition and lemma:

Definition 4.21. (Highest weight) If $v \in M$ is of weight λ and $Ev = 0$, then we say in addition that v is of *highest weight* λ.

Lemma 4.22. *There is a highest weight vector in every finite-dimensional U-module M.*

Proof. The finite-dimensionality of M guarantees that it has only finitely many weights, whence the result follows from Lemma 4.20. (Recall q is not a root of unity!) □

Again, if M is finite-dimensional irreducible, it must be generated (as $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$-module) by a highest weight vector $v \in M$.

In fact, the key idea of Lemma 4.20, *together with the invertibility of K*, allows us to deduce something more.

Lemma 4.23. *E, F act nilpotently on every finite-dimensional U-module M.*

Proof. We handle E; the result for F follows similarly. There is nothing stopping us from 'reversing' the roles of E and K in the idea of Lemma 4.20; given an eigenvector v of E with eigenvalue μ, say, we have that Kv is an eigenvector of E with eigenvalue $q^{-2}\mu$. Observe that *the invertibility of* K guarantees that $Kv \neq 0$ (whenever $v \neq 0$). So if $\mu \neq 0$, together with q not a root of unity, we obtain infinitely many distinct eigenvalues for E, contradicting finite-dimensionality. It follows that the only eigenvalue of E is 0, i.e. E acts nilpotently as desired. □

Comparison with the classical case motivates the introduction of the so-called *quantum integers*, which correspond in a sense to the usual integers of the classical case. Their notation is standard in the literature; we state them here again for completeness:

Definition 4.24. For each $a \in \mathbb{Z}$, denote

$$[a] := \frac{q^a - q^{-a}}{q - q^{-1}}$$

and for $n \in \mathbb{Z}$ with $n \geq 0$ denote

$$[n]! := [n][n-1]\cdots[2][1].$$

For all $a \geq n \geq 0$ we have the so-called *Gaussian binomial coefficients* defined by

$$\begin{bmatrix} a \\ n \end{bmatrix} := \frac{[a]!}{[n]![a-n]!}$$

and for $a < 0$

$$\begin{bmatrix} a \\ n \end{bmatrix} := \frac{[a][a-1]\cdots[a-n+1]}{[n]!}$$

We denote also for all $a \in \mathbb{Z}$

$$[K; a] := \frac{Kq^a - K^{-1}q^{-a}}{q - q^{-1}}$$

and in general

$$[\lambda; a] := \frac{\lambda q^a - \lambda^{-1}q^{-a}}{q - q^{-1}}$$

The similarity to the classical case is apparent in the following sequence of results, which we had earlier obtained for the classical case.

Lemma 4.25. *Let $v \in M$ be of highest weight λ. For $i \geq 0$, denote $v_i = \frac{1}{[i]!}F^i v$ (in particular $v_0 = v$). Then we have $Kv_i = (\lambda q^{-2i})v_i$, $Ev_i = [\lambda; -(p-1)]v_{i-1}$ and $Fv_i = [i+1]v_{i+1}$.*

Proof. Straightforward computations from the defining relations, very similarly to Lemma 4.13. □

Lemma 4.26. *Let $v \in M$ be of highest weight λ, suppose v generates M and M is finite-dimensional. Then $\lambda = \pm q^{\dim(M)-1}$ (in particular, λ is up to a sign a non-negative integer power of q), and M has weight spaces $M_\lambda, M_{\lambda q^{-2}}, \cdots, M_{\lambda^{-1}}$ each of dimension 1.*

Proof. Exactly as in Lemma 4.14. □

Lemma 4.27. *Keeping assumptions and notations of Lemma 4.26, we have also that M is irreducible.*

Proof. Exactly as in Lemma 4.15. □

The point here is that the finite-dimensional irreducible modules over U are completely parameterised by the non-negative integers and a choice of sign.

The nilpotency of E, F (Lemma 4.23) in fact allows us to conclude via a direct computational proof that the linear operator K satisfies a polynomial which splits into distinct linear factors, i.e. the minimal polynomial of K splits into distinct linear factors, i.e. that M is the direct sum of its weight spaces.

We first require a standard lemma on the commutation relation between powers of E and F, which appears in e.g. [Lu].

Lemma 4.28. *For all $a, b \in \mathbb{Z}$, $a, b \geq 0$ we have*

$$E^a F^b = \sum_{j=0}^{\min(a,b)} \begin{bmatrix} a \\ j \end{bmatrix} \begin{bmatrix} b \\ j \end{bmatrix} [j]! F^{b-j} \big(\prod_{i=1}^{j} [K; j-a-b+i] \big) E^{a-j}$$

Proof. The proof is by induction on a, b and direct computation; since it is rather lengthy, we omit it here. □

Proposition 4.29. *Every finite-dimensional U-module M is the direct sum of its weight spaces, and its weights are all of the form $\pm q^a$ for $a \in \mathbb{Z}$.*

Proof. By Lemma 4.23, F acts nilpotently on M, say F^b is the zero operator on M for some $b \in \mathbb{Z}$, $b \geq 0$. Now using Lemma 4.28, we have

$$\begin{aligned} EF^b &= \sum_{j=0}^{1} \begin{bmatrix} 1 \\ j \end{bmatrix} \begin{bmatrix} b \\ j \end{bmatrix} [j]! F^{b-j} \big(\prod_{i=1}^{j} [K; j-1-b+i] \big) E^{1-j} \\ &= \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} [0]! F^b E + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} b \\ 1 \end{bmatrix} [1]! F^{b-1} \big([K; 1-b] \big) \end{aligned}$$

and this implies that $F^{b-1}\big([K; 1-b]\big)$ is zero on M. Again using Lemma 4.28, we have

$$\begin{aligned} E^2 F^b &= \sum_{j=0}^{2} \begin{bmatrix} 2 \\ j \end{bmatrix} \begin{bmatrix} b \\ j \end{bmatrix} [j]! F^{b-j} \big(\prod_{i=1}^{j} [K; j-2-b+i] \big) E^{2-j} \\ &= \begin{bmatrix} 2 \\ 0 \end{bmatrix} \begin{bmatrix} b \\ 0 \end{bmatrix} [0]! F^b E^2 + \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} b \\ 1 \end{bmatrix} [1]! F^{b-1} \big([K; -b] \big) E \\ &\quad + \begin{bmatrix} 2 \\ 2 \end{bmatrix} \begin{bmatrix} b \\ 2 \end{bmatrix} [2]! F^{b-2} \big([K; 1-b][K; 2-b] \big) \end{aligned}$$

Now multiplying both sides on the right by $[K; 3-b]$, and using the easily verified fact that $E[K; 3-b] = [K; 1-b]E$ (this holds in general with $3-b$ replaced with any integer), we see that

$$F^{b-2}\big([K; 1-b][K; 2-b][K; 3-b]\big)$$

is zero on M.

It is clear that we can proceed inductively in this manner; in general, we obtain for all $0 \leq i \leq b$ that

$$F^{b-i}\big([K; 1-b][K; 2-b][K; 3-b] \cdots [K; (2i-1)-b]\big)$$

is zero on M.

