CARLO PETRONIO

AN ALGORITHM PRODUCING HYPERBOLICITY EQUATIONS FOR A LINK COMPLEMENT IN S^3

ABSTRACT. We describe a constructive and effective method for decomposing the complement of an alternating link in the three-sphere into tetrahedra with identifications and vertices removed. Consequently we obtain an algorithm for writing down the hyperbolicity equations associated to such decomposition.

According to the combined results of [Th1], [Th4] and [Ep-Pe] (see also [Ne-Za] and [Be-Pe, Chapter E], in the latter of which one could find a self-contained exposition of the subject) an orientable non-compact hyperbolic three-manifold of finite volume can be decomposed in a finite number of tetrahedra with glued faces and vertices removed; these tetrahedra are isometric to ideal tetrahedra in the hyperbolic three-space and they are parametrized by complex numbers satisfying certain rational equations depending on the combinatorics of the gluings (see Section 3 for definitions and precise statements). Moreover it is well known that such a manifold is the interior of a compact manifold whose boundary consists of tori. The natural question, arising from this topological description, whether the complement of a link in S^3 can be endowed with a hyperbolic structure, was faced and given important answers in [Th3], [Ad1], [Ad2] and [Me2], the results being purely existential as based on Thurston's hyperbolization theorem ([Th2] and [Th4]). On the other hand one can produce examples using the fact that for any three-manifold having a topological decomposition into tetrahedra as the one sketched above the equations can be written down and a solution, if any, actually defines a hyperbolic structure on the manifold.

As suggested by these remarks in this paper we provide an algorithm for writing down hyperbolicity equations (associated to a decomposition into tetrahedra) for the complement of a link in S^3 (not an arbitrary one, but quite a general one; moreover we prove that it is always possible to add components to a link in order to get one to which the algorithm applies). Some basic examples are studied in [Th1] (Thurston's notes are actually the main source for this subject); in [Me1] a description of a more general procedure is partially carried out but in a quite rough way (for instance the conditions under which the procedure works are not explicitly stated; on the contrary we collect those we need in step 1 of our algorithm). We actually

Geometriae Dedicata **44:** 67–104, 1992.

moved from Menasco's paper trying to make the procedure as simple and as effective as possible. Some important differences arise:

(a) we never make use of three-dimensional pictures, all our steps reduce to very elementary operations on plane graphs (this is important in particular in the last steps of the construction, when we add edges);
(b) we prove that the hypothesis that the projection be alternating allows an important simplification in the procedure (Lemma 4.2);
(c) we include in the 'automatic' procedure the calculation of the derivative of the holonomy (giving completeness equations), while Menasco (as well as Thurston in his examples) needed to reconstruct the tori from their triangulation.

Moreover we give a fully detailed description of the geometric construction justifying the steps of the algorithm (in particular the cell decomposition of the link complement is carefully explained, while Menasco just sketched it).

In Section 1 we describe the procedure without any explanation (we call the procedure an 'algorithm' as it is evident that it could be easily implemented on a computer: as to checking if a solution exists, this is another matter). In Section 3 we show how to apply the algorithm on a concrete example (and we prove that the complement of the considered link is actually hyperbolic); by a different application of the algorithm (step 6 actually requires a choice among a specified finite number of possibilities) to the same link we also obtain an interesting fact that we will discuss at the end of the next section. In Section 2 we briefly recall some definitions and the basic results motivating the algorithm, while in Sections 4 and 5 we describe the construction leading to the algorithm. We prove as an appendix the fact that given (the projection of) any link it is possible to add components to it in order to get (the projection of) a link to which the algorithm applies.

For the reader's convenience we have used boldface type to number the effective steps to be performed, both in the description of the algorithm (Section 1) and in its explanation (Sections 4 and 5). Then if one looks for the explanation of Step $n(m)$ of the algorithm (where n and m are integers) he only needs to look for **Step $\boldsymbol{n(m)}$** in Sections 4 and 5 (taking care that it could recur several times). However we remark that explanations of Sections 4 and 5 are necessary for understanding the geometric construction but not for applying the algorithm: with a minor exception, Section 1 is completely self-contained.

1. The algorithm

The starting point of the algorithm is a regular plane projection of a link, i.e. a finite graph $G \subset \mathbb{R}^2$ such that close to each vertex G looks like a cross, in which at each vertex the branch passing over the other one is specified with the usual symbolism. We assume the projection cannot be trivially simplified, i.e. that each vertex of G meets four different components of $S^2 = \mathbb{R}^2 \cup \{\infty\}$ minus G (these components will be called regions) and that if a region has only two vertices then its two edges pass once above and once below at the two vertices. Thus the situations represented in Figure 1 cannot occur.

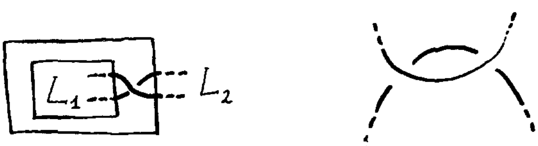

Fig. 1. These situations are excluded by the initial assumptions.

