

A NEW PROOF OF MARKOV'S BRAID THEOREM

PAWEŁ TRACZYK

Institute of Mathematics, Warsaw University

ul. Banacha 2, 02-097 Warszawa, Poland

E-mail: traczyk@mimuw.edu.pl

The purpose of this paper is to introduce a new proof of Markov's braid theorem, in terms of Seifert circles and Reidemeister moves. This means that the proof will be of combinatorial and essentially 2-dimensional nature. One characteristic feature of our approach is that nowhere in the proof will we use or refer to the braid axis. This allows for greater flexibility in various transformations of the diagrams considered. Other proofs of Markov's theorem can be found in [2], [3], [4] and [5].

As in Vogel's paper [6], "closed braid" will be understood to mean a special kind of diagram of an oriented link: the diagram lives on a 2-dimensional sphere and is supposed to have Seifert circles that are nested in one another forming a single chain of concentric circles with compatible orientations. Let us recall that every diagram may be transformed into a closed braid (this is called Alexander's theorem, see [1]). One particularly nice way to do it is by using reducing moves of Vogel, see [6]. Such a move is simply a type II Reidemeister move that is performed on two arcs belonging to different Seifert circles with orientations as in figure 1. Reduction moves will be the main tool in our proof of Markov's theorem.

THEOREM 1 (Markov 1935). *If B and C are closed braids representing the same isotopy class of oriented links, then it is possible to transform B into C by a sequence of braid isotopies and Markov moves (a Markov sequence).*

Proof. We will use notation $h(D)$ (the height of D) to denote a parameter measuring by how much a diagram D fails to be a closed braid. Our function $h(D)$ is a modification of Vogel's function χ . We say that two Seifert circles are incompatible if they have the same orientation with respect to any orientation of the annulus they bound (it should be stressed that this is defined for *any* pair of Seifert circles, we do not assume that the

1991 *Mathematics Subject Classification*: Primary 57M25.

Partially supported by KBN grant No 211399101.

The paper is in final form and no version of it will be published elsewhere.



Fig. 1

two circles are adjacent). We define $h(D)$ as the number of incompatible pairs of Seifert circles. It is easy to see that D is a closed braid if and only if $h(D) = 0$ and that a reducing operation decreases h by precisely one. This solves one of the problems considered in [6]:

THEOREM 2. *The length of any sequence of reducing operations transforming a diagram D into a closed braid is exactly $h(D)$.*

The reducing operations are tightly bound with what we call *defect areas*. A diagram (or rather the underlying projection) divides the projection sphere into several areas. The boundary of every area consists of several arcs, each of them being a part of some Seifert circle. We say that an area is a *defect area* if there exist two incompatible Seifert circles contributing to its boundary. This is equivalent to saying that the area considered supports a reducing move. Let us record the following obvious observation.

PROPOSITION 3. *Smoothing a crossing (or any number of crossings) does not increase the number of defect areas.*

This observation may be used to prove Alexander’s theorem, because it proves that a diagram that is not a closed braid does admit a reducing operation. It will also be useful in our proof of Markov’s theorem.

Another characteristic of a reducing operation is its *reduction curve*—this is a curve connecting two points on the boundary of the defect area such that the operation is realized by dragging the two arcs toward each other along this curve until finally one arc is slid below the other (which arc goes under could be indicated by marking an orientation on the reduction curve, pointing in the direction of the arc going under, see figure 2).

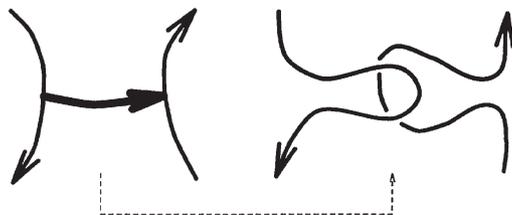


Fig. 2

To prove the Markov theorem let us start with an arbitrary sequence of Reidemeister’s moves transforming B into C . We will pursue a natural strategy of modifying the given Reidemeister sequence gradually until it becomes a Markov sequence. In the first step we observe that the sequence considered may be assumed to contain only moves of the following four types (and their inverses).

1. \mathcal{A} = Reidemeister move of type I (adding a kink),
2. \mathcal{B} = braid-like Reidemeister move of type II,
3. \mathcal{C} = braid-like Reidemeister move of type III,
4. \mathcal{R} = reducing move.