In particular, at the end we obtain (for $i = b$) that

$$[K; 1-b][K; 2-b][K; 3-b] \cdots [K; b-1]$$

is zero on M; some straightforward manipulations then give the desired, which is that K satisfies a polynomial which splits into distinct linear factors of the form $(x \pm q^a)$ for $a \in \mathbb{Z}$. □

Remark 4.30. In the classical case, a theorem of Weyl allows us to conclude (from the semisimplicity of $\mathfrak{sl}_2(\mathbb{C})$) that in fact every finite-dimensional $\mathbf{U}(\mathfrak{sl}_2(\mathbb{C}))$-module is *semisimple*; i.e. it is a direct sum of irreducible modules. Minor modifications of the proof of this theorem allow us to show this also in the quantum case; that is, every finite-dimensional $U = \mathbf{U}_q(\mathfrak{sl}_2(\mathbb{C}))$-module is a direct sum of irreducible modules. Since the proof of this requires a lengthy discussion (e.g. requires the use of the Casimir element), and we do not need the full strength of this result subsequently, we do not discuss this result here.

In comparison with the classical case, the obvious difference is the presence of signs in the weights. Intuitively, we do not lose any information by specifying that all weights have positive sign, since all that matters is a choice of sign. Formally, there is an equivalence between the category of U-modules with all weights positive, and the category of U-modules with all weights negative (although we do not show this here). Therefore, we usually consider only those U-modules with all weights positive, and henceforth we will do just this. These are also known as type **1** modules.

Definition 4.31. We denote by $\mathscr{U}$ the category of finite-dimensional U-modules (of type **1**), with morphisms as U-module homomorphisms.

Lemma 4.32. *($\mathscr{U}$ is a tensor, or monoidal category) $\mathscr{U}$ is a tensor category with the obvious tensor operation of taking tensor products of modules.*

Proof. From $\Delta(K) = K \otimes K$, the tensor product of two weight spaces of positive weight is a (subspace of a) weight space of the product of the two weights, which is again positive. The associativity comes from the associativity of the tensor product of modules (which in turn comes from coassociativity of the Hopf algebra U). The unit object is the trivial one-dimensional module $\mathbb{C}$ with weight $q^0 = 1$; equivalently, we can see this as being the action induced by the counit ϵ of U on the ground field $\mathbb{C}$. □

4.4. **Non-cocommutativity of U.**

Recall in Remark 4.4 that one of the central ideas behind the introduction of $U = \mathbf{U}_q(\mathfrak{sl}_2(\mathbb{C}))$ is its non-cocommutativity as a Hopf algebra. In doing so we lose the natural "flip" isomorphism between $V \otimes W$ and $W \otimes V$. However, we should still expect to have natural isomorphisms between $V \otimes W$ and $W \otimes V$ for any U-modules V, W. In this subsection we do just that, following the approach of Jantzen in [Ja]:

Definition 4.33. We use τ to denote any 'flip" morphism. More precisely, we denote by $\tau : U \otimes U \to U \otimes U$ the 'flip' homomorphism sending $u' \otimes u''$ to $u'' \otimes u'$, and for any two U-modules V, W we also denote by $\tau : V \otimes W \to W \otimes V$ the 'flip' homomorphism sending $v \otimes w$ to $w \otimes v$. It will always be clear from the context which of the two we mean.

Definition 4.34. For each integer $n \geq 0$ denote

$$a_n = (-1)^n q^{-n(n-1)/2} \frac{(q - q^{-1})^n}{[n]!}$$

and

$$\Theta_n = a_n F^n \otimes E^n.$$

Given any finite-dimensional U-modules V, W recall that E, F each act nilpotently on each of V, W (Lemma 4.23). We therefore have a well-defined (linear) map $\Theta : V \otimes W \to V \otimes W$ defined by

$$\Theta = \sum_{n \geq 0} \Theta_n.$$

When there are multiple U-modules involved we denote this map by $\Theta_{V,W}$.

Lemma 4.35. Θ *is bijective.*

Proof. $F \otimes E$ acts nilpotently on $V \otimes W$, so that Θ is in fact unipotent ($\Theta_0 = 1 \otimes 1$) and therefore bijective. □

Definition 4.36. Given any finite-dimensional U-modules V, W recall that V, W are each the direct sum of their weight spaces, and recall that we assume that all weights are of the form $+q^a$ ($a \in \mathbb{Z}$). Define a function f' by $f'(q^a, q^b) = (q^{1/2})^{-ab}$ for all $a, b \in \mathbb{Z}$ (for ease of understanding, we may denote this scalar function also by f when there is no risk of confusion), and define a (linear) map $f : V \otimes W \to V \otimes W$ by $f(v \otimes w) = f'(\lambda, \mu) v \otimes w$ whenever $v \in V_\lambda$ and $w \in W_\mu$. It is clear that f is bijective.

Lemma 4.37. *Given any finite-dimensional U-modules V, W, we have for all $u \in U$ that*

$$\Delta(u) \circ \Theta \circ f = \Theta \circ f \circ (\tau \circ \Delta)(u)$$

as maps from $V \otimes W$ to itself.

Proof. It suffices to verify this for u the generators E, F, K and that the maps agree on each weight space of $V \otimes W$, and the rest is purely computational. □

Proposition 4.38. *Given any finite-dimensional U-modules V, W, the map $R : V \otimes W \to W \otimes V$ defined by*

$$R = \Theta \circ f \circ \tau$$

is a U-module isomorphism. When there are multiple U-modules involved we denote this map by $R_{V,W}$.

Proof. Θ, f, τ are each bijective, so R is bijective also. Now for all $u \in U$ and $x \in V \otimes W$ we have

$$\begin{aligned} R(u \cdot x) &= \Theta \circ f \circ \tau \circ \Delta(u)(x) \\ &= \Theta \circ f \circ (\tau \circ \Delta)(u)(\tau(x)) \\ &= \Delta(u) \circ \Theta \circ f(\tau(x)) \quad \text{(by Lemma 4.37)} \\ &= \Delta(u) \circ R(x) \\ &= u \cdot R(x) \end{aligned}$$

as desired. □

Proposition 4.39. *(Naturality of R) The isomorphism R is a natural isomorphism in the following sense: given any finite-dimensional U-modules V, W, V', W', and any U-module homomorphisms $g : V \to V'$, $h : W \to W'$, the following diagram commutes:*

$$\begin{array}{ccc} V \otimes W & \xrightarrow{R_{V,W}} & W \otimes V \\ \downarrow{\scriptstyle g \otimes h} & & \downarrow{\scriptstyle h \otimes g} \\ V' \otimes W' & \xrightarrow{R_{V',W'}} & W' \otimes V' \end{array}$$

In other words, it is a natural isomorphism between the functors $\otimes$ and $\otimes\tau$ on the category of finite-dimensional U-modules.

Proof. This is to be expected since Θ, f, τ are each defined 'independently' of any U-module. To be more precise: $g \otimes h$ is $(U \otimes U)$-linear, so commutes with $\Theta \in U \otimes U$; $g \otimes h$ preserves weight spaces, so commutes with f which acts as a scalar on each weight space; $g \otimes h$ clearly commutes with the flip τ. □

For *any* two (finite-dimensional) U-modules $V, W \in \mathscr{U}$, therefore, we now have a *natural* isomorphism $R_{V,W} : V \otimes W \to W \otimes V$. While we could discuss further here the important properties it is constructed to possess, we will defer this discussion and return to it in Section 5.5, where greater context can be given to its properties.