We also assume G is not a plain circle, i.e. a single loop without crossings. (Of course once an explicit way of describing the projection is given these conditions can be automatically checked; the same holds for every operation or verification we are going to make, and we will not repeat the remark again.)

We will refer in Section 4 to the assumptions made till now as 'initial assumptions'. The reason for distinguishing them from the further assumptions we are going to make now is that the 'initial assumptions' will be taken without discussion, while the others will be explained in Section 4.

Step 1. Verification that:

(1) the regions are homeomorphic to the open disc (i.e. the boundary of each region is a simple loop);

(2) the projection is alternating (i.e. for all components K of the link, given any orientation to K, each crossing at which K passes below is followed by one at which it passes above, and conversely);

(3) if two edges in G have the same endpoints then they bound a region;

(4) G is not a projection of the type represented in Figure 2;

Fig. 2. If the number of crossings is even this is the projection of two circles wrapping around each other; otherwise of a single knot (e.g. the trefoil).

(5) the unbounded region has at least three vertices.

If one of these conditions is not fulfilled the algorithm does not work.

From now on we use the term 'bigon' to denote a region having only two vertices (in G and in the other graphs we are going to construct).

Step 2. Choice of a name R_1, $R_2, \ldots$ for each region. Choice of a name $g_1, g_2, \ldots$ for each vertex of G. Choice of a symbol $\odot$ or $\otimes$ for each vertex of G, in such a way that if two edges join two vertices then the two corresponding symbols are different (for an effective method for doing this, see the proof of (4) $\Rightarrow$ (2) in Lemma 4.5). For each component of the link, choice of an edge in G belonging to it, named m_i ($i = 1, 2, \ldots,$ number of the components). Whenever possible, the following criteria are used for such a choice:

(a) m_i is chosen as an edge of a bigon;

(b) m_i is not chosen as an edge having one common vertex with a bigon.

We denote the object thus obtained (a labelled graph) by G again.

Step 3. Production of a new labelled plane graph G_+ coinciding with G as an unlabelled graph and labelled according to:

(1) edges and regions are labelled as indicated by the rules of Figure 3;

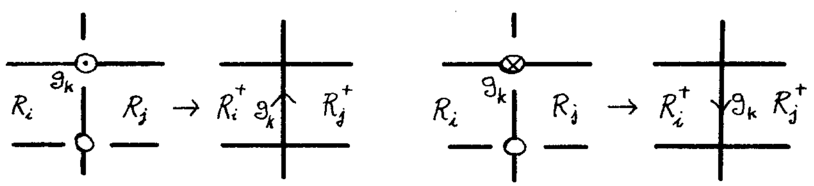

Fig. 3. The rules of **Step 3(1)**.

(2) inside each region R_i^+ we draw a little arrow pointing towards the first endpoint of g_j, where j is the least index appearing on ∂R_i^+;

(3) new symbols labelled by m_i are added as indicated by the rules of Figure 4. (One of the determining aspects of these symbols labelled by m_i is the vertex near which they are drawn.)

Step 4. Production of a new labelled plane graph G_- coinciding with a reversed copy of G (i.e. the image of G under a reflection of the plane relative to a line) as an unlabelled graph, and labelled according to:

(1) edges and regions are labelled as indicated by the rules of Figure 5;

(2) inside each region R_i^- we draw a little arrow pointing towards the first endpoint of g_j, where j is the least index appearing on ∂R_i^-;

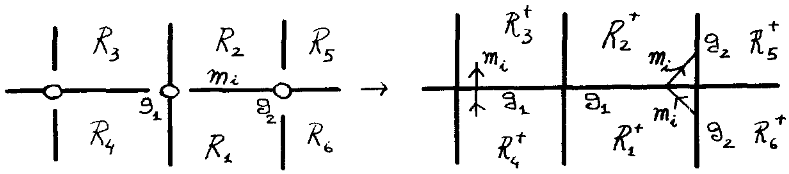

Fig. 4. The rules of **Step 3(3)**.

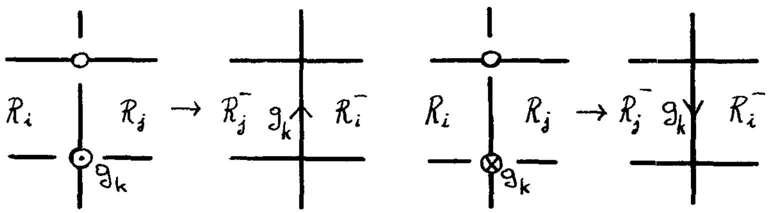

Fig. 5. The rules of **Step 4(1)**.

(3) new symbols labelled by m_i are added as indicated by the rules of Figure 6.