(By a braid-like move is meant a Reidemeister move that *locally* and *up to deformation* looks like a braid isotopy. If a braid-like move is applied to a closed braid, then this move *is* actually a braid isotopy.)

LEMMA 4. *Non-braid Reidemeister moves of type II acting on one Seifert circle and non-braid Reidemeister moves of type III may be obtained as compositions of the 4 moves listed above (and their inverses).*

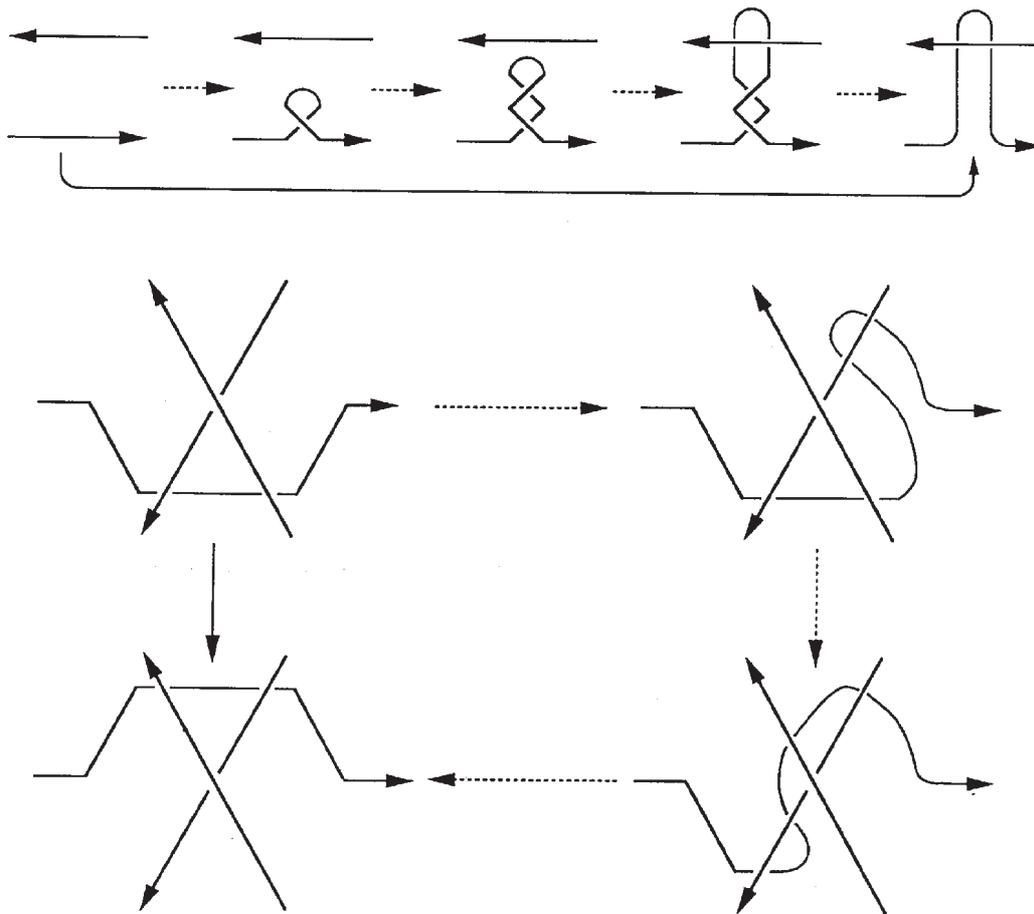


Fig. 3

Proof. See figure 3.

COROLLARY. *If B and D are two diagrams of the same link, then it is possible to transform B into D by a sequence of operations 1–4 and their inverses.*

Proof. We can use directly, or emulate, all Reidemeister moves.

In the second step we further restrict the usage of operations of types \mathcal{A} , \mathcal{B} and \mathcal{C} .