4.5. A word on general semisimple $\mathfrak{g}$.

Everything we have done in this section can be generalised to arbitrary (complex) semisimple Lie algebras $\mathfrak{g}$, with the additional difficulty being only technical rather than conceptual. This is a consequence of Serre's theorem for complex semisimple Lie algebras: given such a Lie algebra $\mathfrak{g}$, consider a root system Φ with base Π. Then there is a presentation of $\mathfrak{g}$ (also $\mathbf{U}(\mathfrak{g})$) by $3|\Pi|$ generators $x_\alpha, y_\alpha, h_\alpha$ for $\alpha \in \Pi$, modulo certain relations. For each $\alpha \in \Pi$, $x_\alpha, y_\alpha, h_\alpha$ generate a 'copy' of $\mathbf{U}(\mathfrak{sl}_2)$ in $\mathbf{U}(\mathfrak{g})$, and this forms the basis for many of the results concerning $\mathbf{U}(\mathfrak{g})$. The quantum case is then obtained similarly; $\mathbf{U}_q(\mathfrak{g})$ is generated by generators $E_\alpha, F_\alpha, K_\alpha, K_\alpha^{-1}$ for all $\alpha \in \Pi$, modulo certain relations derived from those of $\mathbf{U}(\mathfrak{g})$, and for each $\alpha \in \Pi$, $E_\alpha, F_\alpha, K_\alpha, K_\alpha^{-1}$ generate a 'copy' of $\mathbf{U}_q(\mathfrak{sl}_2)$ in $\mathbf{U}_q(\mathfrak{g})$.

5. Reshitikhin-Turaev invariants

We return now to the setting of knots and tangles. As mentioned, our aim is to produce knot invariants in a more motivated manner, by starting with the three Reidemeister moves and formulating them as algebraic relations that are to be satisfied. To this end, we would like to view tangle diagrams as algebraic objects.

5.1. The category of tangles $\mathscr{T}$.

5.1.1. How can a tangle be viewed as an algebraic object? The 'content', or 'information' expressed in a tangle consists of its (fixed) endpoints, as well as the way in which its components are *tangled* between the endpoints. We may therefore view the tangle as encoding a relation (modulo isotopy) between its endpoints. In other words, we may think of the endpoints of each tangle as being *objects* in some category, and of the tangle itself as some *morphism* in this category.

Since each morphism involves a source and a target, the natural choice is to view tangles as *boxed* tangles (Definition 2.4). In this section, therefore, all tangles will be taken to be boxed tangles.

Definition 5.1. (Category of tangles $\mathscr{T}$) The *category of tangles* $\mathscr{T}$ is defined to be the category with:

- objects as all (finite) sequences of $\uparrow$ and $\downarrow$ symbols, corresponding to the endpoints on either the top or bottom boundary of a boxed oriented tangle diagram (modulo isotopy, i.e. free movement of endpoints).
- morphisms between two objects as tangle diagrams (modulo isotopy, or equivalence) with bottom endpoints corresponding to the source object and top endpoints corresponding to the target object.
- composition of morphisms is defined in the obvious way by vertical stacking of tangle diagrams;
- the identity morphism for each object (set of endpoints) is the 'trivial' tangle diagram with no crossings.

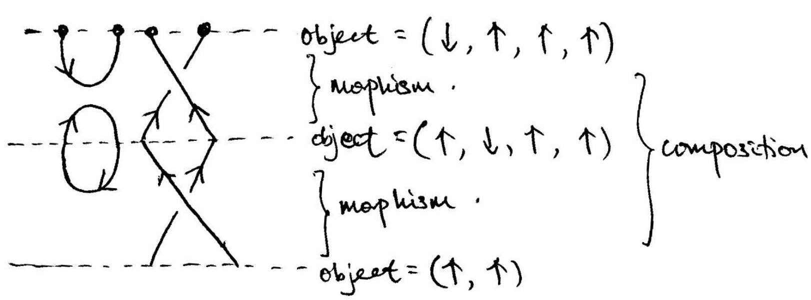


Figure 28. Example of objects, morphisms and composition in $\mathscr{T}$.

Composition in $\mathscr{T}$ is defined by *vertical* composition. One may then naturally ask if we can similarly interpret *horizontal* composition in $\mathscr{T}$. In fact this makes $\mathscr{T}$ naturally a *tensor* (also called *monoidal*) category:

Lemma 5.2. *$\mathscr{T}$ is a tensor (or monoidal) category with the tensor operation $\otimes$: $\mathscr{T} \times \mathscr{T} \to \mathscr{T}$ given by the obvious horizontal stacking for both objects and morphisms.*

Proof. The associativity of $\otimes$ is obvious; the unit object is the empty object with no endpoints. □

5.1.2. The category $\mathscr{T}$ is now defined but is still unmanageable. We would like to again work with local parts of tangle diagrams which we know must take a certain form. The advantage of formulating $\mathscr{T}$ as a category is it allows us to break tangle diagrams down into small, *elementary* pieces via composition and the tensor operation.

Proposition 5.3. *Every tangle diagram, or morphism in $\mathscr{T}$, may be expressed via finite compositions and tensor products of the so-called elementary tangle diagrams (morphisms), which are listed and denoted as follows (here we show only one of the four possible choices of orientation for the crossings $c_{-,-}$):*

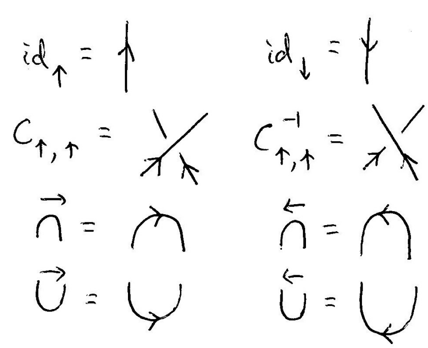


Figure 29. Elementary tangles.

Proof. The idea is intuitively straightforward and runs as follows: given any tangle diagram, first modify the tangle diagram by suitable isotopy (equivalence of diagrams) so that no two crossings in the diagram have the same y-coordinate. Then the tangle diagram is a composition of some number of smaller tangle diagrams each with at most one crossing. It is easy to see then that each of these tangle diagrams with at most one crossing must be a tensor product of elementary tangle diagrams. □

Of course, $\mathscr{T}$ is certainly not *freely* generated by the elementary tangles. One can see this already from the choice of notation for the over- and under-crossings in Figure 29, to represent that they are inverses of each other (under composition in $\mathscr{T}$). This corresponds to the second Reidemeister move (R2).

It is clear that the minimal set of relations satisfied by the elementary tangles to generate $\mathscr{T}$ should come precisely from isotopy/equivalence of diagrams and the Reidemeister moves. Moreover, since the crossings for each choice of orientation each give rise to one generator, we need also to ensure that the crossings for the various orientations are compatible in some sense. A result of Turaev makes this precise:

Proposition 5.4. *$\mathscr{T}$ is generated by the elementary tangles modulo the following relations (and minor modifications of the relations corresponding to change in orientation):*

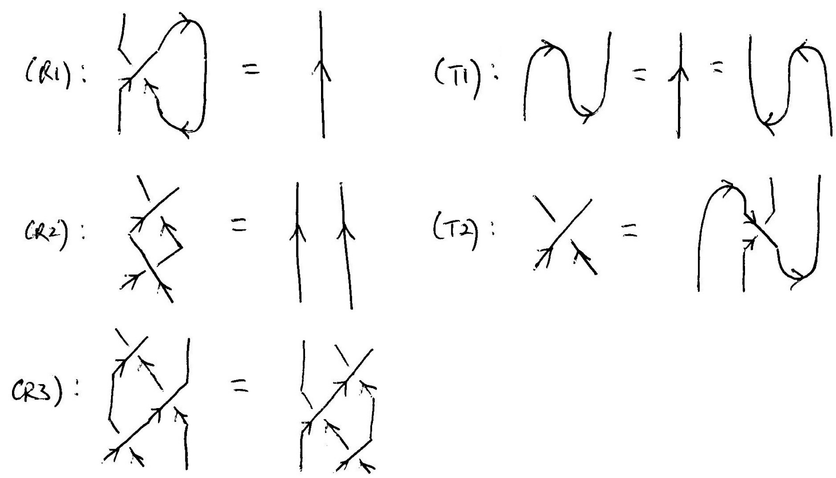


FIGURE 30. Relations satisfied by elementary tangles (up to a choice of orientation).