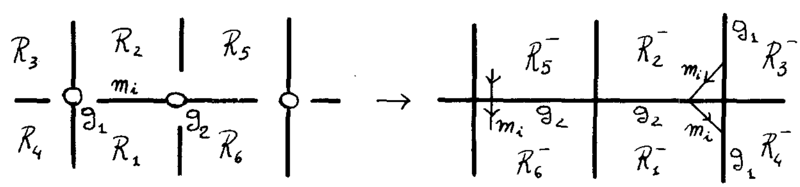

Fig. 6. The rules of **Step 4(3)**.

Since Figure 5 may cause confusion as it depends on the position of the part in question with respect to the line relative to which we are reversing the graph, we rephrase the rule for **Step 4(1)**:

(i) we give the name R_i^- to the region obtained by reversing R_i;

(ii) let R_i and R_j be adjacent along the edge e in G; if the endpoint of e at which e passes above has name g_k, then in G_- we have that R_i^- and R_j^- are adjacent along g_k;

(iii) with the same notation, let i and j be chosen in such a way that a segment going from R_i to R_j leaves g_k on the right (we possibly need to interchange the roles of i and j); then if the symbol $\odot$ is attached to g_k, in G_- we have that g_k runs clockwise with respect to R_j^- and counterclockwise with respect to R_i^-; the converse for the symbol $\otimes$.

Similarly one could rephrase the rule of **Step 4(3)** independently of the line relative to which we are reversing.

From now on the objects represented in Figure 7 will be called respectively a 'passage' and an 'arc' of m_i.

Fig. 7. The object at the left-hand side of the figure will be called a 'passage' of m_i, and the one at the right-hand side an 'arc' of m_i.

Step 5. Production of two new labelled graphs G_+ and G_- starting from the old ones, according to the following rules (the notation refers to G_+, but the rules are the same for G_-):

(1) if in the old G_+ two edges join two vertices, then one of them is deleted (we call this operation 'elimination of the bigon' bounded by the two edges);

(2) the little arrows inside the bigons are forgotten, and the other are copied;

(3) we fix an ordering of the bigons and eliminate them following this ordering; when eliminating a bigon with edges g_i and g_j we give the surviving edge the name g_i where $i = \min\{i, j\}$, and moreover we give the name g_i to all other edges previously having name g_j;

(4) arcs and passages of the m_i's not involving bigons are copied.

In order to establish what one of the objects represented in Figure 8 (or a similar one with the orientation of m_i reversed) causes to appear, a recursive method is used: one first eliminates the bigon in question according to the rules of Figure 9 and similar ones with the orientation of m_i reversed). (Remark that the rule described in the upper-right part of Figure 9 refers to a situation which cannot now occur.)

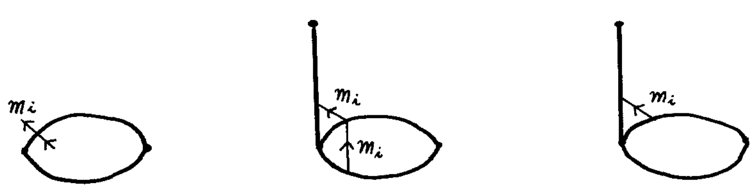

Fig. 8. The objects in the old graph of which the effect on the new graph is described in **Step 5(4)**.

Then the recursive method starts again (and now a situation as in the upper-right rule can occur). The method stops when the arc or passage of m_i we have produced does not touch a bigon, and then it is copied in G_+.

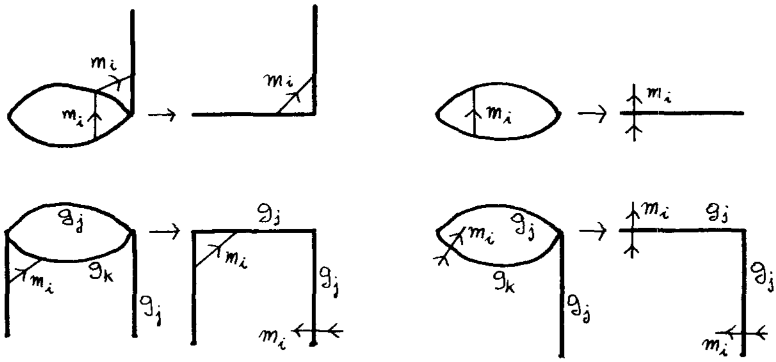

Fig. 9. The rules of **Step 5(4)**.

Remark that this recursive method requires to perform the elimination of the bigons (following the four rules of Figure 9) in several different ways (i.e. according to several different orderings). However if the bigons are isolated (i.e. if any two bigons do not have common vertices) the recursive method stops after the first step, and then we only need to perform the elimination of the bigons following any fixed ordering, without recursive methods; this allows us to simplify much of the procedure.

Step 6. In G_+ we add $p-3$ edges (without adding vertices) inside each region R_i^+ (even in the unbounded one) having $p \geqslant 4$ edges, in order to divide it into triangles (of course there exists a finite number of possibilities for doing this). We correspondingly add edges inside each R_i^- in the following way: (i) starting from the vertex towards which the little arrow inside R_i^+ points and proceeding counterclockwise we number the vertices of R_i^+; (ii) starting from the vertex towards which the little arrow inside R_i^- points and proceeding clockwise we number the vertices of R_i^-; (iii) if in R_i^+ an edge is added having the nth and the mth vertex as endpoints, the same is done in R_i^-.