LEMMA 5. *If B and D are two diagrams of the same link, then it is possible to transform B into D by a sequence of operations 1–4 and their inverses in such a way that operations of type \mathcal{A} , \mathcal{B} and \mathcal{C} (and their inverses) appear only as braid isotopies or Markov moves.*

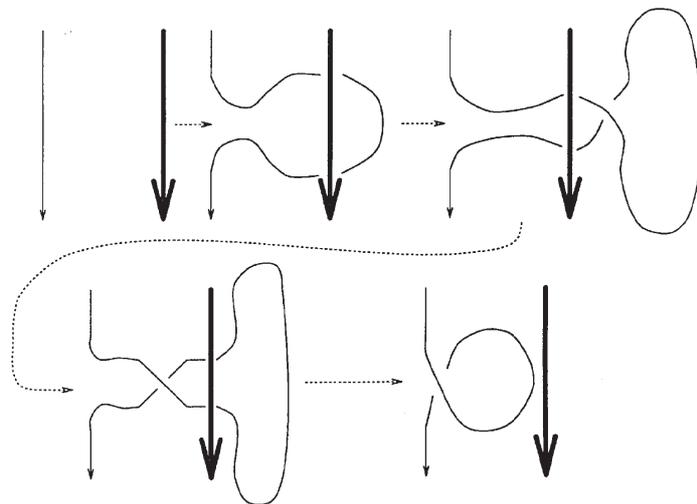


Fig. 4

Proof. Let t be an operation of type \mathcal{A} , \mathcal{B} or \mathcal{C} on a diagram D such that $h(D) > 0$. It is then possible to find a reducing operation r that is performed away from the area at which t changes the diagram. We can replace $D \xrightarrow{h} h(D)$ with a three term sequence: $D \xrightarrow{r} rD \xrightarrow{t} trD \xrightarrow{r^{-1}} r^{-1}trD = tD$ (notice that the operation r^{-1} applied to the diagram trD is of type \mathcal{R}^{-1}). The operation t is now performed at lower height and gradually we can go all the way down to height zero. If t is of type \mathcal{B} or \mathcal{C} , then the operation considered is a braid isotopy. If t is of type \mathcal{A} we may have to consider one additional case, namely that of an operation t of type \mathcal{A} starting from height zero and ending up at height one or more. In other words we start from a closed braid, and a Reidemeister I move is performed which is not a Markov move. Then figure 4 explains how t may be replaced with a sequence of braid isotopies followed by a single Markov move followed by a sequence of braid isotopies. Here, for the first time, we use a certain useful convention. The two lines in the picture are not necessarily single strings, they represent certain bands of parallel strings. In fact in this figure the thin line *is* a single string, while the fat line has a certain multiplicity. Thus the first transformation of the initial diagram is a sequence of several Reidemeister II moves, the second is a Markov move and the third is again a sequence of several Reidemeister III moves. This completes the proof of Lemma 5.

Now, in the third step of the proof of Markov's theorem, it is sufficient to consider sequences containing only operations of type \mathcal{R} and \mathcal{R}^{-1} . We will use a convention of denoting our operations with arrows pointing in the direction of lower height. Thus $X \longleftarrow Y \longrightarrow Z$ means that first X is transformed into Y by an inverse of a reducing operation, and then Y is transformed into Z by a reducing operation. It is clear that in such a case there is a local height maximum at Y . Our strategy is to gradually cancel all such maxima. We want to find a different sequence of operations transforming X into Z , possibly going as far down as height 0, and possibly involving some Markov moves and braid isotopies (in addition to \mathcal{R} and \mathcal{R}^{-1} operations), but never reaching the height of $h(Y)$. It will be convenient to have some classification of such maxima:

A pair of reducing operations $X \xleftarrow{s} Y \xrightarrow{r} Z$ on a diagram Y will be called a \wedge -pair of height $h(Y)$, the shape of \wedge referring to the height maximum occurring at Y .

A \wedge -pair such that the two operations have disjoint reducing curves will be called a d -pair (d for disjoint).

A d -pair acting on more than two Seifert circles will be called a c -pair (c for compatible).

In a d -pair the two Reidemeister II moves do not get in each other's way: they can be performed consecutively in any order, and the result is the same diagram, say Y' . But when the first move is performed, the structure of Seifert circles may be disturbed in such a way that the second Reidemeister move will no longer be a reducing operation. This happens when both operations are performed on the same pair of Seifert circles. However, this is precluded by the condition defining a c -pair. Therefore, if $X \xleftarrow{r} Y \xrightarrow{s} Z$ is a c -pair then r is an \mathcal{R} operation on $s(Y) = Z$ and s is an \mathcal{R} operation on $r(Y) = Y$, again both resulting in the same diagram Y' . Thus $X \xrightarrow{s} Y' \xleftarrow{r} Z$ is what could be called a \vee -pair.