Proof. (R1), (R2), (R3) correspond to the usual Reidemeister moves, written to make the elementary tangles explicit. (T1) accounts for equivalence of diagrams (which does not involve crossings), while (T2) accounts for the compatibility of different orientations of crossings. Repeated application of (T2) will ensure compatibility for all four choices of orientation. ☐

5.2. **Functor between categories.**

Thus far we have two tensor categories: $\mathscr{T}$, the category of tangles, and $\mathscr{U}$, the category of (finite-dimensional) U-modules (of type $\mathbf{1}$). Since a tangle invariant is, roughly speaking, a 'function' on tangle diagrams (modulo equivalence), the natural thing to do is to consider a functor F on the category $\mathscr{T}$.

Since each object in $\mathscr{T}$ is a sequence of endpoints, it is also a tensor product of objects with one endpoint each. Therefore, for the objects, it is enough to specify a U-module, say V, for which $F(\uparrow) = V$ (we will come back to $\downarrow$ later). Where there are no endpoints, i.e. in the case of the empty object, the natural choice is the trivial or unit module of $\mathscr{U}$, which we will denote by k (for the ground field), i.e. $F(\emptyset) = k$.

For the morphisms, it is then enough to specify a U-linear homomorphism in $\mathscr{U}$ for each of the elementary tangles (Figure 29), and check that the relations between elementary tangles (Figure 30) are satisfied, so that this gives a well-defined functor F. The functor F is then an invariant of tangles with 'values' in $\mathscr{U}$.

5.3. **Duality.**

5.3.1. $\uparrow$ and $\downarrow$ are in a sense *dual* to each other: they differ by only their orientation (this is especially so if we consider endpoints of the same component; the full picture will become clear soon). In $\mathscr{U}$, we also have a natural notion of duality given by the dual module (which in turn comes from the antipode of U), cf. Lemma 4.7.

Therefore, having chosen $F(\uparrow) = V$, the natural choice is to choose $F(\downarrow) = V^*$. Now Lemma 4.7 in fact gives us the morphisms that should correspond to the elementary tangles $\overleftarrow{\cap}$ and $\overleftarrow{\cup}$! That is, we specify $F(\overleftarrow{\cap}) = \mathrm{ev}_V : V^* \otimes V \to k$ and $F(\overleftarrow{\cup}) = \delta_V : k \to V \otimes V^*$.

What is the precise way in which they are *dual*? The relation that must be satisfied by $\overleftarrow{\cap}$ and $\overleftarrow{\cup}$ is given by (the second half of) (T1). We require

$$(\mathrm{id}_V \otimes \mathrm{ev}_V) \circ (\delta_V \otimes \mathrm{id}_V) = \mathrm{id}_V$$

and for (the first half of) (T1) with the downward orientation, we require

$$(\mathrm{ev}_V \otimes \mathrm{id}_{V^*}) \circ (\mathrm{id}_{V^*} \otimes \delta_V) = \mathrm{id}_{V^*}$$

Lemma 5.5. *For all (finite-dimensional) U-modules V, we have*

$$(id_V \otimes ev_V) \circ (\delta_V \otimes id_V) = id_V$$

and

$$(ev_V \otimes id_{V^*}) \circ (id_{V^*} \otimes \delta_V) = id_{V^*}.$$

Proof. Straightforward verification. □

In fact this is precisely the notion of *right duality* in a (tensor) category; for each object V there should exist morphisms ev_V, δ_V satisfying the properties of Lemma 5.5.

5.3.2. We then now need also the notion of *left duality*, specifying morphisms for the elementary tangles $F(\overrightarrow{\cap}) = \mathrm{ev}'_V : V \otimes V^* \to k$ and $F(\overrightarrow{\cup}) = \delta'_V : k \to V^* \otimes V$. In the classical case of (finite-dimensional) vector spaces, this may be resolved by appealing to the canonical isomorphism $V \to (V^*)^*$, which composes, say, to give the map $V \otimes V^* \to (V^*)^* \otimes V^* \to k$. However, in the case of $\mathscr{U}$, a computation reveals that the canonical map $V \to (V^*)^*$ is not U-linear, as we will see in the following lemma:

Lemma 5.6. *For all (finite-dimensional) U-modules V, the canonical map*

$$\begin{aligned} e : V &\to (V^*)^* \\ v &\mapsto (f \mapsto f(v)) \end{aligned}$$

is not U-linear; rather, it satisfies for all $u \in U$

$$u \cdot e(v) = e(S^2(u)v).$$

Proof. We have

$$\begin{aligned} u \cdot e(v) &= u \cdot (f \mapsto f(v)) \\ &= (f \mapsto (S(u) \cdot f)(v)) \\ &= (f \mapsto f(S(S(u))v)) \\ &= e(S^2(u)v) \end{aligned}$$

□

Recall, however, from Lemma 4.18, that while S^2 is not the identity, it is an *inner automorphism* given by conjugation by K. So

$$\begin{aligned} u \cdot e(v) &= e(K^{-1}uKv) \\ \implies u \cdot e(K^{-1}v) &= e(K^{-1}uv) \end{aligned}$$

Hence the composition eK^{-1}, which maps $f \mapsto f(K^{-1}v)$, is indeed U-linear. We therefore still have a U-linear composition $V \otimes V^* \to (V^*)^* \otimes V^* \to k$, given as in the following lemma:

Lemma 5.7. *For all (finite-dimensional) U-modules V, the map*

$$\begin{aligned} ev'_V : V \otimes V^* &\to k \\ v \otimes f &\mapsto f(K^{-1}v) \end{aligned}$$

is a U-module homomorphism.

Proof. As above. □

Similarly, since the composition eK^{-1} is bijective with inverse Ke^{-1} which is also U-linear, we have a U-linear composition $k \to V^* \otimes (V^*)^* \to V^* \otimes V$:

Lemma 5.8. *For all (finite-dimensional) U-modules V, the map*

$$\begin{aligned} \delta'_V : k &\to V^* \otimes V \\ 1 &\mapsto \sum_i v_i^* \otimes K v_i \end{aligned}$$

is a U-module homomorphism.

