LECTURES ON TOPOLOGY OF WORDS

VLADIMIR TURAEV

Based on notes by Eri Hatakenaka, Daniel Moskovich, and Tadayuki Watanabe

Abstract. We discuss a topological approach to words introduced by the author in [Tu2]–[Tu4]. Words on an arbitrary alphabet are approximated by Gauss words and then studied up to natural modifications inspired by the Reidemeister moves on knot diagrams. This leads us to a notion of homotopy for words. We introduce several homotopy invariants of words and give a homotopy classification of words of length five.

1. Introduction

Words are finite sequences of symbols, called letters, belonging to a given set α, called an alphabet. In these lectures we discuss an approach to combinatorics of words based on their analogy with curves on the plane. To begin with, consider Figure 1 depicting a plane curve with distinguished origin and orientation. The three crossing points of the curve are labeled by letters $a, b, c \in \alpha$. Now, starting at the origin we move along the curve. The first crossing met is labeled a, the second one b, and so on. Finally we return to the origin and stop. Writing down the labels of the crossings as we encounter them, we obtain the word *abcabc*. This procedure, deriving a word from a closed plane curve with labeled double transversal intersections was introduced by Gauss [Ga] in his attempt to classify such curves. Clearly, every letter appears in the resulting word twice. Words in which every letter appears twice are called Gauss words.

It is easy to see that not all Gauss words can be realized by closed plane curves. The word *abab* for instance is not realizable by such a curve— see Figure 2. Various conditions on Gauss words necessary and sufficient for their realizability by closed plane curves were obtained by several authors, see [Ma], [LM], [Ro], [DT], [CW], [CE], [CR].

The aim of these lectures is to study *arbitrary* words using ideas taken from the topology of curves. To do this, we generalize the above picture

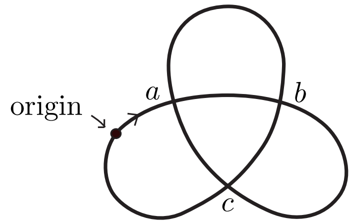


Figure 1: A plane curve associated with the word *abcabc*

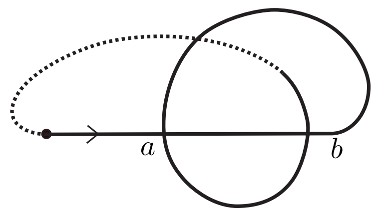


Figure 2: The word *abab* is not realizable by a closed curve on $\mathbb{R}^2$

in the following three directions. First of all, we allow curves with self-crossings of arbitrary multiplicity ≥ 2. For example, the word associated with the curve in Figure 3 is *ababa*. In general, the number of occurrences of the label of a crossing in the associated word is equal to the multiplicity of this crossing. To handle letters appearing only once, we may distinguish a finite set of generic points on the curve and label them as well; we shall not do that here.

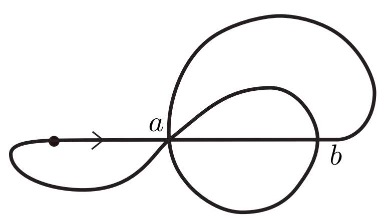


Figure 3: A curve with a triple point yielding the word *ababa*

The second direction in which we may generalize is to allow *unlabeled* or *virtual* crossings. Such crossings do not contribute at all to the associated word. For example, the curve in Figure 4 gives rise to the word *abab*. The idea of unlabeled crossings is inspired by the theory of virtual knots introduced by L. Kauffman [Ka].

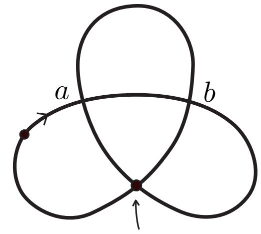


Figure 4: A curve with an unlabeled crossing

Thirdly, we may allow the same label to be used on different crossings. In Figure 5 there are three crossings A_1, A_2, A_3, all labeled by the same letter $a \in \alpha$. This leads us to so-called étale words that are words in an alphabet projecting to the given (fixed) alphabet α. The curve on Figure 5 gives rise to the étale word $A_1 A_2 A_3 A_1 A_2 A_3$ where the letters A_1, A_2, A_3 project to $a \in \alpha$. This étale word is a Gauss word in the alphabet $\{A_1, A_2, A_3\}$; we call such étale words *nanowords*.

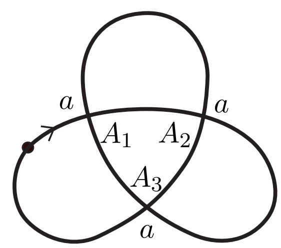


Figure 5: A curve with label a on different crossings

The general scheme of the topology of words is as follows: arbitrary words on the given alphabet α are approximated by nanowords and the latter are studied by methods inspired by the topology of curves. It is certainly interesting for topologists to apply topological methods to study such new objects. Our approach also leads to new questions concerning combinatorics of words. The accent in this theory shifts from words themselves to transformations of words, inspired by topology. The situation is similar to the one in knot theory where one focuses on isotopy classes of knots rather than on specific knot diagrams.

The present exposition follows my lectures given in the Research Institute for Mathematical Sciences (RIMS, Kyoto) in February 2006 and is based on my papers [Tu1]–[Tu4]. The lectures were organized by Prof. Tomotada Ohtsuki. The lecture notes taken by Eri Hatakenaka, Daniel Moskovich and Tadayuki Watanabe served as the basis for this paper. I would like to express my gratitude to Prof. Ohtsuki for inviting me to RIMS and for organizing the lectures and to Eri Hatakenaka, Daniel Moskovich and Tadayuki Watanabe for the preparation of the notes.

2. Words and nanowords

In this section we give formal definitions of words, étale words, and nanowords. Fix a set α called the *alphabet*.

2.1. **Words.**

A *word of length* $n \geq 1$ on α is a mapping

$$w\colon \widehat{n} \to \alpha,$$

where $\widehat{n}$ is the set $\{1, 2, \ldots, n\}$. That is, a word of length n on α is a sequence of n elements of α. For example, the word $w = aba$, with $a, b \in \alpha$, is nothing but the map

$$w\colon \{1, 2, 3\} \to \alpha,$$

defined by $w(1) = a$, $w(2) = b$, and $w(3) = a$. By convention, there is one empty word ϕ of length 0.

The *opposite word* to a word $w\colon \widehat{n} \to \alpha$, is denoted by w^- and defined by $w^-(i) = w(n + 1 - i)$ for all $i \in \widehat{n}$. For instance, if $w = abca$, then $w^- = acba$.

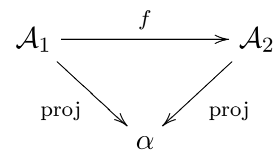
Concatenation of two words is defined by writing down the first word and then the second one. For example, the concatenation of the words $w = abc$ and $v = dbba$ on α is the word $wv = abcdbba$.

One more operation on words is a change of the alphabet. For a map $f\colon \alpha \to \alpha'$ from an alphabet α to an alphabet α' and a word $w\colon \widehat{n} \to \alpha$, set $f_\sharp(w) = f \circ w$. This is a word on α' of the same length n. For example, if $w = abca$, then $f_\sharp(w) = f(a)f(b)f(c)f(a)$.

2.2. **Étale words.**

An α*–alphabet* is a set $\mathcal{A}$ endowed with a mapping to α, called the *projection*. The image in α of any $A \in \mathcal{A}$ will be denoted $|A|$.

A *morphism* of α–alphabets $\mathcal{A}_1$ and $\mathcal{A}_2$ is a mapping $f\colon \mathcal{A}_1 \to \mathcal{A}_2$ such that $|A| = |f(A)|$ for any $A \in \mathcal{A}_1$. This means that the diagram



commutes. An *isomorphism* of α–alphabets is a bijective morphism.

An *étale word* over α is a pair (an α–alphabet $\mathcal{A}$, a word in the α–alphabet $\mathcal{A}$). In particular, every word on α becomes an étale word over α by regarding α as an α–alphabet $\mathcal{A} = \alpha$ with projection to α being the identity map. Another example: let $a \in \alpha$ and $\mathcal{A} = \{A_1, A_2, A_3\}$ with $|A_1| = |A_2| = |A_3| = a$. The pair $(\mathcal{A}, A_1A_2A_3A_1A_2A_3)$ is an étale word over α. It corresponds to the picture on Figure 5.

Two étale words $(\mathcal{A}_1, w_1)$ and $(\mathcal{A}_2, w_2)$ over α are *isomorphic* if there is an isomorphism $f\colon \mathcal{A}_1 \to \mathcal{A}_2$ such that $w_2 = f_\sharp(w_1)$. The relation of isomorphism will be denoted $\approx$. For example, if $\mathcal{B} = \{B_1, B_2, B_3\}$ is an α-alphabet with $|B_1| = |B_2| = |B_3| = a \in \alpha$, then

$$(\mathcal{B}, B_1B_2B_3B_1B_2B_3) \approx (\mathcal{A}, A_1A_2A_3A_1A_2A_3)$$

where the étale word on the right is as in the previous paragraph.

For an étale word $(\mathcal{A}, w)$, the *opposite étale word* is defined by $(\mathcal{A}, w)^- = (\mathcal{A}, w^-)$. The *product* of étale words $(\mathcal{A}_1, w_1)$ and $(\mathcal{A}_2, w_2)$ over α is defined as follows. If $\mathcal{A}_1 \cap \mathcal{A}_2 = \phi$, then the pair $(\mathcal{A}_1 \cup \mathcal{A}_2, w_1w_2)$ is an étale word over α, and we call it the product of $(\mathcal{A}_1, w_1)$ and $(\mathcal{A}_2, w_2)$. If $\mathcal{A}_1 \cap \mathcal{A}_2 \neq \phi$, then we pick an étale word $(\mathcal{A}'_1, w'_1)$ over α isomorphic to $(\mathcal{A}_1, w_1)$ and such that $\mathcal{A}'_1 \cap \mathcal{A}_2 = \phi$. We call then the étale word $(\mathcal{A}'_1 \cup \mathcal{A}_2, w'_1w_2)$, the product of $(\mathcal{A}_1, w_1)$ and $(\mathcal{A}_2, w_2)$. The product of étale words is well defined up to isomorphism.

Beware that concatenation of words on α usually differs from multiplication of the corresponding étale words. For example, for the words $w_1 = abb$ and $w_2 = aa$ on the alphabet $\alpha = \{a, b\}$, the corresponding étale words are $(\mathcal{A}_1 = \{A, B\}, ABB)$ with $|A| = a, |B| = b$, and $(\mathcal{A}_2 = \{A', B'\}, A'A')$ with $|A'| = a, |B'| = b$. Their product is the étale word $(\{A, B, A', B'\}, ABBA'A')$, with $|A| = |A'| = a, |B| = |B'| = b$, while the étale word corresponding to $w_1w_2 = abbaa$ is $(\{A, B\}, ABBAA)$ with $|A| = a, |B| = b$. These two étale words are not isomorphic.

2.3. **Nanowords.** A word w on a finite alphabet is a *Gauss word* if every letter of this alphabet appears in w exactly two times. For instance, the words $AABB, ABAB, ABBA, BAAB, BABA, BBAA$, are all Gauss words on the alphabet $\{A, B\}$. The words AA, ABA are not Gauss words on this alphabet.

An étale word $(\mathcal{A}, w)$ is a *nanoword* over α if w is a Gauss word on $\mathcal{A}$. Then the alphabet $\mathcal{A}$ is finite and the number of its elements is equal to the half of the length of w. For example, let $\mathcal{A} = \{A, B\}$ with $|A| = a \in \alpha$ and $|B| = b \in \alpha$. The étale word $(\mathcal{A}, ABAB)$ is a nanoword. Another example: $\mathcal{A} = \{A, B, C\}$ with $|A| = a \in \alpha$, $|B| = b \in \alpha$, and $|C| = c \in \alpha$. Then the étale word $(\mathcal{A}, ABCBCA)$ is a nanoword.

Two nanowords $(\mathcal{A}_1, w_1)$ and $(\mathcal{A}_2, w_2)$ are *isomorphic* if they are isomorphic as étale words.

We now make a few simple remarks about nanowords. If $(\mathcal{A}, w)$ is a nanoword then its opposite $(\mathcal{A}, w)^-$ is also a nanoword. The concatenation of two nanowords is a nanoword. An empty étale word $(\mathcal{A} = \phi, w = \phi)$ is a nanoword. The set of nanowords over α is infinite provided $\alpha \neq \phi$.