We can now describe the strategy (of cancelling height maxima) more precisely. Given a \wedge -pair we will replace it with a sequence of c -pairs and d -pairs, *avoiding d -pairs as much as possible*. Now, we will replace c -pairs with \vee -pairs—the way to do it is described above. As for remaining d -pairs, these will be replaced with sequences built of reducing operation and Markov sequences, more precisely $X \longleftarrow Y \longrightarrow Z$ will be replaced with $X \longrightarrow \dots \longrightarrow X' \longleftrightarrow Z' \longleftarrow \dots \longleftarrow Z$, where X' and Y' are closed braids and $X' \longleftrightarrow Y'$ is a Markov sequence.

The first step in the above described strategy (of cancelling height maxima) is to replace \wedge -pairs with certain sequences of simpler pairs of operations. In order to do this we will repeatedly replace \wedge -pairs of the form $X \xleftarrow{s} Y \xrightarrow{r} Z$ with certain $X \longleftarrow Y \xrightarrow{r} Y' \xleftarrow{r} Y \longrightarrow Z$ quadruples (here r is a suitably chosen reduction on Y). We will say that the operation r was inserted at Y . Of course the inserted operation r must be chosen properly so that the two \wedge -pairs obtained in the process ($X \longleftarrow Y \longrightarrow Y'$ and $Y' \longleftarrow Y \longrightarrow Z$) are in some way simpler than the original pair. We will use the following two reductions.

1. Reduction to \wedge -pairs with intersection number (= the number of crossings of the reduction curves) at most one: Suppose $X \xleftarrow{s} Y \xrightarrow{r} Z$ is a \wedge -pair with intersection number $n \geq 2$. It is easy to find a reduction curve having no intersections with one curve,

and less than n with the second. Inserting the operation defined by this curve at Y we obtain two \wedge -pairs with smaller intersection numbers. We can repeat this procedure of insertion until all the intersection numbers are smaller than 2.

2. Reduction to d -pairs: Let us consider a \wedge -pair $X \xleftarrow{s} Y \xrightarrow{r} Z$ with intersection number $n = 1$. This time we want to find a reduction curve for the diagram Y disjoint from *both* curves, then we can apply insertion as in 1. The only difference between 1. and 2. is that now it is not that obvious how to find the required curve. Assume that there is no such curve. It is then clear that the diagram Y has only one defect area—the one supporting the two operations in our \wedge -pair. Also, it follows that the two operations act on four different Seifert circles—otherwise we could easily find the required reduction curve as shown in figure 5. Now, let us smooth all crossings of Y , except those between the four Seifert circles involved in our \wedge -pair. By Proposition 3 the number of defect areas in the modified diagram is still one. Let us remove all the Seifert circles, except the four, from the modified diagram. It is clear that the resulting diagram is still an example of one in which the required curve cannot be found. However, by now we have an almost complete description of the diagram. It must look as in figure 6 (the boxes contain certain

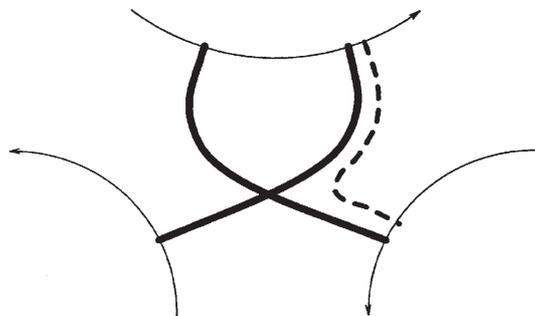


Fig. 5

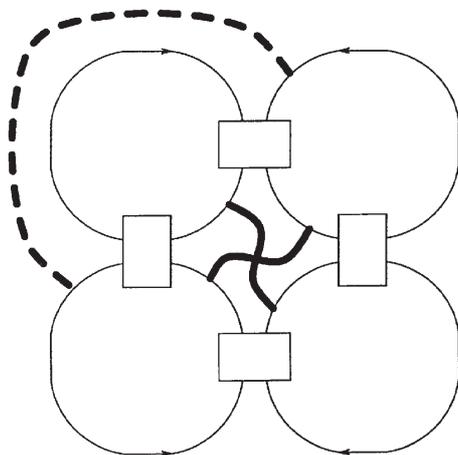


Fig. 6

two-string braids), and this is a contradiction—the diagram shown in figure 6 *does* admit a reduction curve with required properties.