Proof. As above. □

We can now verify

Lemma 5.9. *For all (finite-dimensional) U-modules V, we have*

$$(ev'_V \otimes id_V) \circ (id_V \otimes \delta'_V) = id_V$$

and

$$(id_{V^*} \otimes ev'_V) \circ (\delta'_V \otimes id_{V^*}) = id_{V^*}.$$

Proof. We verify the first equality; the second is similar. We have for all $v \in V$

$$\begin{aligned}(\mathrm{ev}'_V \otimes \mathrm{id}_V) \circ (\mathrm{id}_V \otimes \delta'_V)(v) &= (\mathrm{ev}'_V \otimes \mathrm{id}_V)(\sum_i v \otimes v_i^* \otimes K v_i) \\ &= \sum_i v_i^*(K^{-1}v) K v_i \\ &= K \sum_i v_i^*(K^{-1}v) v_i \\ &= K(K^{-1}v) \\ &= v\end{aligned}$$

as desired. ☐

5.4. Coloured tangles.

The choice $F(\uparrow) = V$ 'forces', by duality, a choice $F(\downarrow) = V^*$. However, nothing has been said about the choice of V *itself.* In fact, all that is required is for the endpoints of each *component* of a tangle to carry the same U-module $\in \mathscr{U}$, with duality then indicating orientation. We may well choose separate U-modules for separate components of a tangle.

This is precisely the idea of *coloured tangles* as considered by Reshitikhin and Turaev; the 'colouring' refers to a choice of (finite-dimensional) U-module for each component of a tangle (which remains the same under isotopy). Our functor F then defines an invariant of *coloured* oriented tangles (again, the orientation comes from duality).

Henceforth we will implicitly work in this more general setting; if all components are 'coloured' with the same U-module V we recover the usual case of plain oriented tangles. (It is worth noting, while obvious, that since the cap and cup morphisms involve only one component, the preceding discussion carries through to this greater generality.)

5.5. Braiding.

5.5.1. Now we have successfully assigned morphisms to the 'cap' and 'cup' elementary tangles $\overleftarrow{\cap}, \overleftarrow{\cup}, \overrightarrow{\cap}, \overrightarrow{\cup}$. The natural next step is to consider the 'crossing' elementary tangles $c_{-,-}$.

The crossing morphism essentially captures a relation between $V \otimes W$ and $W \otimes V$ for objects (endpoints) V, W. The extremely natural choice, therefore, considering also what we have done in Section 4.4, is $F(c_{\uparrow,\uparrow}) = R_{V,W}$, and similarly for the other orientations (associated to V^* and/or W^*)! That is, to a crossing between objects (endpoints) which are associated with U-modules V, W, we associate the morphism $R_{V,W}$ which expresses the natural map between $V \otimes W$ and $W \otimes V$.

Remark 5.10. It is clear now why the usual flip morphism is not desirable, and hence why the need for non-cocommutative Hopf algebra structures (cf. Remark 4.4); we require $R^2 \neq \mathrm{id}$ in order to distinguish between the (clearly non-equivalent) tangles:

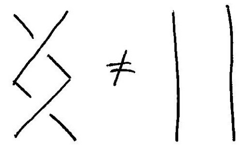


FIGURE 31. Non-equivalent tangles; we require $R^2 \neq \mathrm{id}$ to distinguish them.

5.5.2. There are now two key things that need to be handled regarding the crossings. First, while (R2) is straightforward and has already been handled (over- and under-crossings are inverses of each other), (R3) is the key relation that need to be satisfied by crossings; it is a relation between crossings of three components (objects). Secondly, (T2) expresses compatibility between duality (caps and cups) and crossings, and needs to be checked.

5.5.3. We consider first (R3), since it involves only crossings, as given in Figure 30. Suppose both the LHS and RHS of (R3) has bottom endpoints corresponding to $U, V, W \in \mathscr{U}$ in that order (from left to right); that is, they are both morphisms $U \otimes V \otimes W \to W \otimes V \otimes U$. Then (R3) can be written as

$$(R_{V,W} \otimes \mathrm{id}_U) \circ (\mathrm{id}_V \otimes R_{U,W}) \circ (R_{U,V} \otimes \mathrm{id}_W) = (\mathrm{id}_W \otimes R_{U,V}) \circ (R_{U,W} \otimes \mathrm{id}_V) \circ (\mathrm{id}_U \otimes R_{V,W})$$

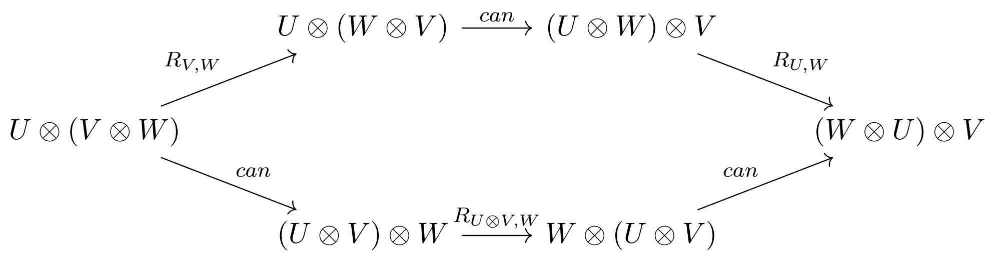
Remark 5.11. In the case $U = V = W$ (all components are coloured with the same U-module V), this relation is also known as the *quantum Yang-Baxter equation* with solution $R_{V,V}$, and is again one of the main reasons motivating the introduction of quantum groups.

Before tackling (R3), we recall first that the crossing R is meant to capture a crossing between *any* two objects V, W. In particular, one should expect that it also captures a crossing between objects which are *themselves* tensor products. This will allow us to work with expressions of the form (R3) much more naturally and easily. For example, the bottom two-thirds of the RHS of (R3) is essentially an (over-)crossing between $U \otimes V$ and W. Therefore, we should expect also a relation of the form

$$R_{U \otimes V, W} = (R_{U,W} \otimes \mathrm{id}_V) \circ (\mathrm{id}_U \otimes R_{V,W})$$

We state this as:

Proposition 5.12. *("Hexagon" identities) For any finite-dimensional* $\mathbf{U}$*-modules* U, V, W*, the following diagram commutes:*



where can denotes the usual canonical isomorphism expressing associativity of the tensor product.

Proof. Naturally the most important thing to ask is: how does $R = \Theta \circ f \circ \tau$ act on the tensor product $(U \otimes V) \otimes W$?

First of all, τ acts by sending $u \otimes v \otimes w \mapsto w \otimes u \otimes v$, which is essentially a permutation which is a composition of transpositions $(\tau \otimes \mathrm{id}) \circ (\mathrm{id} \otimes \tau)$.

Next, since K acts as $\Delta(K) = K \otimes K$ on $U \otimes V$, it follows that the weight spaces satisfy $U_\lambda \otimes V_\mu \subset (U \otimes V)_{\lambda\mu}$, so that f acts on the tensor product of weight spaces $W_\rho \otimes U_\lambda \otimes V_\mu$ as scalar multiplication by $f(\rho, \lambda\mu)$. Let us denote this map by f'.

Finally, $\Theta \in \mathbf{U} \otimes \mathbf{U}$ acts on $W \otimes (U \otimes V)$ by $(\mathrm{id} \otimes \Delta) \circ \Theta \in \mathbf{U} \otimes (\mathbf{U} \otimes \mathbf{U})$.