Each R_i^- is now divided into triangles too. We give the added edges names $g_{m+1}, g_{m+2}, \ldots$ (if m was the greatest index previously appearing) both in G_+ and G_- (of course we give the same name to corresponding edges). The remainder of the labellings of G_+ and G_- (excluding the names of the regions and the little arrows, which no longer make sense) is kept unchanged, with the exception of the arcs labelled by m_i, which transform with the rule described in Figure 10.

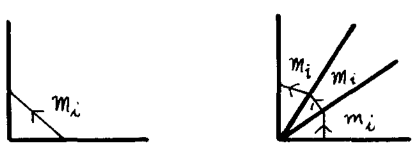

Fig. 10. The rule of transformation of the arcs m_i.

We also delete the orientation of the g_i's. We keep calling G_+ and G_- the labelled graphs thus obtained.

Let us assume now that in G_+ and G_- any two edges have at most one common endpoint. Since there exists a finite number of possibilities for the subdivision of the R_i^+'s it is possible to establish by a finite number of calculations whether we succeed or not in fulfilling this condition: if we do not the procedure fails.

Step 7. In G_+ we choose a vertex v^+ of the unbounded triangle. For each of the triangles not having vertex v^+ we write the symbols z_i, z_i', z_i'' inside the triangle and near the three edges in a counterclockwise order (starting from an arbitrary edge). Near each edge having endpoint v^+ we write the symbol thus obtained: if v is the other endpoint we take the product of all the symbols written inside the triangles containing v and not v^+ near the edge opposite to v. We also adopt the convention that if there are two symbols written near an edge e, when speaking of 'symbol written near e' we actually mean the product of these two symbols; remark however that the phrase 'symbol written near e inside T' (where T is a triangle containing e) has the same meaning as above, i.e. it defines a single symbol and not the product of two.

We repeat the procedure in the very same way for G_-, calling v^- the chosen vertex of the unbounded triangle and using the symbols w_i, w_i', w_i''.

We are going to write now formal equations of the form 'product of symbols $z_i, z_i', z_i'', w_i, w_i', w_i''=1$'; they are translated into true equations in the complex variables $z_1,\dots,w_1,\dots$ by the convention $u'=1/(1-u)$ and $u''=1-1/u$. If a solution of these equations exists with all the z_i's and w_i's in the upper half-plane then the complement of the link considered at the beginning can be endowed with a hyperbolic structure.

First compatibility equations. We describe a procedure referring to G_+ for the notation, to be repeated for G_- too. For all the vertices v being not joined to v^+ we take the product of all the symbols written inside the triangles containing v and not v^+ near the edge opposite to v, and we set this product to 1.

Second compatibility equations. For all the g_i's we set to 1 the product of the symbols written near the various existing copies of g_i in both G_+ and G_-.

Completeness equations. For all i's we select the passages and arcs of m_i leaving the vertex they are referred to on the left, and mark them with a little circle as represented in Figure 11.

Then we take the product of the symbols written near all the edges marked by a little circle, and we set it to 1 if the number of arcs of m_i in G_+ is even, -1 otherwise. We also set the sum of the arguments of these symbols to be π

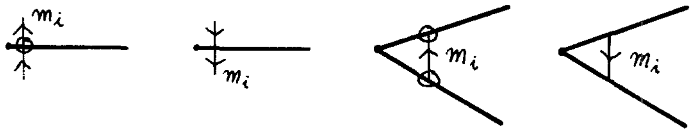

Fig. 11. The little circles we draw if m_i leaves its vertex on the left.

times the number of edges of m_i; if the number of edges of m_i is 1 or 2 then this equation is dispensable.

2. An example

Figures 13 through 16 and the discussion which follows show how to apply the algorithm to the link represented in Figure 12.

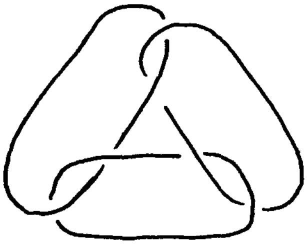

Fig. 12. The projection to which we apply the algorithm.

Step 1: Okay.

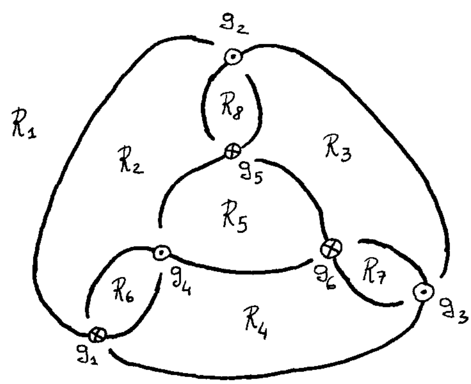

Fig. 13. Application of **Step 2**.