As a result of 1. and 2. we can replace an arbitrary \wedge -pair with a sequence of c - and d -pairs. The c -pairs, are reduced simply by reversing the order in which the two operations appear, this replaces a \wedge -pair with a \vee -pair thus killing one maximum (and possibly producing one or two new maxima, but on a lower level). The problem is what to do with d -pairs. First of all, when describing our strategy (of cancelling \wedge -pairs) we have declared avoiding d -pairs as much as possible. Therefore we will make one last attempt to replace a d -pair with two c -pairs by suitable insertion—a method we have already used several times. The only problem is to find a reducing operation compatible with both operations of the d -pair considered. We do that whenever possible, and what

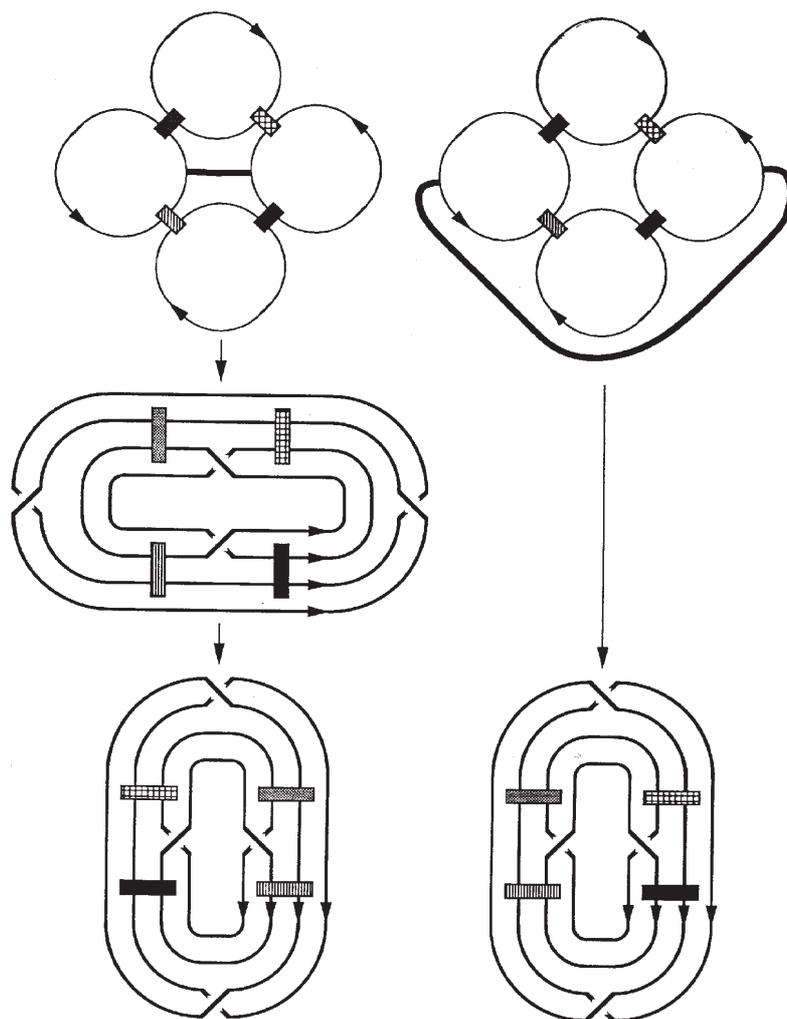


Fig. 7

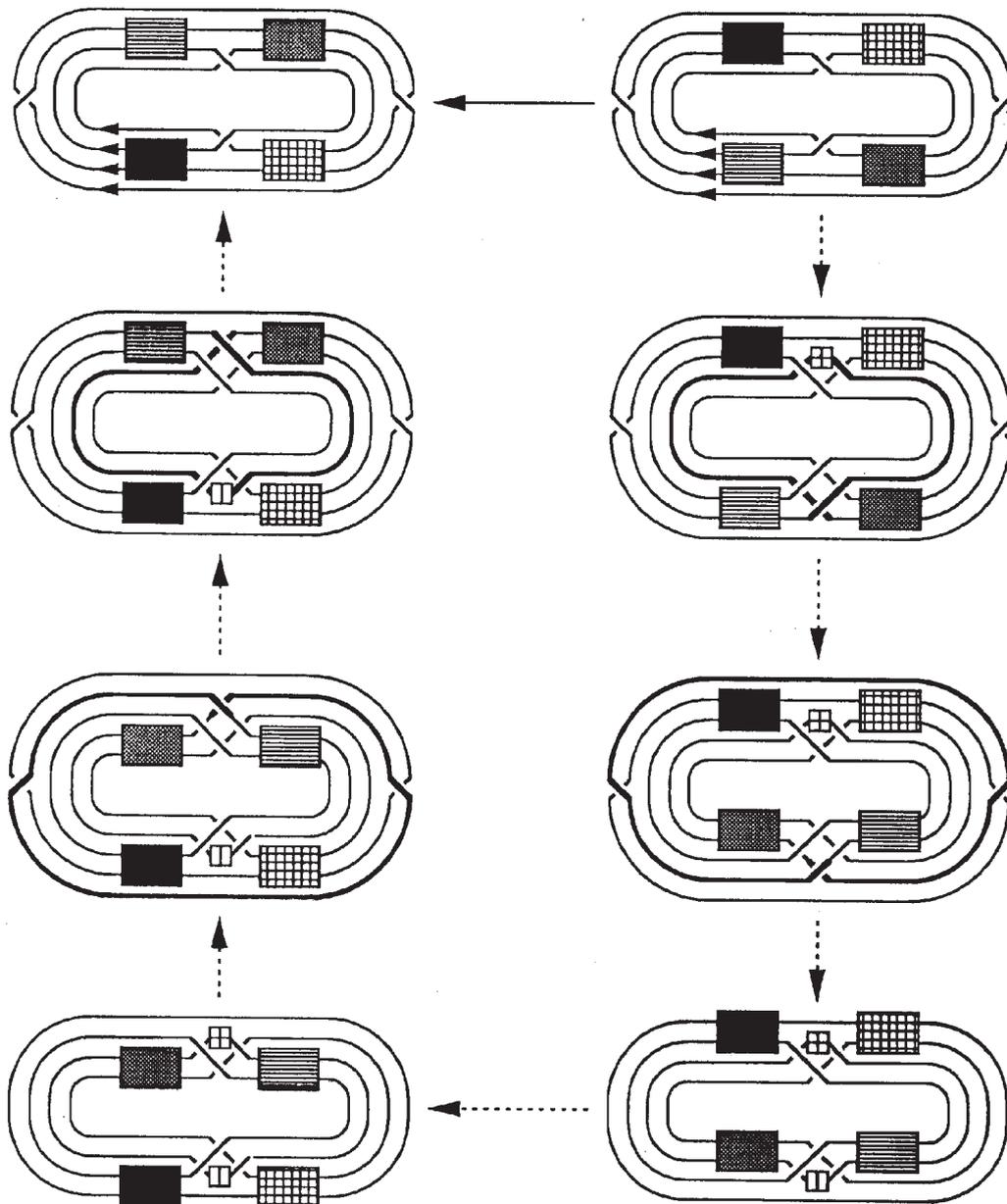


Fig. 8

remains is to identify those cases when it is not possible. This is the final special case in our strategy of cancelling \wedge -pairs. It leads to the situation depicted in figure 7 and this is solved in figure 8. The two figures will be explained later. The first problem is to give some characterization of those special d -pairs that cannot be reduced to c -pairs by insertion. We need some preparations for this description.

We say that two Seifert circles are equivalent if they are compatible and bound an annulus that does not contain any other Seifert circles. This is an equivalence relation. Equivalence classes of this relation are bands of parallel Seifert circles. If all crossings between different bands are smoothed, then what results is a split sum of several closed braids. We claim that in the special diagrams now under consideration there are at most four equivalence classes of Seifert circles. This implies that such a diagram must conform to the pattern shown in figure 7—the circles show equivalence classes of Seifert circles and the boxes are supposed to contain braids on suitable numbers of strings. In other words, each circle denotes a band of n parallel strings (we say that n is the multiplicity of the band) and each box contains a braid on the number of strings equal to the sum of multiplicities of the two bands involved. It should be stressed that while the bands containing the Seifert circles considered are necessarily there, the multiplicities of the other two bands *may* be equal to zero.