Let us take a closer look at the most crucial component, $(\mathrm{id} \otimes \Delta) \circ \Theta$. In particular, this involves the term $\Delta(E^n)$ for each $n \geq 0$. Since $\Delta(E) = E \otimes 1 + K \otimes E$ and Δ is a homomorphism of algebras, we have the following binomial-style identity (easily verified via induction on n):

$$\Delta(E^n) = \sum_{i=0}^{n} q^{i(n-i)} \begin{bmatrix} n \\ i \end{bmatrix} E^{n-i} K^i \otimes E^i$$

so for all $n \geq 0$ we have (keeping notations of Definition 4.34)

$$\begin{aligned} (\mathrm{id} \otimes \Delta) \circ \Theta_n &= \sum_{i=0}^{n} a_n q^{i(n-i)} \begin{bmatrix} n \\ i \end{bmatrix} F^n \otimes E^{n-i} K^i \otimes E^i \\ &= \sum_{i=0}^{n} a_i a_{n-i} (F^{n-i} \otimes E^{n-i} \otimes 1)(F^i \otimes K^i \otimes E^i) \\ &= \sum_{i=0}^{n} (\Theta_{n-i} \otimes 1)\Theta'_i \quad (\Theta'_i := a_i(F^i \otimes K^i \otimes E^i)) \end{aligned}$$

so that, summing over n, we have

$$(\mathrm{id} \otimes \Delta) \circ \Theta = (\Theta \otimes 1) \circ \Theta' \quad (\Theta' := \sum_{i \geq 0} \Theta'_i)$$

In summary, we may write $R_{U\otimes V,W}$ as a composition

$$R_{U\otimes V,W} = (\Theta\otimes 1)\circ\Theta'\circ f'\circ(\tau\otimes\mathrm{id})\circ(\mathrm{id}\otimes\tau)$$

Now we turn to the composition

$$(R_{U,W}\otimes\mathrm{id}_V)\circ(\mathrm{id}_U\otimes R_{V,W}) = (\Theta\otimes 1)\circ(f\otimes\mathrm{id})\circ(\tau\otimes\mathrm{id})\circ(1\otimes\Theta)\circ(\mathrm{id}\otimes f)\circ(\mathrm{id}\otimes\tau)$$

Comparing the two, the plan is clear: we 'push' $(\tau\otimes\mathrm{id})$ to the right and $(1\otimes\Theta)$ to the left.

To commute $(\tau\otimes\mathrm{id})$ and $(1\otimes\Theta)$, we simply define $\overline{\Theta_n} = a_nF^n\otimes 1\otimes E^n$ and $\overline{\Theta} = \sum_{n\geq 0}\overline{\Theta_n}$; then certainly $(\tau\otimes\mathrm{id})\circ(1\otimes\Theta) = \overline{\Theta}\circ(\tau\otimes\mathrm{id})$.

To commute $(\tau\otimes\mathrm{id})$ and $(\mathrm{id}\otimes f)$, we similarly define $\overline{f}$ to get $(\tau\otimes\mathrm{id})\circ(\mathrm{id}\otimes f) = \overline{f}\circ(\tau\otimes\mathrm{id})$.

Now what is left is to commute $(f\otimes\mathrm{id})$ and $\overline{\Theta}$. It suffices to consider the weight spaces; for all $w_\rho\otimes u_\lambda\otimes v_\mu\in W_\rho\otimes U_\lambda\otimes V_\mu$, we have

$$\begin{aligned}(f\otimes\mathrm{id})\circ\overline{\Theta}(w_\rho\otimes u_\lambda\otimes v_\mu) &= (f\otimes\mathrm{id})\sum_{n\geq 0}a_nF^nw_\rho\otimes u_\lambda\otimes E^nv_\mu\\ &= \sum_{n\geq 0}f(\rho q^{-2n},\lambda)a_nF^nw_\rho\otimes u_\lambda\otimes E^nv_\mu \quad \text{(cf. Lemma 4.20)}\end{aligned}$$

Now recall $f(q^a,q^b) = (q^{1/2})^{-ab}$ for all $a,b\in\mathbb{Z}$. Therefore writing $\rho = q^a, \lambda = q^b$ for some $a,b\in\mathbb{Z}$, we have $f(\rho q^{-2n},\lambda) = (q^{1/2})^{-(a-2n)b} = (q^{1/2})^{-ab}(q^b)^n = f(\rho,\lambda)\lambda^n$. Hence

$$\begin{aligned}(f\otimes\mathrm{id})\circ\overline{\Theta}(w_\rho\otimes u_\lambda\otimes v_\mu) &= \sum_{n\geq 0}f(\rho,\lambda)\lambda^na_nF^nw_\rho\otimes u_\lambda\otimes E^nv_\mu\\ &= \sum_{n\geq 0}f(\rho,\lambda)a_nF^nw_\rho\otimes K^nu_\lambda\otimes E^nv_\mu\\ &= \Theta'\circ(f\otimes\mathrm{id})(w_\rho\otimes u_\lambda\otimes v_\mu)\end{aligned}$$

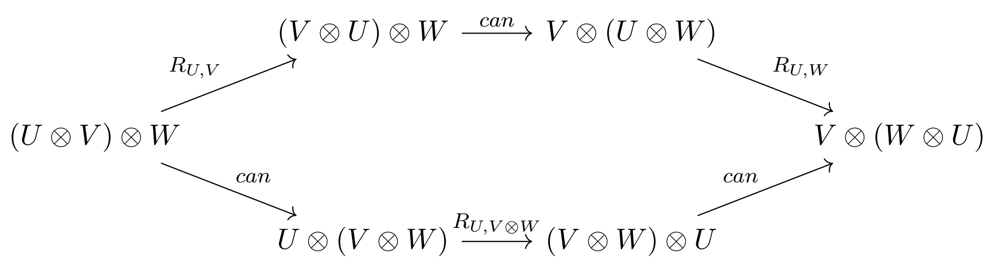
that is, $(f\otimes\mathrm{id})\circ\overline{\Theta} = \Theta'\circ(f\otimes\mathrm{id})$.

It therefore finally remains now to show that $(f\otimes\mathrm{id})\circ\overline{f} = f'$. But again the LHS acts on $w_\rho\otimes u_\lambda\otimes v_\mu\in W_\rho\otimes U_\lambda\otimes V_\mu$ by multiplication by $f(\rho,\lambda)f(\rho,\mu)$, while the RHS acts by $f(\rho,\lambda\mu)$, and it is easily verified that these two quantities are equal.

□

Of course, we have considered $R_{U\otimes V,W}$, with the tensor product in the first object; but the tensor product may equally be in the second object. Considering the bottom two-thirds of the *LHS* of (R3) gives $R_{U,V\otimes W}$, and we also have a similar relation given by:

Proposition 5.13. *("Hexagon" identities) For any finite-dimensional* $\mathbf{U}$*-modules* U, V, W*, the following diagram commutes:*



where can denotes the usual canonical isomorphism expressing associativity of the tensor product.