Note finally that a plane curve with fixed origin and labeled crossings gives rise to a nanoword if and only if this curve is generic— that is all its self-intersections are double transverse crossings. Thus, the nanowords can be thought of as combinatorial analogues of generic curves.

2.4. **Desingularization.** Consider again the curve on Figure 3. By a small deformation near the triple point, we obtain a generic curve whose singularities are double points $A_{1,2}, A_{2,3}, A_{1,3}, B$ shown on Figure 6. We label the points $A_{1,2}, A_{2,3}, A_{1,3}$ by a and the point B by b. Thus, each crossing of the deformed curve has the same label as the corresponding crossing of the original curve. The left curve on Figure 6 gives the word $w = ababa$. The right curve gives the Gauss word $w^d = A_{1,2}A_{1,3}BA_{1,2}A_{2,3}BA_{1,3}A_{2,3}$ in the α–alphabet $\{A_{1,2}, A_{2,3}, A_3, B\}$ with $|A_{1,2}| = |A_{1,3}| = |A_{2,3}| = a$ and $|B| = b$. We view the nanoword w^d over α as the desingularization of the word $w = ababa$.

The desingularization procedure considered in this example can be generalized and leads thus to desingularization of arbitrary étale words. More precisely, for every étale word $(\mathcal{A}, w)$ over α, we define a nanoword $(\mathcal{A}^d, w^d)$ over α called its *desingularization*. For any $A \in \mathcal{A}$, the *multiplicity* $m_w(A)$ is the number of times that A occurs in w. For instance, $m_{ABB}(A) = 1$ and

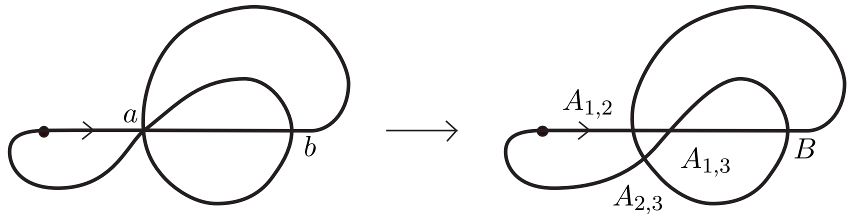


Figure 6: Desingularization of a curve

$m_{ABB}(B) = 2$. We define the set $\mathcal{A}^d$ by

$$\mathcal{A}^d = \{(A, i, j) \mid A \in \mathcal{A},\ 1 \leq i < j \leq m_w(A)\}.$$

Denoting (A, i, j) by $A_{i,j}$, we define the projection $\mathcal{A}^d \to \alpha$ by $|A_{i,j}| = |A| \in \alpha$. This makes $\mathcal{A}^d$ into an α–alphabet. Here every letter of multiplicity $m \geq 1$ in w gives rise to $\frac{m(m-1)}{2}$ letters in $\mathcal{A}^d$. The word w^d on the alphabet $\mathcal{A}^d$ is defined in two steps.

Step 1. Delete from w all letters of multiplicity 1.

Step 2. For each $A \in \mathcal{A}$ with $m_w(A) \geq 2$ and for each $i = 1,\ 2,\ \ldots,\ m_w(A)$, we replace the i-th entry of A in w by

$$A_{1,i}\ A_{2,i}\ \cdots\ A_{i-1,i}\ A_{i,i+1}\ A_{i,i+2}\ \cdots\ A_{i,m_w(A)}.$$

The resulting word w^d on $\mathcal{A}^d$ is a Gauss word and the pair $(\mathcal{A}^d, w^d)$ is a nanoword over α. For example, for the étale word $(\mathcal{A} = \{A, B\}, w = ABABA)$ with $|A| = a, |B| = b$, this procedure gives the nanoword

$$(\mathcal{A}^d, w^d) = (\{A_{1,2}, A_{2,3}, A_{1,3}, B_{1,2}\}, A_{1,2}A_{1,3}B_{1,2}A_{1,2}A_{2,3}B_{1,2}A_{1,3}A_{2,3})$$

with $|A_{1,2}| = |A_{2,3}| = |A_{1,3}| = a$ and $|B_{1,2}| = b$.

We can say that the desingularization of curves is based on viewing the crossings under a strong microscope which allows us to see the "internal structure" of each crossing. This analogy suggested the term nanoword.

3. Knot theory and homotopy of nanowords

3.1. **Knot theory.** We shall study nanowords using the analogy with curves and knots. Recall the classical Reidemeister moves on knot diagrams in $\mathbb{R}^2$, shown in Figure 7. (The inverse moves are also called the Reidemeister moves.) These moves and the isotopy of knot diagrams in $\mathbb{R}^2$ generate an equivalence relation on knot diagrams which we call the *R–equivalence*. We have

$$\{\text{knots in } \mathbb{R}^3\}/\text{ isotopy} = \{\text{knot diagrams}\}/\text{ R–equivalence.}$$

This fundamental equality, due to K. Reidemeister, reduces the study of isotopy classes of knots to a study of R–equivalence classes of knot diagrams.

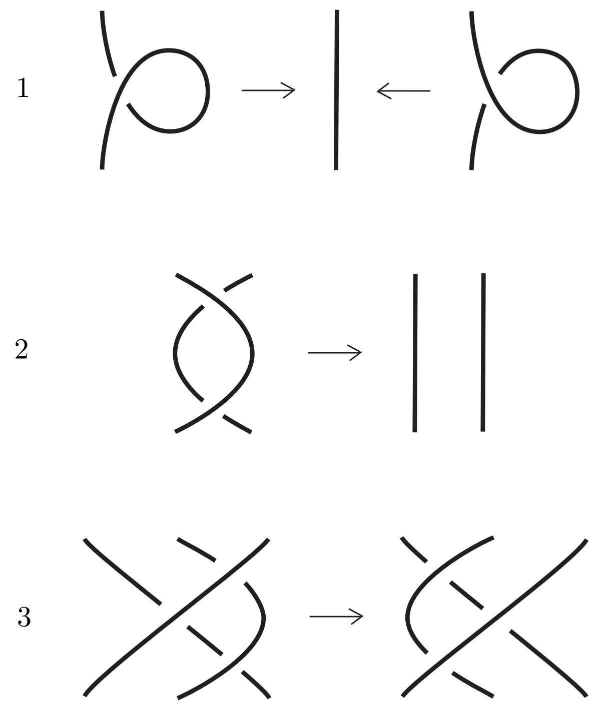


Figure 7: Reidemeister moves on knot diagrams

Similar moves can be considered on pointed curves. For the sake of the following discussion, we switch to curves— that is we make no distinction between under-crossings and over-crossings. We always assume that the moves act away from the origins of the curves.

Let us look at the effect of the Reidemeister moves on words associated with curves. The first Reidemeister move, shown in Figure 8, acts as $xAAy \mapsto xy$, where x and y are words not including the letter A.

Consider the second Reidemeister move with labels, orientations, and the position of the origin as in Figure 9. The move acts on the associated word as $xAByBAz \mapsto xyz$ where x, y, z are words not including the letters A, B. For another choice of orientations, the move may act as $xAByABz \mapsto xyz$.

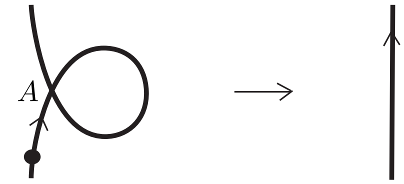


Figure 8: First Reidemeister move on a labeled curve

The first version $xAByBAz \mapsto xyz$ is sufficient for our aims as will be clear from the results below.

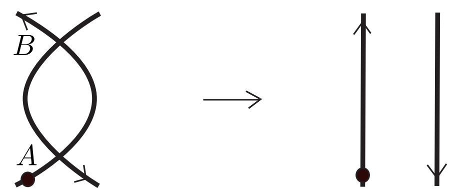


Figure 9: Second Reidemeister move on a labeled curve

Consider the third Reidemeister move with labels, orientations, the position of the origin, and the order of branches as in Figure 10. This move acts on the associated word as $xAByACzBCt \mapsto xBAyCAzCBt$ where x, y, z, t are words not including the letters A, B, C. For other choices of orientations, order of branches etc., the move may act differently but the version shown on Figure 10 is sufficient for our aims.

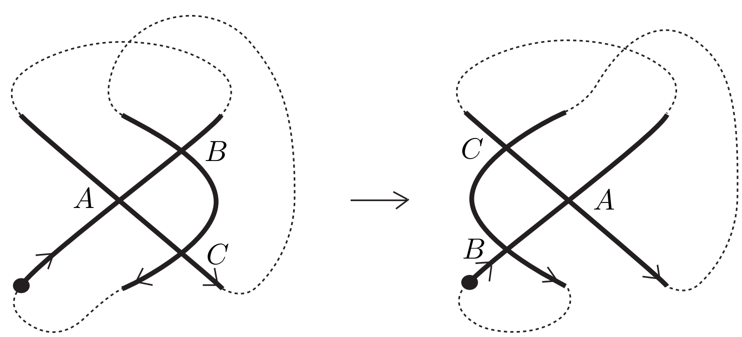


Figure 10: Third Reidemeister move on a labeled curve

3.2. **Homotopy of nanowords.** Let α be an alphabet (a fixed set). We fix *homotopy data* consisting of an involution $\tau\colon\ \alpha \to \alpha$ and an arbitrary set $\mathcal{S} \subset \alpha^3 = \alpha \times \alpha \times \alpha$. The geometric meaning of τ and $\mathcal{S}$ will be discussed in the next section. In the context of knot diagrams, τ switches between positive and negative crossings. The role of $\mathcal{S}$ is to determine the transformations of labels accompanying the third Reidemeister move.

Three *homotopy moves* on nanowords over α are defined as follows.

(1) $(\mathcal{A}, xAAy) \mapsto (\mathcal{A} - \{A\}, xy)$ where x, y are words on the alphabet $\mathcal{A} - \{A\}$. Note that if $(\mathcal{A}, xAAy)$ is a nanoword, then so is $(\mathcal{A} - \{A\}, xy)$. The inverse move adds a new letter A to the α–alphabet and inserts AA into the word.

(2) $(\mathcal{A}, xAByBAz) \mapsto (\mathcal{A} - \{A, B\}, xyz)$ provided $|A| = \tau(|B|)$ and x, y, z are words on the alphabet $\mathcal{A} - \{A, B\}$.

(3) $(\mathcal{A}, xAByACzBCz) \mapsto (\mathcal{A}, xBAyCAzCBt)$, provided $(|A|, |B|, |C|) \in \mathcal{S}$ and x, y, z, t are words on the alphabet $\mathcal{A} - \{A, B, C\}$.

Two nanowords over α are *$\mathcal{S}$–homotopic* if they can be related by a finite sequence of homotopy moves, inverse moves, and isomorphisms. We denote this equivalence relation by $\simeq_{\mathcal{S}}$ and call it *$\mathcal{S}$–homotopy.* This definition readily extends to étale words: étale words w_1 and w_2 are *$\mathcal{S}$–homotopic* if $w_1^d \simeq_{\mathcal{S}} w_2^d$. In particular, the notion of $\mathcal{S}$–homotopy applies to words on α.

We will use the following notation:

$$\mathcal{N}(\alpha, \mathcal{S}) = \{\text{set of nanowords over } \alpha\} \big/ \mathcal{S}\text{–homotopy}.$$

Clearly, $\mathcal{N}(\alpha, \mathcal{S})$ is a monoid, with the empty nanoword as its unit element and concatenation as its product. This monoid depends on τ which is however omitted in the notation $\mathcal{N}(\alpha, \mathcal{S})$ to make it shorter.

The following two lemmas show that our three moves generate a wider set of similar moves. In the context of Figures 9 and 10, the new moves correspond to other choices of orientations, branch connections, etc.

Lemma 3.1. *Let A, B, C be three distinct letters in an α–alphabet $\mathcal{A}$ and let x, y, z, t be words in the alphabet $\mathcal{A} - \{A, B, C\}$ such that $xyzt$ is a Gauss word in this alphabet. Then we have the following $\mathcal{S}$–equivalences:*

(1) $$(\mathcal{A}, xAByCAzBCt) \simeq_{\mathcal{S}} (\mathcal{A}, xBAyACzCBt)$$

if $(|A|, \tau(|B|), |C|) \in \mathcal{S}$,

(2) $$(\mathcal{A}, xAByCAzCBt) \simeq_{\mathcal{S}} (\mathcal{A}, xBAyACzBCt)$$

if $(\tau(|A|), \tau(|B|), |C|) \in \mathcal{S}$, *and*

$$(\mathcal{A}, xAByACzCBt) \simeq_{\mathcal{S}} (\mathcal{A}, xBAyCAzBCt) \tag{3}$$

if $(|A|, \tau(|B|), \tau(|C|)) \in \mathcal{S}$.