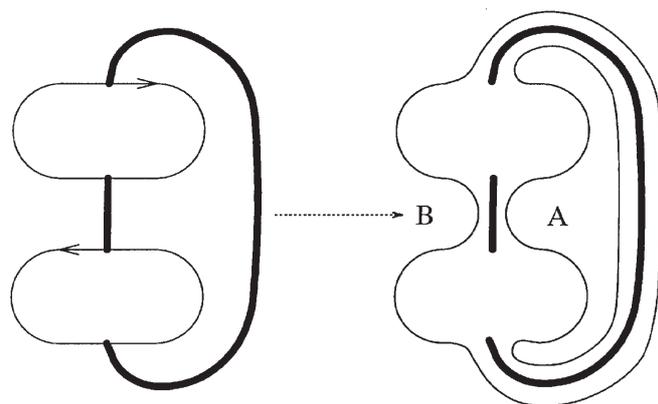


Fig. 9

To prove the claim let us modify the diagram considered in a manner shown in figure 9. There is no defect area in either A or B part of the modified diagram—any reduction curve in such an area would also be a reduction for the original diagram, one with different pair of Seifert circles. But this means that all Seifert circles contained in one of the shaded areas are equivalent. The same argument shows that there is no defect area inside any of the two disks bounded by the Seifert circles involved in our d -pair. Again, the Seifert circles inside each of the disks form a single band of equivalent circles. This shows that there are at most four bands of equivalent Seifert circles in the diagram considered as claimed.

Now we enter the last and most important part of the proof. It is here that we will really see Markov's moves at work (in figures 8 and 11). The problem is to reduce the height maximum occurring at a d -pair ($X \leftarrow Y \rightarrow Z$) of the type shown in figure 7. What is shown in figure 7 in the top row is in fact the diagram Y and the two reduction curves that lead to X and Z . The way to cancel the maximum is shown in figures 7 and 8. First we reduce the diagram Y to closed braids in two ways, as shown in figure 7. Fat lines at the top indicate the reduction curves for initial reductions. In fact multiple reductions are performed, that is, multiple Reidemeister II moves on suitable bands of parallel