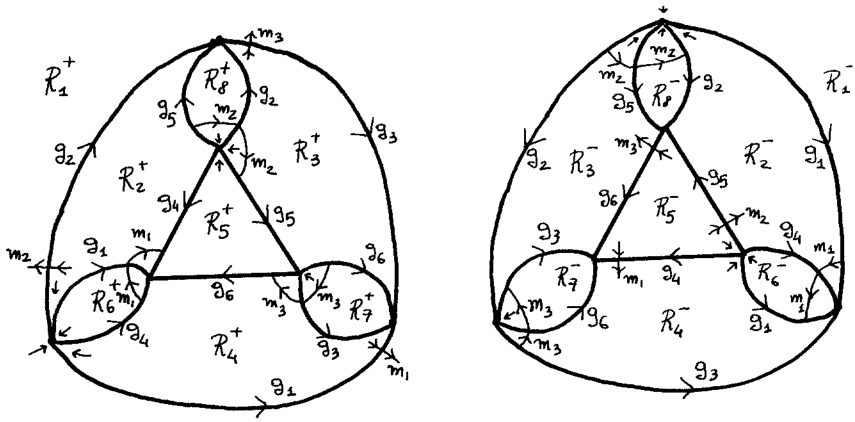

Fig. 14. Application of **Steps 3 and 4**.

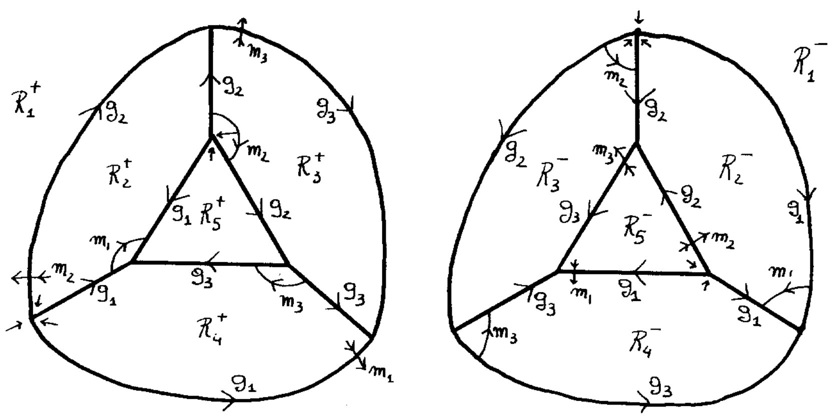

Fig. 15. Application of **Step 5**.

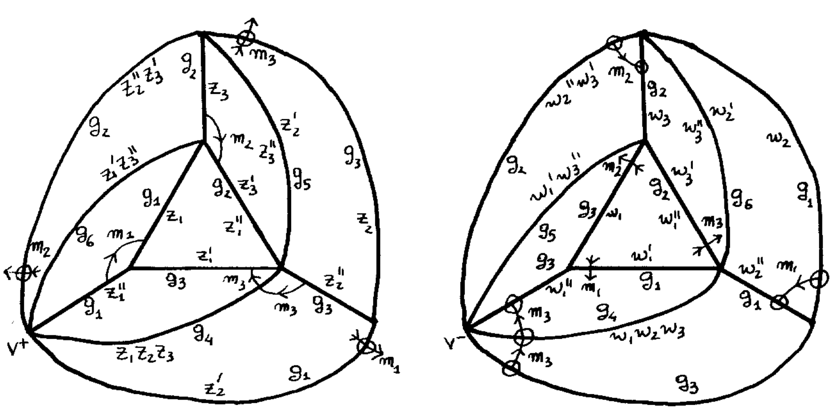

Fig. 16. Application of **Steps 6 and 7**.

First compatibility equations: none.
Second compatibility equations:

$$g_1:\quad z_2'\cdot z_1''\cdot z_1\cdot w_1'\cdot w_2''\cdot w_2=1$$
$$g_2:\quad (z_2''\cdot z_3')\cdot z_3\cdot(z_1''\cdot z_3')\cdot(w_2''\cdot w_3')\cdot w_3\cdot(w_1''\cdot w_3')=1$$
$$g_3:\quad z_1'\cdot z_2\cdot z_2''\cdot w_2'\cdot w_1''\cdot w_1=1$$
$$g_4:\quad (z_1\cdot z_2\cdot z_3)\cdot(w_1\cdot w_2\cdot w_3)=1$$
$$g_5:\quad (z_2'\cdot z_3'')\cdot(w_1'\cdot w_3'')=1$$
$$g_6:\quad (z_1'\cdot z_3'')\cdot(w_2'\cdot w_3'')=1.$$

Completeness equations:

$$m_1:\quad z_2'\cdot w_2\cdot w_2''=-1$$
$$m_2:\quad (z_2''\cdot z_3')\cdot(w_2''\cdot w_3')\cdot w_3=-1$$
$$m_3:\quad z_2\cdot w_2'\cdot(w_1\cdot w_2\cdot w_3)\cdot w_1''=1.$$

It is not so hard to check that all these equations (translated into true equations by the stated convention) have the unique solution in the product of six upper half-planes

$$z_1=z_2=w_1=w_2=\frac{1}{2}+i\frac{\sqrt{7}}{2},\qquad z_3=w_3=\frac{3}{8}+i\frac{\sqrt{7}}{8}.$$

In order to give an example of the first compatibility equations we consider now another way to perform **Step 6**, as represented in Figure 17. This will also lead us to an important fact we will comment on at the end of Section 3.