Proof. Extremely similar to Proposition 5.12; we will not repeat it here! □

Propositions 4.39, 5.12, 5.13 together form precisely the basis of the notion of a *braided* (tensor) category. It is this precise notion that gives rise to useful applications (e.g. the Yang-Baxter equation), such as to (R3):

Proposition 5.14. *For any finite-dimensional* $\mathbf{U}$*-modules* U, V, W*, we have the (R3) relation*

$$(R_{V,W} \otimes id_U) \circ (id_V \otimes R_{U,W}) \circ (R_{U,V} \otimes id_W) = (id_W \otimes R_{U,V}) \circ (R_{U,W} \otimes id_V) \circ (id_U \otimes R_{V,W})$$

Proof. The top two-thirds of the LHS of (R3) is itself a crossing between $V \otimes U$ and W, while the bottom two-thirds of the RHS is a crossing between $U \otimes V$ and W. By Proposition 5.12, it suffices to show

$$R_{V\otimes U,W} \circ (R_{U,V} \otimes \mathrm{id}_W) = (\mathrm{id}_W \otimes R_{U,V}) \circ R_{U\otimes V,W}.$$

But this is just the commutativity of the diagram

$$\begin{CD} (U \otimes V) \otimes W @>{R_{U\otimes V,W}}>> W \otimes (U \otimes V) \\ @V{R_{U,V}\otimes \mathrm{id}_W}VV @VV{\mathrm{id}_W \otimes R_{U,V}}V \\ (V \otimes U) \otimes W @>{R_{V\otimes U,W}}>> W \otimes (V \otimes U) \end{CD}$$

which is Proposition 4.39! □

Remark 5.15. The name *braiding* comes from the corresponding theory for *braids*, which are a special kind of tangle with no copies of S^1 (so all components are copies of $[0, 1]$) and with all strands always oriented upward. Therefore no cup or cap morphisms are required; the braids are generated by the crossings. The braids with n endpoints (n each on top and bottom) form a group under the usual composition (vertical stacking), with a presentation given by generators which are the crossings at each of the $(n - 1)$ possible positions and relations corresponding to (R2) and (R3).

This relation has remarkable similarity to the presentation of the symmetric group on n elements given by generators as transpositions; the only exception being that the square of each crossing (generator) is not the identity in the braid group. Our algebraic approach to tangles can be seen as a natural extension of the algebraic theory of braid groups.

An alternate motivation for the algebraic approach is the following, which bears more relation to the concept of categorification: just as a monoid is essentially a one-object category (and vice versa), a tensor or monoidal category may be viewed essentially as a one-object 2-category. Viewing our tensor category $\mathscr{T}$ as a one-object 2-category, then, our approach is equivalent to the use of the standard *string diagrams* for visualising 2-categories, with our objects in $\mathscr{T}$ being the 1-morphisms and our morphisms (tangles) in $\mathscr{T}$ being the 2-morphisms.

5.5.4. We turn now to (T2), which expresses the compatibility between crossings and cups/caps. *Surprisingly, the compatibility between crossings and cups/caps is a consequence of the generality of duality and braiding that we have developed!*

Proposition 5.16. *For any (finite-dimensional) U-modules V, W, we have the (T2) relation (cf. Figure 30)*

$$R_{V,W} = (ev'_V \otimes id_{W\otimes V}) \circ (id_V \otimes R^{-1}_{V^*,W} \otimes id_V) \circ (id_{V\otimes W} \otimes \delta'_V)$$

Proof. First, from Lemma 5.9, we have

$$\begin{aligned} R_{V,W} &= R_{V,W} \circ (\mathrm{id}_V \otimes \mathrm{id}_W) \\ &= (\mathrm{id}_k \otimes R_{V,W}) \circ (\mathrm{ev}'_V \otimes \mathrm{id}_{V\otimes W}) \circ (\mathrm{id}_V \otimes \delta'_V \otimes \mathrm{id}_W) \\ &= (\mathrm{ev}'_V \otimes \mathrm{id}_{W\otimes V}) \circ (\mathrm{id}_{V\otimes V^*} \otimes R_{V,W}) \circ (\mathrm{id}_V \otimes \delta'_V \otimes \mathrm{id}_W) \end{aligned}$$

Graphically, this may be visualised as in the first equality of Figure 32. The motivation is to produce the cap morphism as appears in the RHS as $(\mathrm{ev}'_V \otimes \mathrm{id}_{W\otimes V})$.

Now we would like to produce the crossing $R^{-1}_{V^*,W}$. We have

$$\begin{aligned} R_{V,W} = (\mathrm{ev}'_V \otimes \mathrm{id}_{W\otimes V}) &\circ (\mathrm{id}_V \otimes R^{-1}_{V^*,W} \circ \mathrm{id}_V) \circ (\mathrm{id}_V \otimes R_{V^*,W} \circ \mathrm{id}_V) \\ &\circ (\mathrm{id}_{V\otimes V^*} \otimes R_{V,W}) \circ (\mathrm{id}_V \otimes \delta'_V \otimes \mathrm{id}_W) \end{aligned}$$

Graphically, this is visualised in the second equality of Figure 32.

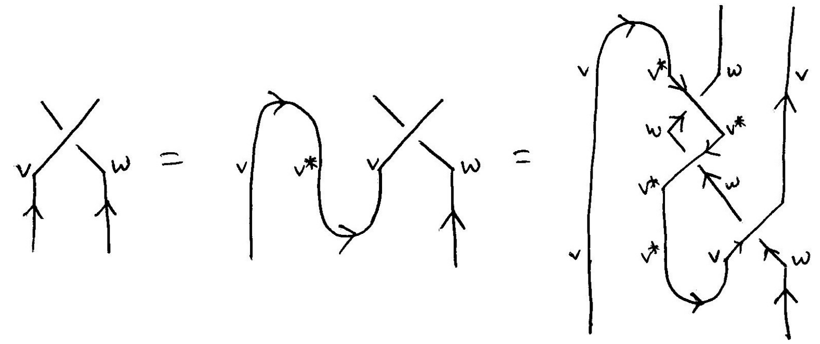


FIGURE 32. Graphical visualisation of proof.

Therefore all that remains is to show

$$(R_{V^*,W} \circ \mathrm{id}_V) \circ (\mathrm{id}_{V^*} \otimes R_{V,W}) \circ (\delta'_V \otimes \mathrm{id}_W) = (\mathrm{id}_W \otimes \delta'_V)$$

Visually, this is as in Figure 33.

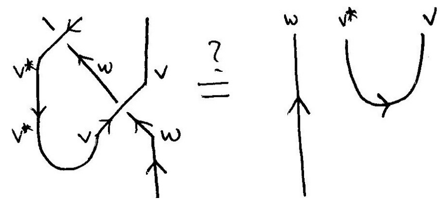


FIGURE 33. Graphical visualisation of proof.

But now the top two-thirds of Figure 33 is precisely the crossing $R_{V^* \otimes V,W} = (R_{V^*,W} \circ \mathrm{id}_V) \circ (\mathrm{id}_{V^*} \otimes R_{V,W})$ (cf. Proposition 5.12)! So all we want is

$$R_{V^* \otimes V,W} \circ (\delta'_V \otimes \mathrm{id}_W) = (\mathrm{id}_W \otimes \delta'_V)$$

which is just the commutativity of the diagram

$$\begin{array}{ccc} k \otimes W & \xrightarrow{R_{k,W} = \mathrm{id}} & W \otimes k \\ \downarrow{\scriptstyle \delta'_V \otimes \mathrm{id}_W} & & \downarrow{\scriptstyle \mathrm{id}_W \otimes \delta'_V} \\ (V^* \otimes V) \otimes W & \xrightarrow{R_{V^* \otimes V,W}} & W \otimes (V^* \otimes V) \end{array}$$

which is Proposition 4.39! □

5.6. **Deframing.**

All that remains now is (R1). Recall in Remark 2.9 that we mentioned that (R1) differs from the other moves in that it contains ‘non-trivial’ content; that is, it removes a ‘3D’ twist from the 2D diagram. In fact, we will see in the next section that our functor F does *not* in general satisfy (R1), much as the Kauffman bracket does not satisfy

(R1) (but only up to a scalar). Our functor F is therefore a well-defined invariant of (coloured) *framed* oriented tangles, but not yet of *unframed* oriented tangles. In order to obtain an invariant of unframed tangles we will have to use the approach of 'deframing' as previously used in Section 2.3.3 and for the Khovanov homology (Section 3.8.1).