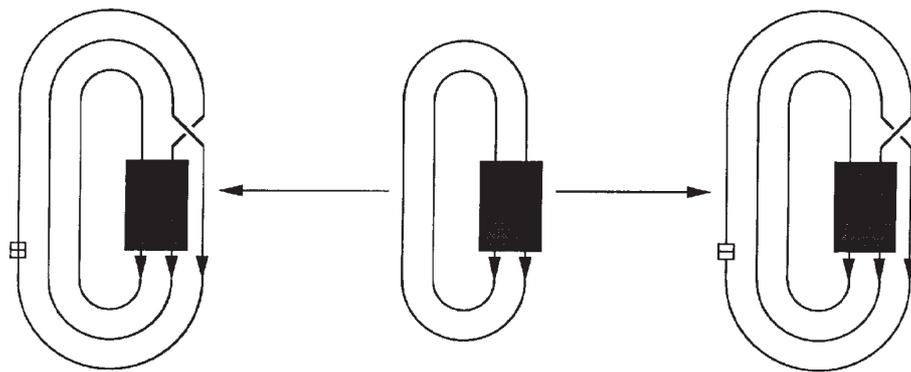


Fig. 10

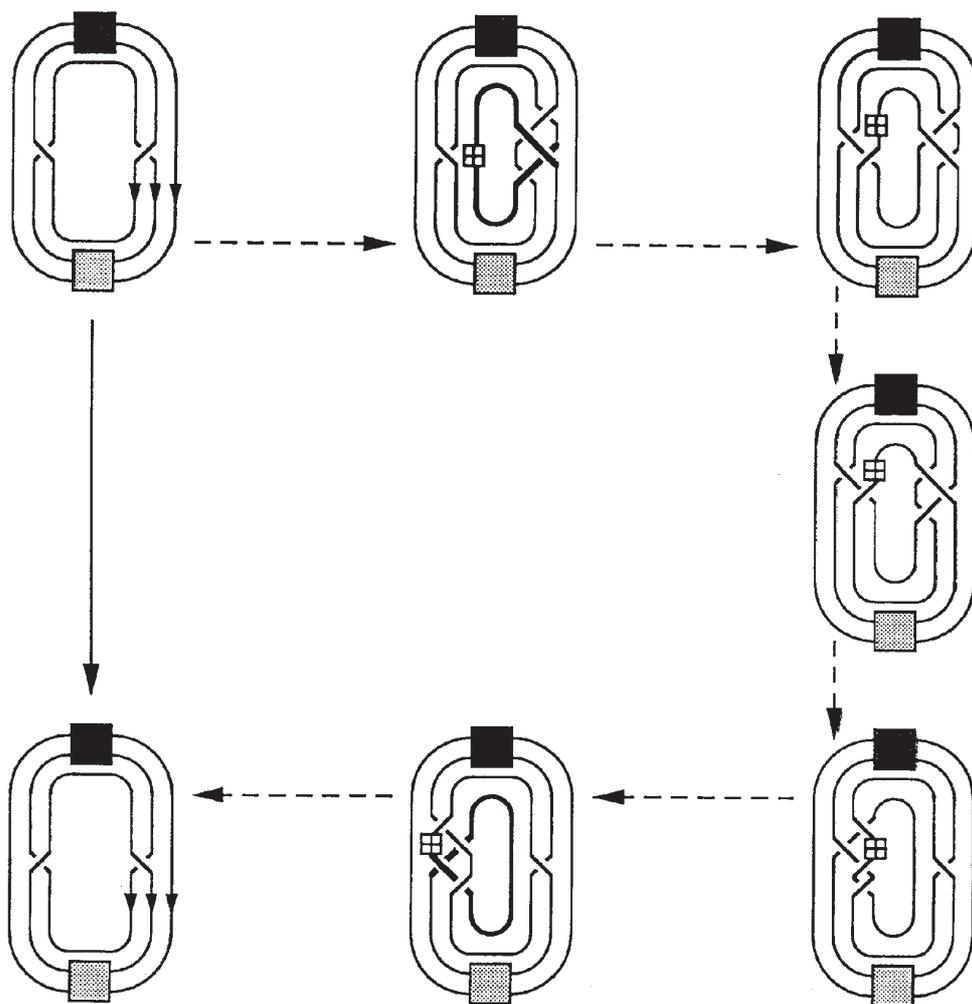


Fig. 11

strings. Obviously, a multiple reduction is in fact equivalent to a sequence of ordinary reducing operations. The result of the first individual reduction is X in one case and Z in the second case (remember we started from a d -pair of the form $X \leftarrow Y \rightarrow Z$). The initial multiple reduction is then (in both cases) followed by another multiple reduction, which is not specially marked in the figure, because at this last stage there is essentially just one obvious possibility to proceed. In the bottom row of the picture the two resulting braids are shown. The reader should remember that we consider diagrams on a sphere rather than on a plane. The use of this is made several times in figure 7. In particular it is due to this that the two braids in the left column are in fact identical. Having reduced the diagrams X and Z to two different closed braids ($Y \rightarrow X \rightarrow \dots \rightarrow X'$ and $Z' \leftarrow \dots \leftarrow Z \leftarrow Y$) we now need to show that X' and Z' are related via a Markov sequence. Such a sequence is shown in figure 9. In this figure (and figure 11) we adopt the additional convention that a square with a $+$ or $-$ sign stands for a positive or negative full-twist on the band of strings. In addition to multiple Reidemeister II and III moves some *multiple Markov moves* of figure 10 are present in the picture (implicitly, for example the first modification of the given braid is in fact a combination of a sequence of braid isotopies followed by a multiple Markov move, then again followed by a sequence of braid isotopies). We leave it as an exercise that such multiple Markov moves are equivalent to sequences of braid isotopies and single Markov moves. Thus the maximum considered is cancelled as required. But there is one last catch. In figure 7 we have chosen arbitrarily how the reducing moves along the given curves are performed, and there is some ambiguity here: we made the decision which arc goes under and which one goes over as was convenient. But clearly the resulting braid is unique up to the exchange move of figure 11, and the exchange move is equivalent to a Markov sequence as shown in the same figure. This completes the proof of Markov's theorem.

Acknowledgements. This is a reworked version of a preprint written in Isaac Newton Institute for Mathematical Sciences in the summer of 1992. I would like to express my gratitude for the Institute's support and hospitality.

References

- [1] J.W. Alexander, *A Lemma on Systems of Knotted Curves*, Proc. Nat. Acad. Sci. 9 (1923), 93–95.
- [2] D. Bennequin, *Entrelacements et équations de Pfaff*, Astérisque 107-108 (1983), 87–161.
- [3] J. Birman, *Braids, links and the mapping class groups*, Annals of Math. Stud. 82, Princeton University Press, 1974.
- [4] A.A. Markov, *Über die freie Äquivalenz geschlossener Zöpfe*, Recueil Mathématique Moscou 1 (1935).
- [5] H.R. Morton, *Threading knot diagrams*, Math. Proc. Camb. Phil. Soc. 99 (1986), 247–260.
- [6] P. Vogel, *Representation of links by braids: A new algorithm*, Comment. Math. Helvetici 65 (1990), 104–113.
- [7] S. Yamada, *The minimal number of Seifert circles equals the braid index of a link*, Invent. Math. 89 (1987), 347–356.