Lemma 3.2. *Suppose that* $\mathcal{S} \cap (\alpha \times b \times b) \neq \phi$ *for all* $b \in \alpha$. *Let* $(\mathcal{A}, xAByABz)$ *be a nanoword over* α *where* $A, B \in \mathcal{A}$ *with* $|A| = \tau(|B|)$ *and* x, y, z *are words on the alphabet* $\mathcal{A} - \{A, B\}$. *Then*

$$(\mathcal{A}, xAByABz) \simeq_{\mathcal{S}} (\mathcal{A} - \{A, B\}, xyz).$$

Proof. Set $b = |B| \in \alpha$. By assumption, there is $e \in \alpha$ such that $(e, b, b) \in \mathcal{S}$. Pick a letter E not belonging to $\mathcal{A}$ and set $|E| = \tau(e) \in \alpha$. Then

$$\begin{aligned}
(\mathcal{A}, xAByABz) &\overset{(\text{Move } 1)^{-1}}{\simeq_{\mathcal{S}}} (\mathcal{A} \cup \{E\}, xAEEByABz) \\
&\simeq_{\mathcal{S}} \quad (\mathcal{A} \cup \{E\}, xEABEyBAz) \\
&\overset{(\text{Move } 2)}{\simeq_{\mathcal{S}}} (\mathcal{A} \cup \{E\} - \{A, B\}, xEEyz) \\
&\overset{(\text{Move } 1)}{\simeq_{\mathcal{S}}} (\mathcal{A} - \{A, B\}, xyz)
\end{aligned}$$

In the second line we use the $\mathcal{S}$–homotopy of Lemma 3.1.(2), where A is replaced by E, B by A, and C by B. This homotopy applies since $(\tau(|E|), \tau(|A|), |B|) = (e, b, b) \in \mathcal{S}$. □

3.3. **Typical questions.** In analogy with knot theory, the main objective of the homotopy theory of words is to classify étale words and nanowords up to $\mathcal{S}$–homotopy. Putting it differently, the goal is to compute the monoid $\mathcal{N}(\alpha, \mathcal{S})$ at least for some choices of α, τ, and $\mathcal{S}$. We are very far from reaching this goal. Available results are outlined in the rest of the paper.

Taking knot theory as a model, we list here several typical questions concerning the homotopy of words.

(Q1) Is a given nanoword $\mathcal{S}$–contractible, *i.e.*, $\mathcal{S}$–homotopic to the empty nanoword ? This question corresponds to the question of whether or not a given knot diagram presents an unknot.

(Q2) Is a given nanoword w homotopically symmetric, that is $\mathcal{S}$–homotopic to w^- ? Note that opposite words corresponds to knots with opposite orientations.

(Q3) Define the *length norm* of an étale word w by

$$||w||_{\mathcal{S}} = \frac{1}{2} \text{ (minimal length of a nanoword } \mathcal{S}\text{-homotopic to } w).$$

Note that $||w||_{\mathcal{S}} = 0$ if and only if w is contractible. For any étale words w_1, w_2,

$$||w_1 w_2||_{\mathcal{S}} \leq ||w_1||_{\mathcal{S}} + ||w_2||_{\mathcal{S}}.$$

Compute the length norm.

4. Curves and knots as nanowords

In this section we clarify the relations between curves, knots, and nanowords.

4.1. **Curves as nanowords.** In the sequel, the word "curve" means the image of a generic immersion of an oriented circle into an oriented surface. Here "generic" means that the curve has only a finite set of self-intersections, which are all double and transversal. The curve may be immersed into any oriented surface, of any genus, compact or not, with boundary or not. Note that all self-intersections of a curve look locally like × . Triple points and self-tangencies are not allowed. Every curve has a *regular neighborhood.* This is a narrow band around the curve inside the surface, see Figure 11. Note that the orientation of the ambient surface induces an orientation of the regular neighborhood.

A curve is *pointed* if it is provided with a base-point which is not a self-intersection. An example of a pointed curve is drawn on Figure 11:

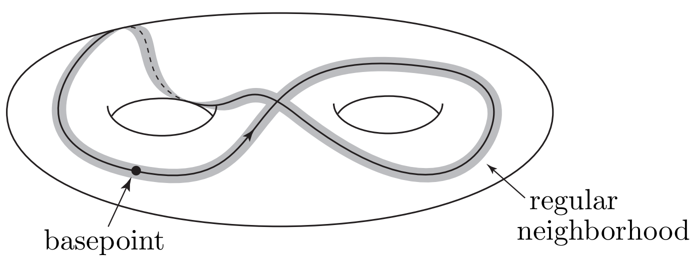


Figure 11: A curve on a surface of genus two

Two pointed curves are *stably homeomorphic* if their regular neighborhoods look exactly the same including the position of the curves in these neighborhoods. Here is a more precise definition.

Definition 4.1. *Two pointed curves are stably homeomorphic if there is an orientation-preserving homeomorphism of their regular neighborhoods mapping the first curve onto the second one and preserving the origin and the orientations of the curves.*

The stable homeomorphism class of a curve is determined solely by the germ of the ambient surface near the curve; what happens outside a regular neighborhood does not matter. In particular, adding handles and punctures to the surface away from a neighborhood of the curve does not change its stable homeomorphism class.

Recall the definition of the *stable equivalence* of curves from [KK], [CKS].

Definition 4.2. *Two pointed curves are stably equivalent if they can be related by a finite sequence of the following transformations:*

(1) Stable homeomorphism.
(2) Homotopy of the curve in its ambient surface away from the origin.

The homotopy in (2) may push a branch of the curve across another branch or across a double point but not across the origin of the curve.

Pointed curves related by any sequence of moves (1), (2) are stably equivalent. Thus, we may start with a curve, transform it by a stable homeomorphism, deform the resulting curve, add handles, deform again, puncture the surface, etc. All these transformations preserve the stable equivalence class of the curve. As an exercise, the reader may show that any two pointed curves on the 2–sphere are stably equivalent. The same is true for curves on the 2–torus. Pointed curves on surfaces of higher genus are not necessarily stably equivalent. The classification of stable equivalence classes of pointed curves is an interesting topological problem.

We note here three geometric invariants of pointed curves preserved under the stable equivalence: the minimal crossing number, the genus, and the virtual number. The minimal crossing number $\|c\|$ of a pointed curve c is the minimal number of crossings of a pointed curve stably equivalent to c. The genus $g(c)$ is the minimal integer $g \geq 0$ such that c is stably equivalent to a pointed curve on a closed surface of genus g. To define the virtual number of c, note that any pointed curve on $\mathbb{R}^2$ with a distinguished set of "virtual" crossings represent a stable equivalence class of pointed curves. One simply does not look at the virtual crossings or, equivalently, trades a branch of c near each virtual crossing for a branch going along a small 1-handle attached to $\mathbb{R}^2$ and avoiding the rest of the curve. The virtual number $v(c)$ is the minimal integer $v \geq 0$ such that there is a pointed curve on $\mathbb{R}^2$ with v virtual crossings representing the stable equivalence class of c. It is clear that $v(c) \geq g(c)$.

Denote by $\mathcal{C}$ the set of stable equivalence classes of pointed curves. The elements of $\mathcal{C}$ are called long flat knots [Ka] or open virtual strings [Tu1].

We now relate the theory of curves with the theory of nanowords. Consider the following homotopy data:

$$\alpha_0 = \{a, b\}, \ \tau_0(a) = b, \tau_0(b) = a, \ \mathcal{S}_0 = \{(a, a, a), (b, b, b)\} \subset \alpha_0^3.$$

Theorem 4.3. *There is a canonical bijection*

$$\mathcal{C} \xrightarrow{\approx} \mathcal{N}(\alpha_0, \mathcal{S}_0).$$

This theorem shows that the theory of nanowords includes the theory of pointed curves as a special case. We outline a construction of the bijection $\mathcal{C} \to \mathcal{N}(\alpha_0, \mathcal{S}_0)$. Consider a pointed curve on a surface. Label its crossings in an arbitrary way by different letters $A_1, A_2, \ldots, A_n$ where n is the number of crossings. The Gauss word of the curve is obtained by moving along the curve starting at the origin and writing down the letters as we encounter them, finishing when we get back to the origin. The resulting word, w, on the alphabet

$$\mathcal{A} = \{A_1, A_2, \ldots, A_n\}$$

contains every letter $A_1, A_2, \ldots, A_n$ twice. We provide $\mathcal{A}$ with the projection to α_0 as follows. Consider the crossing of the curve labeled A_i. If when moving as above along the curve, we first traverse this crossing from the bottom-left to the top-right, then $|A_i| = a$, otherwise $|A_i| = b$; see Figure 12, where the orientation of the ambient surface is counterclockwise. The dot on the left (resp. right) picture is the bottom-left (resp. bottom-right) entry of the crossing. In this way the set $\mathcal{A}$ becomes an α_0–alphabet. We assign to our curve the class of this nanoword in $\mathcal{N}(\alpha_0, \mathcal{S}_0)$. We must prove that stably equivalent curves give rise to $\mathcal{S}_0$–homotopic nanowords. A different choice of the labeling of the crossings gives an isomorphic nanoword. If the curve is changed by a stable homeomorphism, then the associated nanoword does not change, since it is defined entirely by the behavior of the curve in its regular neighborhood. A homotopy of the curve may be split into a composition of local Reidemeister moves and the inverse moves. Then one verifies that under these moves the associated nanoword changes via the $\mathcal{S}_0$–homotopy moves and the transformations in Lemmas 3.1 and 3.2. The resulting mapping $\mathcal{C} \to \mathcal{N}(\alpha_0, \mathcal{S}_0)$ is bijective, see [Tu3].

Figure 12: On the left $|A_i| = a$ and on the right $|A_i| = b$

Under the identification $\mathcal{C} = \mathcal{N}(\alpha_0, \mathcal{S}_0)$ the minimal crossing number of curves corresponds to the length norm on $\mathcal{N}(\alpha_0, \mathcal{S}_0)$. The genus and the virtual number yield interesting geometric invariants of nanowords over α_0.

One can suppress all references to the origin in the definitions above. This gives a relation of stable homotopy for non-pointed curves. To obtain a corresponding notion for the nanowords, one has to introduce an additional move on nanowords, the so-called circular shift. Briefly speaking, the shift moves the last letter of the word to the first position. For details, see [Tu3].

4.2. **Knots as nanowords.** The constructions of the previous section can be upgraded to the setting of knot diagrams. By a (pointed) knot diagram we mean a (pointed) curve on an oriented surface with additional data at each crossing: one of the branches lies "over" and the other one lies "under". A pointed knot diagram on $\mathbb{R}^2$ is shown on Figure 13.

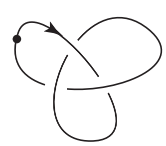


Figure 13: A pointed knot diagram

Two pointed knot diagrams are said to be *stably homeomorphic* if there is an orientation-preserving homeomorphism of the regular neighborhoods of the underlying curves, sending the first diagram onto the second one and preserving the origin, the orientation, and the over-/under-crossing data.

Recall the *stable equivalence* of knot diagrams from [KK], [CKS].

Definition 4.4. *Two pointed knot diagrams are stably equivalent if they can be related by a finite sequence of the following transformations:*

(1) Stable homeomorphism.

(2) The Reidemeister moves on a knot diagram in its ambient surface away from the origin.

The moves in (2) may push a branch of the diagram above or below a double point or another branch but not across the origin. Thus, the origin may not lie inside the neighborhoods where the Reidemeister moves are performed.

Let $\mathcal{K}$ denote the set of stable equivalence classes of pointed knot diagrams. Elements of $\mathcal{K}$ are called *long virtual knots*, see [Ka], [GPV].

The set $\mathcal{K}$ includes the set of isotopy classes of classical knots. Classical knots are oriented knots in S^3. There is a map

$$\{\text{Classical knots}\}/\text{isotopy} \hookrightarrow \mathcal{K}$$

obtained by picking an arbitrary diagram of the given classical knot, picking an arbitrary origin on the diagram (not a crossing), and taking the stable equivalence class of the resulting pointed knot diagram. This gives a well-defined mapping from the set of isotopy classes of classical knots into $\mathcal{K}$. This mapping is known to be injective, see [Ka], [GPV].