5.7. **Jones polynomial.**

5.7.1. Our objective now is to show how the Jones polynomial may be recovered from this approach and the representation theory of $\mathbf{U} = \mathbf{U}_q(\mathfrak{sl}_2)$.

Definition 5.17. (cf. Lemma 4.26) In this subsection, let V denote the (unique) irreducible $\mathbf{U}$-module of dimension 2 with highest weight q; this is also known as the *standard* representation of $\mathbf{U}_q(\mathfrak{sl}_2)$. With respect to a basis $\{v_0, v_1\}$ of weights q, q^{-1} respectively, $E, F, K \in \mathbf{U}$ respectively act as

$$E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad K = \begin{pmatrix} q & 0 \\ 0 & q^{-1} \end{pmatrix}$$

Since we are only interested in uncoloured tangles (Section 5.4), we assign to each component of every tangle this standard module V.

The first order of business is certainly to compute $R_{V,V}$ with respect to the standard basis $\{v_0 \otimes v_0, v_0 \otimes v_1, v_1 \otimes v_0, v_1 \otimes v_1\}$ of $V \otimes V$. We have

Lemma 5.18. *With respect to the standard basis* $\{v_0 \otimes v_0, v_0 \otimes v_1, v_1 \otimes v_0, v_1 \otimes v_1\}$, $R_{V,V}$ *acts as*

$$R_{V,V} = \begin{pmatrix} q^{-1/2} & 0 & 0 & 0 \\ 0 & 0 & q^{1/2} & 0 \\ 0 & q^{1/2} & q^{-1/2} - q^{3/2} & 0 \\ 0 & 0 & 0 & q^{-1/2} \end{pmatrix}$$

Proof. Straightforward verification on each of the basis elements. □

Together with $\mathrm{ev}_V, \delta_V, \mathrm{ev}'_V, \delta'_V$ which have all been defined explicitly, as well as all the preceding discussion, we now obtain a well-defined invariant of *framed* tangles.

Now we can consider what happens under (R1).

Lemma 5.19. *The LHS of (R1) (over-crossing) is, as an endomorphism of V, scalar multiplication by $q^{-3/2}$. Similarly, the same diagram as in (R1) but with the over-crossing replaced by an under-crossing acts as scalar multiplication by $q^{3/2}$.*

Proof. Straightforward verification on each of the basis elements v_0, v_1. □

5.7.2. However, recall that the Jones polynomial (and Kauffman bracket) are defined only on *links*, and not tangles. Since a link is essentially a tangle with no endpoints, our functor F assigns to each link a U-linear, or essentially just a $\mathbb{C}$-linear map from $k = \mathbb{C}$ to $k = \mathbb{C}$, which is nothing but a scalar in $\mathbb{C}$ depending on the parameter q. Equivalently, we may view this as a $\mathbb{C}(q)$-linear map from $k = \mathbb{C}(q)$ to $\mathbb{C}(q)$, which is nothing but a scalar in $\mathbb{C}(q)$. The value of this scalar is then the value of the invariant that we assign to each link.

We now go through with the deframing process:

Proposition 5.20. *Given any oriented link diagram D, the quantity*

$$J'(D) := (q^{3/2})^{n_+(D)-n_-(D)}F(D)$$

is a well-defined invariant of oriented links.

5.7.3. In fact, the claim now is that this invariant J' is, up to a cosmetic replacement of q by $-q$, precisely the Jones polynomial that we have defined in Section 2!

From Proposition 2.23, we need only check two things: the value assigned to n closed loops, and that the skein relation is satisfied.

Lemma 5.21. *A single closed loop is assigned the value $(q+q^{-1})$; more generally, n closed loops are assigned the value $(q+q^{-1})^n$.*

Proof. For a single loop, it suffices to check that $\mathrm{ev}'_V \circ \delta_V$ (or $\mathrm{ev}_V \circ \delta'_V$) acts as the scalar $(q+q^{-1})$, which is straightforward. The case of multiple loops is immediate. ☐

Lemma 5.22. *We have the following 'local' relation:*

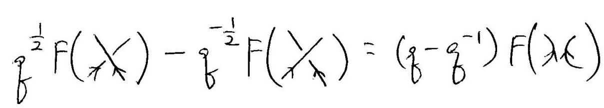


FIGURE 34. Skein relation for F.

that is, we have the equality of endomorphisms on $V \otimes V$

$$q^{1/2}R_{V,V}^{-1} - q^{-1/2}R_{V,V} = (q-q^{-1})id_{V\otimes V}$$

Then since the value on an entire link diagram is obtained by tensor products and compositions which preserve the form of the above relation, we obtain the desired skein relation

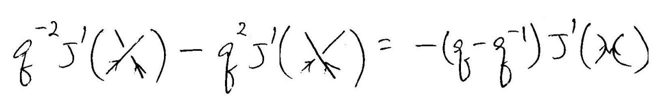


FIGURE 35. Skein relation for J'.

(cf. Proposition 2.23).

Proof. The verification of

$$q^{1/2}R_{V,V}^{-1} - q^{-1/2}R_{V,V} = (q - q^{-1})\mathrm{id}_{V\otimes V}$$

is a straightforward computation using Lemma 5.18; the second relation (for J') is proved in the same way as in Figure 13. □

Remark 5.23. We see from the preceding result that, although the $R_{V,V}$ are constructed so that $R^2 \neq \mathrm{id}$, their minimal polynomial is still quadratic, and gives rise precisely to the crucial defining skein relation.

Theorem 5.24. *$J' = J$ up to a change in sign of q; that is, we have recovered the Jones polynomial from the standard representation of $\mathbf{U}_q(\mathfrak{sl}_2)$.*

Proof. As above. □

References

[Kh] M. Khovanov, *A categorification of the Jones polynomial,* Duke Mathematical Journal **101-3** (2000) 359–426, arXiv:math.QA/9908171.

[BN] D. Bar-Natan, *Khovanov's Homology for Tangles and Cobordisms,* Geometry and Topology **9-33** (2005) 1443–1499, arXiv:math.GT/0410495.

[RT] Reshetikhin, N.Y., Turaev, V.G. *Ribbon graphs and their invaraints derived from quantum groups,* Commun.Math. Phys. 127, **1–26** (1990), https://doi.org/10.1007/BF02096491.

[Ja] Jantzen, J.C., *Lectures on Quantum Groups,* Graduate Studies in Mathematics Vol. 6 (1996).

[Lu] Lusztig, G., *Introduction to Quantum Groups,* Modern Birkhäuser Classics (2010).

[St] C. Stroppel. *Parabolic category O, perverse sheaves on Grassmannians, Springer fibres and Khovanov homology,* Compositio Mathematica **145(4)** (2009) 954–992, arXiv:math.RT/0608234.

[KL] M. Khovanov, A.D. Lauda, *A diagrammatic approach to categorification of quantum groups III,* Quantum Topology (2010) 1–92, arXiv:math.QA/0807.3250.

Department of Mathematics, National University of Singapore, Singapore 119076, Singapore.

Email address: `bwang@u.nus.edu`