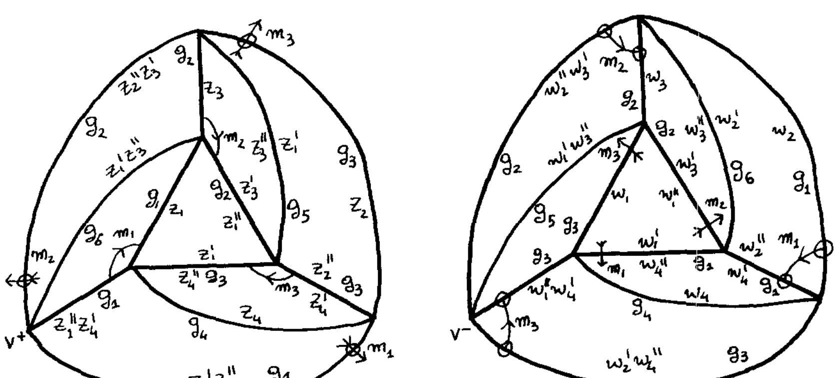

Fig. 17. Application of **Steps 6 and 7** (second version).

First compatibility equations:

$$z_1 \cdot z_2 \cdot z_3 \cdot z_4 = 1$$
$$w_1 \cdot w_2 \cdot w_3 \cdot w_4 = 1.$$

Second compatibility equations:

$$g_1\colon\ (z_2' \cdot z_4'') \cdot z_1 \cdot (z_1'' \cdot z_4') \cdot (w_1' \cdot w_4'') \cdot w_2 \cdot (w_2'' \cdot w_4') = 1$$
$$g_2\colon\ (z_2'' \cdot z_3') \cdot z_3 \cdot (z_1'' \cdot z_3') \cdot (w_2'' \cdot w_3') \cdot w_3 \cdot (w_1'' \cdot w_3') = 1$$
$$g_3\colon\ (z_1' \cdot z_4'') \cdot z_2 \cdot (z_2'' \cdot z_4') \cdot (w_2' \cdot w_4'') \cdot w_1 \cdot (w_1'' \cdot w_4') = 1$$
$$g_4\colon\ z_4 \cdot w_4 = 1$$
$$g_5\colon\ (z_2' \cdot z_3'') \cdot (w_1' \cdot w_3'') = 1$$
$$g_6\colon\ (z_1' \cdot z_3'') \cdot (w_2' \cdot w_3'') = 1.$$

Completeness equations:

$$m_1\colon\ (z_2' \cdot z_4'') \cdot w_2 \cdot (w_2'' \cdot w_4') = -1$$
$$m_2\colon\ (z_2'' \cdot z_3') \cdot (w_2'' \cdot w_3') \cdot w_3 = -1$$
$$m_3\colon\ z_2 \cdot (w_2' \cdot w_4'') \cdot (w_1'' \cdot w_4') = -1.$$

It is easily checked that if one uses the first two equations to eliminate z_4 and w_4 the result is the system of 9 equations in 6 unknowns we obtained above. Then we must have that $z_1, \ldots, w_3$ are as above, and hence

$$z_4 = w_4 = (z_1^2 \cdot z_3)^{-1} = -1$$

so that there is no solution in the product of upper half-planes.

3. Three-manifolds decomposing into tetrahedra and hyperbolicity equations

Let M be a (connected and orientable) non-compact 3-manifold without boundary being the interior of a compact manifold with boundary made of tori. It is possible to prove (see [Ma-Fo] and [Be-Pe]) that M can be obtained according to the following procedure: a finite number of copies of the standard 4-simplex (the tetrahedron) is considered, the faces of these tetrahedra are glued in pairs (via simplicial maps) and the vertices are removed. (In other words, M can be 'triangulated' by tetrahedra without vertices.) We will denote by $\mathscr{T}_3$ the 'set' of all these manifolds together with their realization as above (the procedure is not unique in general). Elements of $\mathscr{T}_3$ will be denoted as the manifold they represent, a fixed realization as glued tetrahedra being understood.

From now on with the terms tetrahedron, face and edge we will often understand that vertices are removed. Given $M \in \mathscr{T}_3$ obtained by gluing the tetrahedra $\Delta_1, \ldots, \Delta_n$ we reserve the term projection for the natural mapping from the disjoint union of $\Delta_1, \ldots, \Delta_n$ onto M; a tetrahedron (face, edge) in M is just the projection of a tetrahedron (face, edge). Remark that each tetrahedron in M is the projection of precisely one tetrahedron and each face is the projection of precisely two faces, while the edges may be the projection of several edges.