The set $\mathcal{K}$ can be interpreted in terms of nanowords as follows, Set

- $\alpha_* = (a_+, a_-, b_+, b_-)$,
- $\tau_* : \alpha_* \to \alpha_*$ the involution defined by $\tau_*(a_+) = b_-$ and $\tau_*(a_-) = b_+$,
- $\mathcal{S}_\star = \{(a_\pm, a_\pm, a_\pm), (a_\pm, a_\pm, a_\mp), (a_\mp, a_\pm, a_\pm), (b_\pm, b_\pm, b_\pm), (b_\pm, b_\pm, b_\mp), (b_\mp, b_\pm, b_\pm)\} \subset \alpha_*^3$.

Theorem 4.5. *There is a canonical bijection*

$$\mathcal{K} \xrightarrow{\approx} \mathcal{N}(\alpha_*, \mathcal{S}_*)$$

such that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{K} & \xrightarrow{\approx} & \mathcal{N}(\alpha_*, \mathcal{S}_*) \\ \downarrow & & \downarrow \\ \mathcal{C} & \xrightarrow{\approx} & \mathcal{N}(\alpha_0, \mathcal{S}_0). \end{array}$$

Here the map $\mathcal{K} \to \mathcal{C}$ is given by forgetting the over-/under-crossing data, and the map $\mathcal{N}(\alpha_*, \mathcal{S}_*) \to \mathcal{N}(\alpha_0, \mathcal{S}_0)$ is given by $a_\pm \mapsto a$, $b_\pm \mapsto b$.

This theorem shows that the theory of nanowords includes the theory of long virtual knots as a special case. The definition of the bijection $\mathcal{K} \to \mathcal{N}(\alpha_*, \mathcal{S}_*)$ goes similarly to the one for curves. The difference is that now we project to α_* rather than to α_0. The rule is shown on Figure 14.

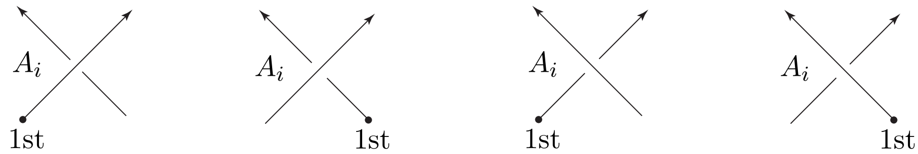


Figure 14: From left to right:
$|A_i| = a_+$, $|A_i| = b_+$, $|A_i| = a_-$, and $|A_i| = b_-$

To extend these ideas to links, one has to involve phrases, *i.e.*, sequences of words, see [Tu3], [Tu4].

4.3. **An extension to general α.** Let α be an arbitrary alphabet with homotopy data $\tau, \mathcal{S}$. Nanowords over α can be geometrically interpreted as follows. Pick a mapping $f\colon \alpha \to \alpha_0$ such that $f\tau = \tau_0 f$ and $f(a) = f(b) = f(c)$ for any triple $(a,b,c) \in \mathcal{S}$. Every nanoword $(\mathcal{A}, w\colon \widehat{n} \to \alpha)$ over α determines a nanoword $f_\sharp(w) = (\mathcal{A}, fw\colon \widehat{n} \to \alpha_0)$ over α_0. The latter can be represented by a pointed curve on a surface. Thus, w gives rise to a family of pointed curves $\{f_\sharp(w)\}_f$ *underlying* w and numerated by f as above. Geometric invariants of these curves provide geometric information about w. Stable equivalence classes of these curves depend only on the $\mathcal{S}$–homotopy class of w.

This geometric representation of w seems to be especially efficient in the case where $\mathcal{S}$ is the diagonal of α^3 so that the conditions on f reduce to the equivariance relation $f\tau = \tau_0 f$. One interesting question: when does a given family of stable equivalence classes of pointed curves numerated by equivariant maps $\alpha \to \alpha_0$ arise from a nanoword over α ?

A similar geometric interpretation of nanowords in terms of knot diagrams can be obtained by replacing α_0 with α_*.

5. Invariant γ and self-linking of nanowords

In this section and in the sequel, the symbol α denotes a (fixed) alphabet with involution $\tau\colon \alpha \to \alpha$.

5.1. **The set $\mathcal{S}$.** For the rest of this paper, we set

$$\mathcal{S} = \text{diagonal} = \{(a,a,a)\}_{a\in\alpha}$$

The third homotopy move is $xAByACzBCt \mapsto xyz$ provided $|A| = |B| = |C|$. In the sequel, we leave $\mathcal{S}$ out of notation. By *homotopy* of nanowords, we mean $\mathcal{S}$–homotopy when $\mathcal{S}$ is the diagonal as above. The homotopy relation is denoted $\simeq$. The case of knots is excluded by this choice of $\mathcal{S}$, but the case of curves is covered.

We give now an example of homotopic words. Pick $a \in \alpha$ such that $\tau(a) \neq a$. Set $b = \tau(a)$. Consider the words:

$$w_1 = aabab, \; w_2 = babaa, \; w_3 = baaab$$

and the nanoword

$$w_4 = (\{A, A'\}, AA'AA') \text{ where } |A| = |A'| = a.$$

Claim. $w_1 \simeq w_2 \simeq w_3 \simeq w_4$

Proof. We prove only that $w_1 \simeq w_4$. The proofs that $w_2 \simeq w_4$ and $w_3 \simeq w_4$ are similar.

The desingularization of w_1 gives

$$w_1^d = A_{1,2}A_{1,3}A_{1,2}A_{2,3}BA_{1,3}A_{2,3}B$$

where $|A_{1,2}| = |A_{1,3}| = |A_{2,3}| = a$ and $|B| = b$. By Lemma 3.2, we can strike out the two occurrences of $A_{2,3}B$. This gives the nanoword $A_{1,2}A_{1,3}A_{1,2}A_{1,3}$ isomorphic to w_4. □

This example shows that the relation of homotopy is quite non-trivial.

5.2. **A group–theoretic homotopy invariant.** We construct here a homotopy invariant of nanowords, γ. First, define a group Π by generators and relations:

$$\Pi = \langle \{z_a\}_{a\in\alpha} \mid z_a z_{\tau(a)} = 1 \text{ for all } a \in \alpha \rangle.$$

Note that if $a \neq \tau(a)$, then we have a free generator $z_a = (z_{\tau(a)})^{-1}$ of Π, and if $a = \tau(a)$, then we have a generator z_a of order 2.

For a nanoword $(\mathcal{A}, w\colon \widehat{n} \to \mathcal{A})$ of length n, we define n elements γ_1, γ_2, ..., γ_n of Π by:

$$\gamma_i = \begin{cases} z_{|w(i)|}, & \text{if } w(i) \neq w(j) \text{ for all } j < i; \\ z_{|w(i)|}^{-1}, & \text{otherwise.} \end{cases}$$

Here $w(i) \in \mathcal{A}$ is the i-th letter of w and $|w(i)| \in \alpha$ is its projection to α. The sequence $\gamma_1, \gamma_2, \ldots, \gamma_n$ may be also described as follows. Since w is a nanoword, each letter of $\mathcal{A}$ appears twice in the sequence $w(1), w(2), \ldots, w(n)$. The first time it appears we write at this place the corresponding generator of Π. At its second appearance, we write down the inverse of that generator. This procedure gives the sequence $\gamma_1, \gamma_2, \ldots, \gamma_n$. Set

$$\gamma(w) = \gamma_1\gamma_2 \cdots \gamma_n \in \Pi.$$

Since each generator appears in this product twice with opposite powers, the abelianization of $\gamma(w)$ is zero. Thus, $\gamma(w)$ lies is in the commutator subgroup $[\Pi, \Pi] \subset \Pi$.

For example, consider the nanoword $w = ABAB$ with $|A| = a \in \alpha$ and $|B| = b \in \alpha$. Then $\gamma(w) = z_a z_b z_a^{-1} z_b^{-1} \in [\Pi, \Pi]$.

Theorem 5.1. *$\gamma(w)$ is a homotopy invariant of w.*

Proof (outline). Under the first homotopy move, $\gamma(xAAy) = \gamma(xy)$ because the first appearance of A contributes z_a and the second appearance of A contributes z_a^{-1} where $a = |A|$. So we have the invariance under the first homotopy move. The other two moves are treated similarly. □

The mapping

$$\gamma\colon\ \mathcal{N}(\alpha) = \mathcal{N}(\alpha, \mathcal{S}) \longrightarrow [\Pi, \Pi]$$

is a monoid homomorphism ($\mathcal{S}$ is the diagonal). It is easy to check that it is surjective.

The group $[\Pi, \Pi]$ can be shown to be free for any τ. If τ has at least two orbits, then this group is non-trivial and by the results above, $\mathcal{N}(\alpha)$ is an infinite monoid. If τ has at least three orbits, then $[\Pi, \Pi]$ has rank ≥ 2, and by the results above, $\mathcal{N}(\alpha)$ is a non-abelian monoid.

Let us consider two examples where γ does not work. The interesting case in topology is the case of curves, where $\alpha = \{a, b\}$ with $\tau(a) = b$. In this case:

$$\Pi = \langle z_a, z_b \mid z_a z_b = 1 \rangle = \mathbb{Z}$$

and $[\Pi, \Pi] = 0$. So, for topology the invariant γ is of no interest. Another example: $\alpha = \{a\}$ with $\tau(a) = a$. In this case $\Pi = \mathbb{Z}/2\mathbb{Z}$ and $[\Pi, \Pi] = 0$.

5.3. **Self-linking of nanowords.** We introduce here another homotopy invariant of nanowords, the so-called self-linking. We begin with the following observation. Consider the words $ABAB$ and $AABB$. The letters A, B are obviously linked or interlaced in the first word and unlinked in the second one. Consider now an arbitrary nanoword $(\mathcal{A}, w)$ over α. We say that two letters $A, B \in \mathcal{A}$ are *w–interlaced* if

$$w = \cdots A \cdots B \cdots A \cdots B \cdots \qquad \text{or} \qquad w = \cdots B \cdots A \cdots B \cdots A \cdots .$$

In the first case set $n_w(A,B) = 1$ and in the second case set $n_w(A,B) = -1$. In all other cases set $n_w(A,B) = 0$. The function n_w is skew-symmetric in the sense that for all $A, B \in \mathcal{A}$,

$$n_w(A,B) = -n_w(B,A) \quad \text{and} \quad n_w(A,A) = n_w(B,B) = 0.$$

Now consider the abelian group $\pi = \Pi/[\Pi,\Pi]$. The group operation in π will be written multiplicatively. Each generator $z_a \in \Pi$ with $a \in \alpha$ projects to an element of π denoted a. Thus,

$$\pi = \langle \{a\}_{a\in\alpha} \mid ab = ba \text{ and } a\tau(a) = 1 \text{ for all } a, b \in \alpha \rangle.$$

For a nanoword $(\mathcal{A}, w)$ over α and every $A \in \mathcal{A}$, set

$$[A]_w = \prod_{B\in\mathcal{A}} |B|^{n_w(A,B)} \in \pi.$$

For $a \in \alpha$, set

$$[a]_w = \sum_{\substack{A\in\mathcal{A},\ |A|=a \\ [A]_w\neq 1}} [A]_w \in \mathbb{Z}\pi.$$

The function $\alpha \mapsto \mathbb{Z}\pi, a \mapsto [a]_w$ is called the *self-linking* of w. The following theorem derives from this function a homotopy invariant of w.

Theorem 5.2.

(1) *For any $a \in \alpha$, the difference $[a]_w - [\tau(a)]_w \in \mathbb{Z}\pi$ is a homotopy invariant of w.*

(2) *If $\tau(a) = a$, then $[a]_w \bmod 2 \in (\mathbb{Z}/2\mathbb{Z})\pi$ is a homotopy invariant of w.*

For a proof, we refer to [Tu2]. By this theorem, the letters of α give rise to homotopy invariants of nanowords over α. These invariants reflect the linking of letters in nanowords. Here is a simple application of this invariant.

Pick $a, b \in \alpha$ and consider the nanoword $w = w_{a,b} = ABAB$ with $|A| = a$ and $|B| = b$. By Lemma 3.2, if $a = \tau(b)$, then w is contractible. We can use the invariant γ and the self-linking invariant to show the converse: if w is contractible, then $a = \tau(b)$. Indeed, suppose that $a \neq \tau(b)$. We have $\gamma(w) = z_a z_b z_a^{-1} z_b^{-1}$. If $a \neq b$, then a, b lie in different orbits of τ and therefore z_a does not commute with z_b in Π. Then $\gamma(w) \neq 1$ and w is non-contractible. If $a = b \neq \tau(a)$, then $\gamma(w) = 1$. However, in this case $|A| = |B| = a = b$ and

$$[A]_w = |A|^{n_w(A,A)}|B|^{n_w(A,B)} = a,$$

$$[B]_w = |A|^{n_w(B,A)} \, |B|^{n_w(B,B)} = a^{-1}.$$

Then $[a]_w = a + a^{-1} \neq 0 \in \mathbb{Z}\pi$. Since there are no letters in $\{A, B\}$ projecting to τ_a, we have $[\tau(a)]_w = 0$. These computations and the previous theorem imply that w is non-contractible.