It is quite easily checked that the edges in an element of $\mathscr{T}_3$ are as many as the tetrahedra.

Given $M \in \mathscr{T}_3$ obtained by gluing n tetrahedra and removing k vertices we will describe a system of $n+k$ rational equations in the product of n copies of the upper half-plane Π^+ whose solution, if any, essentially represents a hyperbolic structure on M (i.e. a complete Riemannian structure with sectional curvatures -1; the word 'essentially' means that there is another minor verification to carry out). It is worth remarking that the topological assumptions imply that M has finite volume with respect to any such structure, and moreover by the rigidity theorem the structure, if any, is unique.

The basic idea is to realize the tetrahedra as ideal tetrahedra in the hyperbolic 3-space $\mathbb{H}^3$ and try to globalize and make complete the hyperbolic structure naturally defined on the interior of the tetrahedra. The structure always naturally extends to the interior of the faces; n equations will come from the requirement that the structure extend to the n edges, and k more equations from the requirement that the structure be complete on a neighborhood of the k removed vertices (this fact being equivalent to completeness, as a compact set is left out).

We recall that if we fix a preferred pair of opposite edges on an ideal tetrahedron in $\mathbb{H}^3$, its oriented isometry class is parametrized by Π^+. We use the notion of modulus of an ideal tetrahedron along an edge with the obvious meaning. Figure 18 shows the moduli of a tetrahedron along the different edges, where z' stands for $1/(1-z)$ and z'' stands for $1-1/z$.

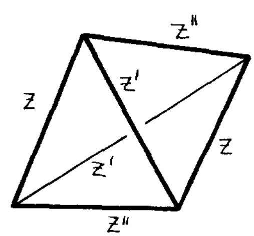

Fig. 18. Moduli of an ideal tetrahedron.

Now, let us turn to our $M \in \mathscr{T}_3$ obtained by gluing n tetrahedra; we choose on each of them a preferred pair of opposite edges and realize them as ideal tetrahedra in $\mathbb{H}^3$; we denote the corresponding moduli by $z_1, \ldots, z_n$. The condition that the hyperbolic structure extend to a certain edge e is easily translated in terms of $z_1, \ldots, z_n$: we need that the product of the moduli along the various edges projecting onto e is 1 and that the sum of the arguments of these moduli is 2π. Moreover a lemma to be found in [Be-Pe] implies that it is sufficient to impose the first condition ('product of the moduli = 1') for all the n edges. Remark that these are in fact n rational equations in $z_1, \ldots, z_n$.

Now we turn to completeness. Let $z_1, \ldots, z_n$ define a hyperbolic structure extending to the whole M. Then the toric links of the k removed vertices are obtained by gluing in pairs the edges of plane Euclidean triangles (well-defined up to similarity, each triangle being the intersection of one of the tetrahedra with a suitably small horosphere centered at one of the vertices) and the condition that the structure extend to the whole M implies that the similarity structure globalizes to the toric link.

It turns out (see [Be-Pe]) that the hyperbolic structure is complete if and only if these similarity structures on the k tori are actually Euclidean. In turn this is equivalent to the fact that with respect to the similarity structure the holonomy of a meridian of each torus is a translation (i.e. has derivative 1) and is not the identity.

Then, going back to our M, we must compute the derivative of the holonomy of a meridian m of each of the k tori with respect to the similarity structure induced by $z_1, \ldots, z_n$. This is achieved by representing m as a simple simplicial loop: if $\# m_0$ denotes the number of vertices of m then the derivative of the holonomy of M is given by $(-1)^{\# m_0}$ times the product of the moduli of the triangles m leaves on the right along the vertices on m. (The modulus of a triangle T along a vertex v is the modulus of the tetrahedron containing T along the edge containing v.) So if the structure is complete we have that $z_1, \ldots, z_n$ satisfy k more rational equations. The converse is not precisely true as we have to check that the holonomy is not the identity; this condition is equivalent to having the sum of the arguments of the moduli m leaves on the right equal to $\# m_0 \cdot \pi$. Remark that if the previous equations are satisfied this sum must be $(\# m_0 + 2k)\pi$ for some k; so we can think to these equations involving the arguments as a 'final verification': we first solve the rational equations and then we check that the arguments of the solutions satisfy the last equations. Remark as well that if $\# m_0$ is 1 or 2 then the condition about the arguments is automatic.