5.4. **Applications of the self-linking.** Recall the norm on the nanowords

$$\|w\| = \frac{1}{2}(\text{minimal length of a nanoword homotopic to } w).$$

We can use the self-linking to estimate this norm from below. The idea is that elements of a group ring may be treated like polynomials and for a polynomial we can consider its degree. When w is not too long, there are not so many factors in the self-linking invariant, and the degree can not be too big. Instead of stating here general theorems, we give an example of the resulting estimate for a specific nanoword. Consider the monoliteral word

$$a^m = aa \cdot \cdots \cdot a$$

formed by m copies of a letter $a \in \alpha$. If $\tau(a) = a$ or $m = 1, 2$, then this word is contractible. If $a \neq \tau(a)$ and $m \geq 3$, then the self-linking invariant gives us the following estimate:

$$\|a^m\| = \left\|(a^m)^d\right\| \geq \left[\frac{m}{2}\right] \times \left[\frac{m-1}{2}\right] + 1$$

where $[x]$ denotes the greatest integer which is smaller than or equal to x. In particular, a^m has a positive norm and is non-contractible. The estimate of $\|a^m\|$ given above is very rough. I suppose that it gives approximately twice the actual value of the norm. My conjecture is that

$$\|a^m\| = \frac{m(m-1)}{2}.$$

Theorem 5.3. *Let $a, b \in \alpha$ such that $a \neq \tau(a)$ and $b \neq \tau(b)$. The words a^m and b^n with $m, n \geq 3$ are homotopic if and only if $a = b$ and $m = n$.*

So such monoliteral words are not homotopic unless they coincide. The proof goes by comparing the self-linking invariants.

5.5. **Geometric interpretation of the self-linking.** We give a geometric interpretation of the self-linking in the case of curves, that is in the case where α consists of two letters a, b permuted by τ. The group π is then the infinite cyclic group with generators a, b satisfying $ab = 1$.

Consider a curve c on an oriented surface with crossings $A_1, A_2, ..., A_n$. Each crossing A_i gives rise to two sub-curves of c as follows. Start from A_i

and go along c in the positive direction until coming back to A_i for the first time. The resulting closed curve is a sub-curve of c. The two branches of c passing through A_i give rise in this way to two sub-curves of c. One of them passes through the origin of c, we call this sub-curve the *thin curve*. The other, complementary sub-curve is called the *thick curve*, cf. Figure 15.

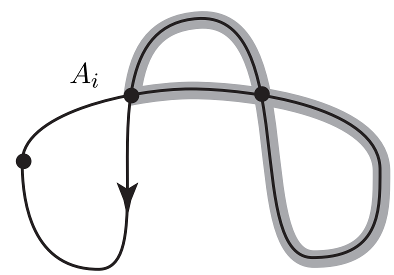


Figure 15: Thin and thick curves associated with a crossing A_i

Consider the homological intersection number:

$$k_i = \text{thin curve} \cdot \text{thick curve} \in \mathbb{Z}.$$

Recall that the homological intersection number of two curves on an oriented surface is obtained by deforming the curves into a transversal position and then counting their intersections with appropriate signs.

Consider the nanoword $(\{A_1, ..., A_n\}, w)$ corresponding to the curve c. Then for all $i = 1, ..., n$,

$$[A_i]_w = |A_i|^{k_i} \in \pi.$$

This formula gives a geometric interpretation of the symbol $[A_i]_w$. The self-linking is obtained by taking the sum of these symbols over the crossings A_i with fixed projection to the alphabet $\{a, b\}$. To ensure the invariance under the first Reidemeister move, we restrict the summation to A_i such that $k_i \neq 0$ or, equivalently, $[A_i]_w \neq 1$.

6. Linking pairings of nanowords

The invariants defined so far, γ and the self-linking, are insufficient to classify even short words. We need more invariants. One idea is to consider again the geometric situation of nanowords associated with curves. With each crossing A_i we associated a 'thick curve' on the ambient surface. We can consider the intersection numbers of these curves with each other. This gives an $n \times n$ integral matrix where n is the number of crossings and the

(i, j) entry is the intersection number of the thick curves determined by A_i and A_j. This matrix can be computed directly from the nanoword. This leads us to so-called linking pairings of nanowords.

In this section, the symbol π denotes the multiplicative abelian group associated with α, τ in Section 5.3.

6.1. α–**pairings.** We begin with purely algebraic definitions, whose connection to nanowords will be explained later.

Definition 6.1. *An α–pairing is a tuple consisting of a set S, a distinguished element $s \in S$, a mapping $S - \{s\} \to \alpha$, and a skew-symmetric pairing $b\colon S \times S \to \pi$.*

By skew-symmetric, we mean that $b(A, B) = b(B, A)^{-1}$ for all $A, B \in S$ and $b(A, A) = 1$ for all $A \in S$.

An α–pairing can be shortly written as

$$(S, s, b\colon\ S \times S \to \pi).$$

The mapping $S - \{s\} \to \alpha$ will be encoded by saying that the set $S - \{s\}$ is an α–alphabet. The image of any $A \in S - \{s\}$ under this mapping will be denoted $|A|$.

The notion of isomorphism for α–pairings is defined in the obvious way.

Given an α–pairing $(S, s, b\colon\ S \times S \to \pi)$, we define *annihilating elements* of S and *twins* as follows.

Definition 6.2. *An element $A \in S - \{s\}$ is annihilating if $b(A, C) = 1$ for all $C \in S$.*

Definition 6.3. *Elements $A, B \in S - \{s\}$ are twins if $b(A, C) = b(B, C)$ for all $C \in S$ and $|A| = \tau(|B|)$.*

An α–pairing is *primitive* if it has no annihilating elements and no twins. For example, the *trivial α–pairing* $(S = \{s\}, b(s, s) = 1)$ is primitive.

We introduce two moves M_1, M_2 on α–pairings.

M_1**:** Delete an annihilating element.

M_2**:** Delete a pair of twins.

The moves M_1, M_2, the inverse moves M_1^{-1}, M_2^{-1}, and isomorphisms of α–pairings generate an equivalence relation on the class of α–pairings, called *homology.* The following theorem classifies α–pairings up to homology.

Theorem 6.4. *Every α–pairing is homologous to a primitive α–pairing. Two homologous primitive α–pairings are isomorphic.*

Thus, in each homology class of α–pairings there is a primitive one unique up to isomorphism. Starting with an arbitrary α–pairing, we can delete annihilating elements and twins and get a primitive α–pairing. The latter is uniquely determined by the homology class of the original α–pairing at least up to isomorphism.

6.2. **From nanowords to α–pairings.** The connection between α–pairings and nanowords is this: to each nanoword w over α we shall assign an α–pairing b_w. Its homology class will be a homotopy invariant of w.

Let $(\mathcal{A}, w\colon \widehat{n} \to \mathcal{A})$ be a nanoword over α. Set $S = \{s\} \cup \mathcal{A}$. We have a projection $S - \{s\} = \mathcal{A} \to \alpha$. The skew-symmetric pairing $b_w\colon S \times S \to \pi$ is defined in four steps.

Step 1. For every $A \in \mathcal{A}$, we can write uniquely $w^{-1}(A) = \{i_A, j_A\} \subset \widehat{n}$ where $i_A < j_A$. Thus i_A is the position in which the letter A appears in w for the first time, and j_A is the position in which the letter A appears in w for the second time. Thus w is of the form:

$$w = \cdots \underset{i_A}{A} \cdots \underset{j_A}{A} \cdots .$$

Step 2. Given two letters $D, E \in \mathcal{A}$, set

$$D \circ E = \prod_{\substack{F \in \mathcal{A} \\ i_D < i_F < j_D \text{ and } i_E < j_F < j_E}} |F| \in \pi.$$

Step 3. The *w–linking* of $D, E \in \mathcal{A}$ is defined by

$$\mathrm{lk}_w(D, E) = (D \circ E)(E \circ D)^{-1} \in \pi.$$

Step 4. Finally, the form $b_w\colon S \times S \to \pi$ is defined as follows: $b_w(s, s) = 1$,

$$b_w(A, s) = [A]_w = \prod_{B \in \mathcal{A}} |B|^{n_w(A,B)} \in \pi,$$

$$b_w(s, A) = ([A]_w)^{-1} \in \pi,$$

$$b_w(A, B) = (\mathrm{lk}_w(A, B))^2 \, |A|^{n_w(A,B)} \, |B|^{n_w(A,B)} \in \pi,$$

for any $A, B \in \mathcal{A} = S - \{s\}$.

The following theorem justifies this definition and relates the homotopy of nanowords to the homology of α–pairings.

Theorem 6.5. *Homotopic nanowords have homologous α–pairings.*

This theorem together with Theorem 6.4, give an efficient method to distinguish nanowords. Given a nanoword w, we first compute the associated α–pairing b_w and then apply the moves M_1 and M_2 to get a primitive α–pairing. The isomorphism class of the latter is a homotopy invariant of w.

6.3. **Applications.** One application of the α–pairings is the following estimate of the length norm of nanowords: if $(S_+, s, b_+ : S_+ \times S_+ \to \pi)$ is a primitive α–pairing homologous to the α–pairing (S, s, b_w) of a nanoword w, then

$$\|w\| \geq \mathrm{card}(S_+) - 1.$$

Indeed, if w is homotopic to a nanoword w' of length $2m$, then the α–pairing (S', s, b') of w' is homologous to (S, s, b_w) and therefore reduces by the moves M_1, M_2 to the same primitive α–pairing (S_+, s, b_+). Hence

$$m = \mathrm{card}(S') - 1 \geq \mathrm{card}(S_+) - 1.$$

In particular, if the α–pairing (S, s, b_w) is primitive, then w has minimal length in its homotopy class.

The α–pairing (S, s, b_w) can be used to estimate the geometric genera of surfaces carrying the underlying curves of w. Pick an equivariant map $f\colon \alpha \to \alpha_0$ where $\alpha_0 = \{a, b\}$ is the 2–letter alphabet with involution permuting a, b. The nanoword $f_\sharp(w)$ over α_0 (defined in Section 4.3) corresponds to a pointed curve on a compact surface. We can estimate the genus g of this surface by $g \geq (1/2)\,\mathrm{rank}(M)$ where M is the skew-symmetric integral matrix obtained from the matrix $\{b_w(s_1, s_2)\}_{s_1, s_2 \in S}$ by the group homomorphism $\pi \to \mathbb{Z}$ sending the generators of π belonging to $f^{-1}(a)$ to $1 \in \mathbb{Z}$ and the generators of π belonging to $f^{-1}(b)$ to $-1 \in \mathbb{Z}$. This estimate follows from the geometric interpretation of the α_0-pairing of $f_\sharp(w)$ in terms of the intersection numbers of curves.

Another area of applications of α–pairings is the homotopy classification of nanowords. With the help of α–pairings we can establish the following theorem. Recall the nanoword $w_{a,b} = ABAB, |A| = a, |B| = b$ defined for any $a, b \in \alpha$. As we know, $w_{a,b}$ is non-contractible if and only if $a \neq \tau(b)$.

Theorem 6.6. *Two non-contractible nanowords $w_{a,b}$ and $w_{a',b'}$ with a, b, a', $b' \in \alpha$ are homotopic if and only if $a = a'$ and $b = b'$.*

Using the α–pairings and the invariants introduced in further sections, we establish the following theorem. It gives a complete homotopy classification of words of length 5 in which one letter, a, occurs 3 times, and another letter, b, occurs 2 times.

Theorem 6.7. *Let a, b be two distinct letters of the alphabet α. Then:*

(1) *The words aaabb, aabba, abbaa, bbaaa are homotopic to each other; they are contractible if and only if $\tau(a) = a$.*
(2) *The word baaab is contractible if and only if $\tau(a) = a$.*
(3) *The word ababa is contractible if and only if $\tau(a) = b$.*
(4) *The words abaab, baaba, aabab, babaa are never contractible.*
(5) *A non-contractible word from (2) – (4) is never homotopic to a word from (1).*
(6) *Two non-contractible words from (2) – (4) are homotopic to each other if and only if they coincide letterwise (i.e., if and only if they are the same word written twice) with the following exceptions:*

$$aabab \simeq babaa \simeq baaab \qquad \text{for } \tau(a) = b.$$

A more general homotopy classification of all words of length ≤ 5 is given in [Tu2]. We can think of such classification theorems as analogues of knot tables. First we draw all possible knot diagrams and then decide which diagrams represent isotopic knots. The same kind of problem arises for the homotopy of words.

6.4. **Examples.** 1. We show how to compute the α–pairing associated with the word $w = abaab$ where $a \neq b$. We have

$$w^d = A_3A_2BA_3A_1A_2A_1B, \ |A_1| = |A_2| = |A_3| = a, \ |B| = b$$

where to simplify notation we write A_1, A_2, A_3 for $A_{2,3}, A_{1,3}, A_{1,2}$ respectively. The matrix for n_w is computed by

$$\begin{bmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{bmatrix}$$

where the rows and columns correspond to A_1, A_2, A_3, B respectively. The matrix for lk_w is computed by

$$\begin{bmatrix} 1 & 1 & a^{-1} & 1 \\ 1 & 1 & 1 & a \\ a & 1 & 1 & a \\ 1 & a^{-1} & a^{-1} & 1 \end{bmatrix}$$

where the rows and columns correspond to A_1, A_2, A_3, B, respectively. Finally, we compute the α–pairing b_w:

$$b_w = \begin{bmatrix} 1 & a & b^{-1} & a^{-1}b^{-1} & a^2 \\ a^{-1} & 1 & a^{-2} & a^{-2} & 1 \\ b & a^2 & 1 & a^{-2} & a^3b \\ ab & a^2 & a^2 & 1 & a^3b \\ a^{-2} & 1 & a^{-3}b^{-1} & a^{-3}b^{-1} & 1 \end{bmatrix}$$

where the rows and columns correspond to s, A_1, A_2, A_3, B, respectively. Recall that $\pi = \langle \{c\}_{c\in\alpha} \mid c\tau(c) = 1\rangle$. In particular a and b are non-trivial elements of π. This implies that the elements A_1 and A_2 of $S = \{s, A_1, A_2, A_3, B\}$ are non-annihilating. If A_3 is annihilating, then $a^2 = ab = 1$. Then $a = b$, which contradicts the assumption $a \neq b$. A similar argument shows that B is non-annihilating. It is also easy to check that b_w does not have twins. Thus, the α–pairing b_w is primitive. Therefore w is non-contractible, and $\|w\| = 4$.

2. The α–pairings are strong enough to distinguish short words and nanowords in many cases. The following example shows however that in some cases the α–pairings are powerless.

Consider the word $w = ababa$ where $\tau(a) = a \neq b = \tau(b)$. A direct computation shows that the α–pairing of w^d is given by the following matrix over π:

$$\begin{bmatrix} 1 & ab & 1 & a^{-1}b^{-1} & 1 \\ a^{-1}b^{-1} & 1 & a^{-2}b^{-2} & a^{-2}b^{-2} & a^{-1}b^{-1} \\ 1 & a^2b^2 & 1 & a^{-2}b^{-2} & 1 \\ ab & a^2b^2 & a^2b^2 & 1 & ab \\ 1 & ab & 1 & a^{-1}b^{-1} & 1 \end{bmatrix}.$$

As above, the rows and columns correspond to s, A_1, A_2, A_3, B respectively. The equality $a = \tau(a)$ implies that $a^2 = 1$ and similarly $b^2 = 1$. Therefore the matrix above simplifies to the following matrix:

$$\begin{bmatrix} 1 & ab & 1 & ab & 1 \\ ab & 1 & 1 & 1 & ab \\ 1 & 1 & 1 & 1 & 1 \\ ab & 1 & 1 & 1 & ab \\ 1 & ab & 1 & ab & 1 \end{bmatrix}.$$

Since the third row and the third column consist only of $1's$, the element A_2 is annihilating. Eliminating it, we observe next that A_1 and A_3 are twins.

What remains after their elimination are two elements s and B. The matrix becomes as follows:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

where the rows and columns correspond to s and B. Now B is an annihilating element. Its elimination gives the trivial α–pairing. Therefore the α–pairing associated with w^d gives no information at all and does not allow to decide whether $w = ababa$ is contractible or not. In fact all invariants of nanowords considered so far are trivial for this word (under the assumptions that $\tau(a) = a \neq b = \tau(b)$).

This example shows that we need more invariants to prove Theorem 6.7. We shall introduce further invariants in the next sections.

7. Further invariants of nanowords

We outline here several ideas inspired by knot theory and leading to homotopy invariants of nanowords over α.

7.1. **Tricolorings.** In knot theory one may treat any knot diagram as consisting of disjoint arcs. A *coloring* of the diagram assigns a residue mod 3 to each arc, such that for every crossing, the sum of the three residues assigned to the adjacent arcs is equal to 0. The number of such colorings of a knot diagram is a knot invariant. This definition is due to R. Fox.

We can introduce similar definitions for words. Fix a set $\beta \subset \alpha$ such that $\tau(\beta) = \beta$ (the resulting invariants may depend on β). Consider a nanoword $w = (\mathcal{A}, w : \widehat{n} \to \mathcal{A})$ over α. For any letter $A \in \mathcal{A}$, let $i_A < j_A$ be the first and the second indices enumerating the positions of A in w as in Section 6.2. A *tricoloring* of w is a function $f\colon \{0, 1, 2, \ldots, n\} \to \mathbb{Z}/3\mathbb{Z}$ such that for any $A \in \mathcal{A}$, if $|A| \in \beta$, then

$$f(i_A) = f(i_A - 1) \quad \text{and} \quad f(j_A - 1) + f(j_A) + f(i_A) = 0$$

and if $|A| \in \alpha - \beta$, then

$$f(j_A) = f(j_A - 1) \quad \text{and} \quad f(i_A - 1) + f(i_A) + f(j_A) = 0.$$

The residues $f(0)$ and $f(n)$ are called respectively the *input* and the *output* of f.

Tricolorings of w may be alternatively described as follows. We first write w with dashes between consecutive letters:

$$- w(1) - w(2) - \cdots - w(n) -$$

Enumerate the dashes from left to right by the numbers $0, 1, \ldots, n$. Then the function f as above can be seen as an assignment of a residue mod 3 to every dash. The conditions above mean that for any $A \in \mathcal{A}$, the coloring has the following form near the two entries of A in w: if $|A| \in \beta$, then it looks like

$$\cdots \overset{c}{—} A \overset{c}{—} \cdots \overset{c'}{—} A \overset{c''}{—} \cdots$$

with $c + c' + c'' = 0$ and if $|A| \in \beta - \alpha$, then it looks like

$$\cdots \overset{c'}{—} A \overset{c''}{—} \cdots \overset{c}{—} A \overset{c}{—} \cdots$$

with $c + c' + c'' = 0$. The input of the coloring is the residue assigned to the leftmost dash and the output is the residue assigned to the rightmost dash.

For example, the function assigning one and the same residue to all dashes is a coloring. It is called the *trivial coloring*.

Theorem 7.1. *For any $k, l \in \mathbb{Z}/3\mathbb{Z}$, the number of tricolorings of a nanoword w with input k and output l is a homotopy invariant of w.*

Note that this number may depend on k, l, and on the choice of β. We can easily compute this number for the empty nanoword (it has only one dash). If $k = l$, then this nanoword admits one coloring with input k and output l. If $k \neq l$, then there are no such colorings. By Theorem 7.1, the same is true for any contractible nanoword.

Now we give an example of a nanoword which admits a coloring with distinct input and output. Consider the nanoword

$$w = A_1 A_2 B A_3 A_1 B A_2 A_3,$$

where $|A_1| = |A_2| = |A_3| = a \in \alpha$, $|B| = b \in \alpha$ such that a and b lie in different orbits of τ. Set $\beta = \{a, \tau(a)\} \subset \alpha$. The nanoword w has the following non-trivial coloring with input 0 and output 1:

$$\overset{0}{—} A_1 \overset{0}{—} A_2 \overset{0}{—} B \overset{1}{—} A_3 \overset{1}{—} A_1 \overset{2}{—} B \overset{2}{—} A_2 \overset{1}{—} A_3 \overset{1}{—} .$$

By the remarks above, this nanoword is non-contractible.

7.2. **Module of a nanoword.** A related but stronger invariant of knots is the Alexander module. It can be computed from a knot diagram via an explicit presentation by generators and relations. We apply a similar idea to nanowords. First, we introduce the group

$$\Psi = \langle \{a, a_\bullet\}_{a \in \alpha} \,|\, a a_\bullet = a_\bullet a,\ a\tau(a) = 1,\ a_\bullet \tau(a)_\bullet = 1 \rangle.$$

We already considered groups Π and π given by similar presentations. In Ψ, each letter $a \in \alpha$ gives rise to two commuting generators a and $a_\bullet$. In the case of knot diagrams on surfaces, this phenomenon of doubling of the number of generators was already observed by Sawollek [Sa] who studied generalizations of the Alexander polynomial, see also [SW1].

Let $\Lambda = \mathbb{Z}\Psi$ be the integral group ring of Ψ. This ring will play the role of the ground ring for our modules.

Fix a set $\beta \subset \alpha$ such that $\tau(\beta) = \beta$. Consider a nanoword $w = (\mathcal{A}, w\colon \widehat{n} \to \mathcal{A})$ over α. We derive from w a $(n+1) \times n$ matrix over Λ whose rows are numerated by the dashes of w. Each letter $A \in \mathcal{A}$ gives rise to two rows. To write them down, set $a = |A| \in \alpha$ and assume that A appears in w for the first time at the i-th position and for the second time at the j-th position. If $a \in \beta$, then the two rows determined by A are

$$\begin{array}{ccccc} & \overset{i}{} & & \overset{j}{} & \\ \cdots & - \ A \ - & \cdots & - \ A \ - & \cdots \\ \cdots 0 & a \quad\ -1 & 0 \cdots 0 & 0 \quad 0 & 0 \cdots \\ \cdots 0 & 1 - aa_\bullet \quad 0 & 0 \cdots 0 & a_\bullet \quad -1 & 0 \cdots \end{array}$$

If $a \in \alpha - \beta$, then the two rows determined by A are

$$\begin{array}{ccccc} & \overset{i}{} & & \overset{j}{} & \\ \cdots & - \ A \ - & \cdots & - \ A \ - & \cdots \\ \cdots 0 & 0 \quad 0 & 0 \cdots 0 & a \quad\ -1 & 0 \cdots \\ \cdots 0 & a_\bullet \quad -1 & 0 \cdots 0 & 1 - aa_\bullet \quad 0 & 0 \cdots \end{array}$$

All unspecified entries of the rows are 0. The resulting $n \times (n+1)$ matrix over Λ determines a Λ-homomorphism $\psi\colon\ \Lambda^n \to \Lambda^{n+1}$ whose cokernel

$$K_\beta(w) = \Lambda^{n+1}/\psi(\Lambda^n).$$

is a Λ–module. This module has distinguished elements: the "input" v_- and the "output" v_+ represented by the leftmost dash and the rightmost dash, respectively.

Theorem 7.2. *The triple $(K_\beta(w), v_-, v_+)$ considered up to isomorphism is a homotopy invariant of w.*

One can derive further homotopy invariants from the triple $(K_\beta(w), v_-, v_+)$ or directly from the presentation matrix of $K_\beta(w)$ introduced above. For example, one may remove the first (or the last) column and consider the resulting $n \times n$ matrix over Λ. Then we can take its determinant over the ring Λ^{ab} obtained by abelianization of Λ. Following this line of thought and with a little more work, one obtains two "polynomial" homotopy invariants of w belonging to Λ^{ab} (see [Tu2] for details). They are denoted $\nabla^-_\beta(w)$ and $\nabla^+_\beta(w)$ and satisfy the following duality:

$$\nabla^+_\beta(w) = \overline{\nabla^-_{\alpha-\beta}(w^-)}.$$

The bar on the right-hand side is the ring involution on Λ^{ab} given by $\overline{a} = \tau(a)$ and $\overline{a_\bullet} = \tau(a)_\bullet$ for all $a \in \alpha$.

7.3. **The invariant λ.** We now focus on the case where $\beta = \alpha$. This will lead us to a homotopy invariant λ of nanowords taking values in the ring Λ. This invariant is a generalization of an invariant introduced by Silver and Williams [SW2] for curves.

Consider the Λ–module $K_\beta(w) = K_\alpha(w)$ associated above with a nanoword $w \colon \widehat{n} \to \mathcal{A}$. Let $x_0, x_1, \ldots, x_n$ be the generators of $K_\alpha(w)$ given by the dashes. Each letter $A \in \mathcal{A}$ gives rise to two relations

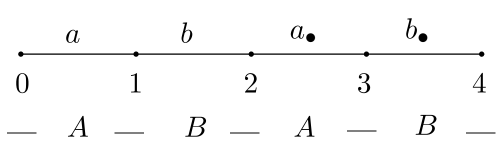
$$x_{i+1} = a x_i, \quad x_j = a_\bullet x_{j-1} + (1 - a a_\bullet) x_i,$$

where $a = |A| \in \alpha$ and $i = i_A < j = j_A$ are the elements of $w^{-1}(A)$. Each of these relations expresses a generator via the previous generators. Therefore $K_\alpha(w)$ is a rank one free Λ-module generated by the input $v_- = x_0$. The output $v_+ = x_n \in K_\alpha(w)$ has the form $v_+ = \lambda' v_-$ for a unique $\lambda' \in \Lambda$. Theorem 7.2 implies that $\lambda' = \lambda'(w)$ is a homotopy invariant of w. This invariant is a non-commutative polynomial. It admits an equivalent but more convenient version $\lambda(w)$ defined as follows. Consider the involutive anti-automorphism ι of Λ keeping fixed all the generators $\{a, a_\bullet\}$ of Λ. Thus, ι acts on monomials by reading them from right to left. For instance $\iota(aab_\bullet) = b_\bullet aa$. Set

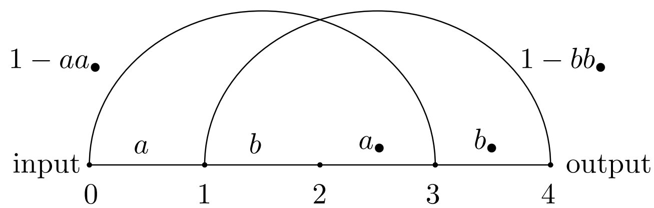
$$\lambda(w) = \iota(\lambda'(w)) \in \Lambda.$$

We describe a method allowing to compute $\lambda(w)$ and generalizing a method due to Silver and Williams [SW2] in the context of curves. We do it here for a few examples, the general method [Tu2] should be clear.

Example 7.3. Consider the nanoword $w = ABAB$, $|A| = a \in \alpha$, $|B| = b \in \alpha$. First draw the following graph:



Each vertex of this graph corresponds to a dash in w and each edge corresponds to a letter in w. Recall that every letter appears twice. The edge corresponding to the first (leftmost) appearance of A is labeled with a; the edge corresponding to the second (rightmost) appearance of A is labeled with $a_\bullet$. Connect the left vertex of the first edge with the right vertex of the second edge by an arc in the upper half-plane and label this arc with $1 - aa_\bullet \in \Lambda$. Do the same for the letter B replacing everywhere $a = |A|$ by $b = |B|$. The resulting picture is drawn on the next figure.



Consider all paths starting at the input and going to the output along the edges and arcs, always from left to right. We record the elements of Λ labeling the arcs and edges on the path and multiply them following the order determined by the path. The polynomial $\lambda(w)$ is obtained as the sum of the resulting elements of Λ over all paths. In this case there are three such paths:

(1) The path $0 - 1 - 2 - 3 - 4$ contributes $aba_\bullet b_\bullet$.
(2) The path $0 - 1 \frown 4$ contributes $a(1 - bb_\bullet)$.
(3) The path $0 \frown 3 - 4$ contributes $(1 - aa_\bullet)b$.

Then

$$\lambda(w) = aba_\bullet b_\bullet + a(1 - bb_\bullet) + (1 - aa_\bullet)b_\bullet.$$

The ring Λ has a natural grading as follows. Recall that

$$\Lambda = \mathbb{Z}[a, a_\bullet]_{a\in\alpha}/aa_\bullet = a_\bullet a,\ a\tau(a) = 1,\ a_\bullet\tau(a)_\bullet = 1.$$

The defining relations are homogeneous with respect to degrees mod 2. Therefore

$$\Lambda = \Lambda_{0,0} \oplus \Lambda_{0,1} \oplus \Lambda_{1,0} \oplus \Lambda_{1,1},$$

where $\Lambda_{i,j}$ is generated by monomials in which generators without bullets appear i times mod 2 and generators with bullets appear j times mod 2. Every $\lambda \in \Lambda$ expands uniquely as the sum

$$\lambda = \lambda_{0,0} + \lambda_{0,1} + \lambda_{1,0} + \lambda_{1,1},$$

where $\lambda_{i,j} \in \Lambda_{i,j}$ for all i, j. For $w = ABAB$, this expansion of $\lambda(w)$ gives:

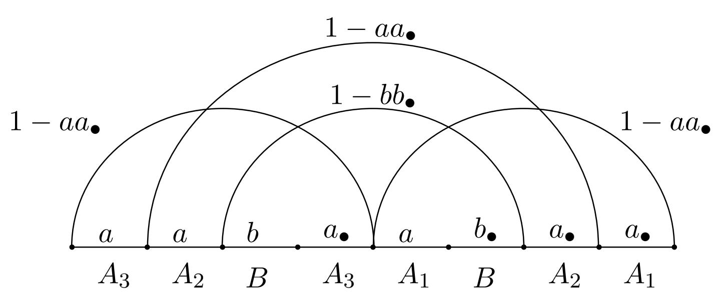
$$\begin{aligned}\lambda_{0,0}(w) &= aba_\bullet b_\bullet,\\ \lambda_{0,1}(w) &= -abb_\bullet + b_\bullet,\\ \lambda_{1,0}(w) &= a - aa_\bullet b_\bullet,\\ \lambda_{1,1}(w) &= 0.\end{aligned}$$

These computations allow us to give another proof of the fact that w is contractible if and only if $a = \tau(b)$. Indeed, if w is contractible, then $\lambda(w) = 1$ and hence $\lambda_{0,1}(w) = 0$. This implies that $abb_\bullet = b_\bullet$. Hence $ab = 1$ and $a = \tau(b)$.

Example 7.4. We apply λ to the word $ababa$ with $a = \tau(a) \neq b = \tau(b)$. As we saw above, the corresponding α–pairing gives no information about the homotopy properties of w. By definition, $\lambda(w) = \lambda(w^d)$. The desingularization of w is the nanoword

$$w^d = A_3A_2BA_3A_1BA_2A_1,\ |A_1| = |A_2| = |A_3| = a,\ |B| = b.$$

To compute $\lambda(w)$, we draw the following graph:



Then

$$\begin{aligned}\lambda(w) = \quad & (1 - aa_\bullet)^2 && \frown\frown\\ & +(1 - aa_\bullet)ab_\bullet a_\bullet^2 && \frown - - - -\\ & +a(1 - aa_\bullet)a_\bullet && - \frown -\\ & +a^2(1 - bb_\bullet)a_\bullet^2 && - - \frown - -\\ & +a^2ba_\bullet(1 - aa_\bullet) && - - - - \frown\\ & +a^2baa_\bullet b_\bullet a_\bullet^2. && - - - - - - - -\end{aligned}$$

The assumptions $a = \tau(a)$ and $b = \tau(b)$ imply that $a^2 = b^2 = a_\bullet^2 = b_\bullet^2 = 1$. After simplification, we obtain that

$$\lambda_{0,0}(w) = 2 - ba - a_\bullet b_\bullet + baa_\bullet b_\bullet.$$

If $\lambda_{0,0}(w) = 1$, then one of the two elements ba and $a_\bullet b_\bullet$ of the group Ψ must be equal to 1. This is possible only if $a = \tau(b) = b$, which contradicts the assumptions. So, $\lambda_{0,0}(w) \neq 1$ and w is non-contractible.

Example 7.5. Consider the nanowords

$$w_1 = ABACBC,\ |A| = |C| = a,\ |B| = \tau(a),$$
$$w_2 = ACAC,\ |A| = |C| = a,$$

where $a \in \alpha$ satisfies $\tau(a) \neq a$. These two nanowords are not distinguished by λ. In fact, all the techniques described so far fail to distinguish these nanowords up to homotopy. This can be done using the methods introduced in the next section.

8. α–KEIS AND WORDS

8.1. α**–keis.** Keis were introduced in 1942 by a Japanese mathematician, M. Takasaki, see S. Kamada [Kam] for a comprehensive survey of keis, their generalizations, and connections with knot theory. A *kei* is a set X with multiplication $*$ which satisfies a few axioms, the main axiom being

$$(x * y) * z = (x * z) * (y * z)$$

for all $x, y, z \in X$. One may think of $x * y$ as of a kind of conjugation of x by y.

To produce homotopy invariants of words, we introduce a notion of an α–kei, where α is a set with involution τ. An *α–kei* is a non-empty set X with maps

$$X \to X,\ x \mapsto ax \quad \text{and} \quad X \times X \to X,\ (x, y) \mapsto x *_a y \in X$$

numerated by $a \in \alpha$ such that the following axioms are satisfied:

(1) $ax *_a x = x$,
(2) $a(x *_a y) = ax *_a ay$,
(3) $(x *_a y) *_a z = (x *_a az) *_a (y *_a z)$,
(4) $a\tau(a)x = x$,
(5) $(x *_a y) *_{\tau(a)} ay = x$,

for all $a \in \alpha$ and $x, y, z \in X$. Arbitrary α–keis can be presented by generators and relations as groups in group theory.

Example 8.1. Recall the non-commutative ring

$$\Lambda = \mathbb{Z}[a, a_\bullet]_{a\in\alpha}/aa_\bullet = a_\bullet a,\ a\tau(a) = 1,\ a_\bullet\tau(a)_\bullet = 1.$$

Any left Λ–module X becomes an α–kei with kei operations $x \mapsto ax$ and

$$x *_a y = a_\bullet x + (1 - a_\bullet a)y.$$

The α–keis obtained by this construction are said to be *abelian*.

8.2. **α–keis of nanowords.** The theory of keis can be applied to produce homotopy invariants of nanowords. Fix a set $\beta \subset \alpha$ such that $\tau(\beta) = \beta$. For any nanoword $(\mathcal{A}, w\colon \widehat{n} \to \mathcal{A})$ over α, we define an α–kei $\mathcal{K}_\beta(w)$. It is generated by $n+1$ symbols $X_0, X_1, \ldots, X_n$ satisfying the following n defining relations. Each letter $A \in \mathcal{A}$ gives two relations. To write them down, assume that A appears in w for the first time at the i-th position and for the second time at the j-th position, where $i < j$. If $a = |A| \in \beta$, then the relations are

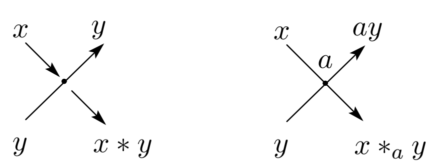
$$X_i = aX_{i-1}, \quad X_j = X_{j-1} *_a X_{i-1}.$$

If $a = |A| \in \alpha - \beta$, then the relations are

$$X_i = X_{i-1} *_a X_{j-1}, \quad X_j = aX_{j-1}.$$

The elements $V_- = X_0 \in \mathcal{K}_\beta(w)$ and $V_+ = X_n \in \mathcal{K}_\beta(w)$ are called the *input* and the *output*, respectively.

The idea behind these formulas comes from knot theory. In knot theory, every knot diagram gives rise to a so-called quandle. Quandles are generalizations of keis and also have only one operation, the binary operation $*$. The quandle associated with a knot diagram is determined by generators, associated with the arcs of the diagrams, and relations, associated with the crossings, cf. the picture on the left hand side of the following figure.



In the setting of nanowords the situation is somewhat different. First, each crossing is labeled by a letter, $a \in \alpha$, which allows us to involve the operation $y \mapsto ay$ absent for knots. The binary operation $*_a$ also depends on a. Also, the two incoming branches are ordered. This leads us to the defining relations as above, whose geometric interpretation is shown on the right hand side of the figure.

Theorem 8.2. *The triple $(\mathcal{K}_\beta(w), V_-, V_+)$, considered up to isomorphism, is a homotopy invariant of w.*

The Λ–module $K_\beta(w)$, viewed as an α–kei, can be computed from $\mathcal{K}_\beta(w)$. Namely, there is a homomorphism of α–keis $\mathcal{K}_\beta(w) \to K_\beta(w)$ such that for any homomorphism from $\mathcal{K}_\beta(w)$ to an abelian α–kei X, the following diagram is commutative:

$$\begin{array}{ccc} \mathcal{K}_\beta(w) & \longrightarrow & K_\beta(w) \\ & \searrow & \downarrow \\ & & X. \end{array}$$

8.3. **Characteristic sequences.** Consider in more detail the case $\beta = \alpha$. Looking at the defining relations, we easily observe that $\mathcal{K}_\beta(w) = \mathcal{K}_\alpha(w)$ is a free α–kei generated by the input V_-. The output $V_+ \in \mathcal{K}_\alpha(w)$ is a homotopy invariant of w. The structure of free α–keis is poorly understood, which prevents us from deriving further invariants of w from V_+. We focus on a special case where more information is available.

Suppose that the involution $\tau\colon\ \alpha \to \alpha$ is fixed-point-free, that is $\tau(a) \neq a$ for all $a \in \alpha$. Fix a set $\alpha_+ \subset \alpha$ meeting every orbit of τ in exactly one element. Thus,

$$\alpha = \alpha_+ \cup \tau(\alpha_+),\ \ \alpha_+ \cap \tau(\alpha_+) = \emptyset.$$

Recall the group Ψ introduced in Section 7.2. We show how to derive from any nanoword w over α a finite sequence $(\varepsilon_1\psi_1, \varepsilon_2\psi_2, \ldots, \varepsilon_m\psi_m)$ with $m \geq 0$, $\psi_1, \ldots, \psi_m \in \Psi$, and $\varepsilon_1, \ldots, \varepsilon_m \in \{\pm 1\}$. This sequence is a homotopy invariant of w (possibly depending on α_+). It determines $\lambda(w)$ by

$$\lambda(w) = \sum_{i=1}^{m} \varepsilon_i \psi_i \in \Lambda = \mathbb{Z}\Psi.$$

In the setting of curves, this sequence was introduced by Silver and Williams [SW2].

We first define an α–kei F as follows. Let F be the free group generated by the group Ψ, viewed as a set. Each element $\psi \in \Psi$ gives rise to a generator of F, denoted $\underline{\psi}$. In particular, the unit $1 \in \Psi$ gives rise to a generator $\underline{1} \in F$ which is by no means the unit of F. A typical element of F has the form

$$(\underline{\psi_1})^{\varepsilon_1} (\underline{\psi_2})^{\varepsilon_2} \cdots (\underline{\psi_m})^{\varepsilon_m}$$

where $m \geq 0$, $\psi_1, \psi_2, \ldots, \psi_m \in \Psi$, and $\varepsilon_1, \ldots, \varepsilon_m \in \{\pm 1\}$. Such an element is the unit of F if either $m = 0$ or it can be reduced to the case $m = 0$ by applying the relations $\underline{\psi}(\underline{\psi})^{-1} = (\underline{\psi})^{-1}\underline{\psi} = 1$. The left action of Ψ on itself extends to a group action of Ψ on F by group automorphisms. The generators $a, a_\bullet \in \Psi$ act on F by

$$a((\underline{\psi_1})^{\varepsilon_1}(\underline{\psi_2})^{\varepsilon_2}\cdots(\underline{\psi_m})^{\varepsilon_m}) = (\underline{a\psi_1})^{\varepsilon_1}(\underline{a\psi_2})^{\varepsilon_2}\cdots(\underline{a\psi_m})^{\varepsilon_m},$$
$$a_\bullet((\underline{\psi_1})^{\varepsilon_1}(\underline{\psi_2})^{\varepsilon_2}\cdots(\underline{\psi_m})^{\varepsilon_m}) = (\underline{a_\bullet\psi_1})^{\varepsilon_1}(\underline{a_\bullet\psi_2})^{\varepsilon_2}\cdots(\underline{a_\bullet\psi_m})^{\varepsilon_m}.$$

This defines in particular the mapping $F \to F$, $x \mapsto ax$ for all $a \in \alpha$. The binary operation $x *_a y$ for $x, y \in F$ is defined by

$$x *_a y = y(a_\bullet x)(a_\bullet a y)^{-1} \in F,$$

if $a \in \alpha_+$ and

$$x *_a y = (\tau(a)_\bullet^{-1}\tau(a)^{-1}y)^{-1}(a_\bullet^{-1}x)y \in F,$$

if $a \in \alpha - \alpha_+$. These operations make F into an α–kei.

Recall that starting with a nanoword w, we obtained a homotopy invariant element V_+ of the free α–kei $\mathcal{K}_\alpha(w)$ on one generator V_-. Since $\mathcal{K}_\alpha(w)$ is free, there is a unique α–kei homomorphism $f : \mathcal{K}_\alpha(w) \to F$ such that $f(V_-) = \underline{1} \in F$. Then $f(V_+) \in F$ is a homotopy invariant of w. We can expand

$$f(y) = (\underline{\psi_1})^{\varepsilon_1}\cdots(\underline{\psi_m})^{\varepsilon_m} \in F,$$

where $\psi_1, \ldots, \psi_m \in \Psi$ and $\varepsilon_1, \ldots, \varepsilon_m \in \{\pm 1\}$. The resulting sequence $(\varepsilon_1\psi_1, \ldots, \varepsilon_m\psi_m)$ is well-defined up to insertion or deletion of consecutive pairs $(+\psi, -\psi)$ and $(-\psi, +\psi)$. Deleting all such pairs, we obtain a uniquely defined sequence $(\varepsilon_1\psi_1, \ldots, \varepsilon_{m'}\psi_{m'})$ with $m' \leq m$ which is a homotopy invariant of w. This is the *characteristic sequence* of w.

8.4. **Examples.** 1. Pick $a, b \in \alpha_+$ and consider the nanoword $w = ABAB$ with $|A| = a$ and $|B| = b$. It is easy to compute from the relations that $V_+ = (baV_- *_a V_-) *_a aV_-$. The characteristic sequence of w is computed to be

$$(a, b_\bullet, b_\bullet a_\bullet ba, -b_\bullet a_\bullet a, -b_\bullet ba).$$

In particular, if $a = b$, then this sequence is $(a, a_\bullet, a_\bullet^2 a^2, -a_\bullet^2 a, -a_\bullet aa)$.

2. Consider the nanoword $w_1 = ABACBC$, $|A| = |C| = a$, $|B| = \tau(a) \neq a$. Its characteristic sequence (determined by any $\alpha_+ \subset \alpha$ as above such that $a \in \alpha_+$) is:

$$(1, a_\bullet, -aa_\bullet, -1, a, aa_\bullet, -a^2a_\bullet, aa_\bullet, a^2a_\bullet^2, -aa_\bullet^2, -aa_\bullet).$$

Comparing with the previous example (for $a = b$), we obtain that w_1 is not homotopic to the nanoword $w_2 = ACAC$ with $|A| = |C| = a$. This result was claimed at the end of Section 7.

3. One might think that such a powerful invariant as the characteristic sequence should distinguish arbitrary non-homotopic nanowords. However this is not true, as shows the following example. Pick four letters $a, b, c, d \in \alpha$ (possibly coinciding) and consider the nanoword

$$w = ABCDCDAB, \ \ |A| = a, \ |B| = b, \ |C| = c, \ |D| = d.$$

An inspection shows that if $a \neq \tau(b)$ and $c \neq \tau(d)$, then the α-pairing of w is primitive. Then $\|w\| = 4$ and w is non-contractible. However, a direct computation shows that for $a, b \in \alpha_+$ and $c = \tau(b), d = \tau(a)$, the characteristic sequence of w is the same as the one of the empty nanoword. Both consist of a single term $1 \in \Psi$.

9. Open questions and further directions

Question 9.1. Classify nanowords of length ≤ 10 up to homotopy.

In [Tu2] we give a homotopy classification of nanowords up to length 6. The next step is to handle the nanowords of length 8. Does one need new homotopy invariants already for length 8 ?

Question 9.2. Classify words of length ≤ 7 up to homotopy.

In [Tu2] we give a homotopy classification of words up to length 5. One may try to classify words by first classifying nanowords. However, short words may desingularize into quite long nanowords. For example, the word *aababb*, desingularizes into a nanoword of length 12. Still, a classification of words of length ≤ 7 does not look unrealistic because they desingularize into a quite particular set of nanowords.

Question 9.3. What (primitive) α–pairings can be realized as α–pairings of nanowords?

Question 9.4. What polynomials $\lambda \in \Lambda$ arise from nanowords?

There are some simple known conditions, see [Tu2]. All new conditions are welcome.

Question 9.5. Is it true that all nanowords over the alphabet consisting of a single element are contractible ?

At the moment, nothing contradicts the conjecture that the answer is yes.

Question 9.6. Give a normal form for elements of a free α–kei on one generator.

Such a normal form (or at least an algorithm to distinguish elements of this α–kei) would help to distinguish words up to homotopy.

One further direction is the study of cobordisms of words. Cobordism is an equivalence relation generated by surgery on words which consists in deleting or inserting symmetric subwords or subphrases. There are difficult problems concerning the classification of words up to cobordism. This is studied in [Tu4].

Another interesting direction is a study of higher dimensional words over an alphabet α. Knot theory and other topological ideas used above generalize to higher dimensions. What can be said about similar generalizations of words? From the topological perspective, an n–dimensional nanoword is an immersion of a connected n-dimensional manifold into an $(n+1)$–dimensional manifold. The double points of the immersion split as a union of connected $(n-2)$-dimensional manifolds labeled with letters of α. The case $n = 1$ is treated in the present paper. The next case $n = 2$ is quite mysterious. What are the appropriate analogues of the homotopy moves for $n = 2$? Although a study of high-dimensional words is tempting, it is hard to imagine intelligent beings communicating with such words.

References

[CE] G. Cairns and D. M. Elton, *The planarity problem for signed Gauss words*, J. Knot Theory Ramifications 2 (1993), 359–367.

[CKS] J. S. Carter, S. Kamada, M. Saito, *Stable equivalence of knots on surfaces and virtual knot cobordisms*. J. Knot Theory Ramifications 11 (2002), 311–322.

[CW] N. Chaves and C. Weber, *Plombages de rubans et problème des mots de Gauss*, Exposition. Math. 12 (1994), 53–77 and 124.

[CR] H. Crapo and P. Rosenstiehl, *On lacets and their manifolds*, Discrete Math. 233 (2001), 299–320.

[DT] C. H. Dowker and M.B. Thistlethwaite, *Classification of knot projections*, Topology Appl. 16 (1983), 19–31.

[Ga] C. F. Gauss, Werke, Vol. VIII, Teubner, Leipzig, 1900, pp. 272, 282–286.

[GPV] M. Goussarov, M. Polyak, and O. Viro, *Finite-type invariants of classical and virtual knots*. Topology 39 (2000), 1045–1068.

[Kam] S. Kamada, *Knot invariants derived from quandles and racks*. Invariants of knots and 3-manifolds (Kyoto, 2001), 103–117 (electronic), Geom. Topol. Monogr., 4, Geom. Topol. Publ., Coventry, 2002. 57M27

[KK] N. Kamada and S. Kamada, *Abstract link diagrams and virtual knots*. J. Knot Theory Ramifications 9 (2000), 93–106.

[Ka] L. Kauffman, *Virtual knot theory*, European J. Combin. 20 (1999), 663–690.

[LM] L. Lovász and M. L. Marx, *A forbidden substructure characterization of Gauss codes*. Acta Sci. Math. (Szeged) 38 (1976), 115–119.

[Ma] M. L. Marx, *The Gauss realizability problem*, Proc. Amer. Math. Soc. 22 (1969), 610–613.

[Ro] P. Rosenstiehl, *Solution algébrique du problème de Gauss sur la permutation des points d'intersection d'une ou plusieurs courbes fermées du plan*, C. R. Acad. Sci. Paris Sér. A-B 283 (1976), A551–A553.

[Sa] J. Sawollek, *On Alexander-Conway polynomials for virtual knots and links*, math.GT/9912173

[SW1] D. Silver and S. Williams, *Polynomial invariants of virtual links*, J. Knot Theory Ramifications 12 (2003), 987–1000.

[SW2] D. Silver and S. Williams, *An invariant for open virtual strings*, J. Knot Theory Ramifications 15 (2006), 143–152.

[Tu1] V. Turaev, *Virtual strings*, Ann. Inst. Fourier 54 (2004), 2455–2525.

[Tu2] V. Turaev, *Topology of words*, math.CO/0503683.

[Tu3] V. Turaev, *Knots and words*, math.CO/math.GT/0506390.

[Tu4] V. Turaev, *Cobordisms of words*, math.CO/0511513.

IRMA, Université Louis Pasteur - C.N.R.S.,
7 rue René Descartes
F-67084 Strasbourg
France
e-mail: turaev@math.u-strasbg.fr