As we said, given $M \in \mathscr{T}_3$, the existence of a solution for the compatibility and completeness equations (which depend only on the combinatorics) provides a hyperbolic structure on M. Then one may wonder if a solution of the equations must exist whenever a hyperbolic structure is defined on M. In [Be-Pe] it is shown that a solution must exist with the unknowns in $\mathbb{C}\setminus\{0,1\}$, but the question is not settled whether it is possible or not to find a solution with the unknowns in Π^+. Since we are going to see in the next sections that the equations described in the algorithm are those associated to a realization of the link complement in $\mathscr{T}_3$, the example of Section 2 implies the following:

THEOREM 3.1. *Let an element M of $\mathscr{T}_3$ be given and consider the compatibility and completeness equations associated to the realization of M. Existence of a solution for these equations in a product of half-planes is a sufficient but in general not a necessary condition for M to be hyperbolic.*

4. EXPLANATION OF THE ALGORITHM: GEOMETRIC CONSTRUCTION AND COMPATIBILITY EQUATIONS

Let us consider a link L represented by a regular projection satisfying the initial assumptions stated at the beginning of Section 1. We start by describing a cell decomposition of the space S^3/L obtained from S^3 by collapsing each component of L to a point, in which the 0-cells are precisely the collapsed components; later we will discuss how (and when) to such a cell decomposition it is possible to associate a realization in $\mathscr{T}_3$ of $S^3\setminus L = (S^3/L)\setminus$0-cells. In this and the following section we will use several three-dimensional pictures by looking at the projection plane of L as a horizontal one in $\mathbb{R}^3$; however we will show that everything can be represented by 2-dimensional pictures as we actually need for the algorithm.

We start by assuming that all the regions (the components $R_1, R_2, \ldots$ of $S^2 = \mathbb{R}^2 \cup \{\infty\}$ minus G, where G is the projection in question, a graph) are homeomorphic to the open disc (**Step 1(1)**). Remark that by the initial assumption that each vertex be in the closure of four different regions we obtain that the closure of each region is homeomorphic to the closed disc. From now on we will often use the term 'region' to denote the closure of the components of $S^2\setminus G$; remark that the regions are polygons, so that it is natural to speak of their edges and vertices.

The assumption that the regions be discs is necessary for associating to G a cell decomposition of S^3 in which the 0-cells are the vertices of G, the 1-cells are the edges of G, the 2-cells are the regions and the 3-cells are the two components of $S^3\setminus S^2$, denoted by B_+ and B_- as they are 3-dimensionally viewed as the upper and lower half-spaces.

Remark that the assumption that the regions be discs is automatic for knots but in general not for links; moreover this condition is obviously necessary for $S^3\backslash L$ to be irreducible and hence to possess a hyperbolic structure.

The first step is to construct a new 2-complex $C \subset S^3$ by modifying the cell decomposition of S^3 up to the 2-skeleton. C is constructed in such a way that (a link equivalent to) L is a subset of its 1-skeleton: this is achieved by replacing each vertex in G with a short vertical segment as suggested in Figure 19.

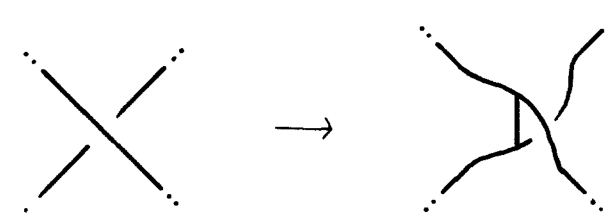

Fig. 19. How to replace a crossing by a vertical segment.

The 0-cells and the 1-cells of this new 2-complex C are evident; the 2-cells are obtained by modifying the previous regions as suggested by Figure 20.

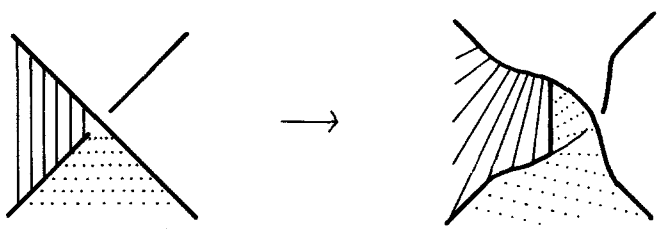

Fig. 20. How to modify the old regions and get the 2-cells.

We keep denoting the modified regions by R_1, R_2.... Remark that these new 2-cells are nicely chosen, so that S^3 minus C still consists of two balls we keep denoting by B_+ and B_-. Remark, however, that C together with these two balls does not define in general a cell decomposition of S^3 (B_+ and B_- are not bounded by spheres now).

We describe how to associate to C a cell decomposition of S^3/L. We start by giving names $g_1, g_2, \ldots$ to the added vertical segments and attaching to them an orientation (an arbitrary one at the moment: we will come back to this later). The following is the description of the cell decomposition (becoming more and more complicated as the dimension of the cells grows).

(a) The 0-cells are the collapsed components $p_1, p_2, \ldots$ of the link.

(b) The 1-cells are segments we keep denoting by $g_1, g_2, \ldots$ coming from the vertical segments; the gluing function is obvious: the first (second) endpoint of g_j is mapped to p_i if the first (second) endpoint of g_j as a vertical segment lies on the component collapsing to p_i.

(c) The 2-cells come from the above modified regions and they are denoted by $R_1, R_2, \ldots$ again; the right gluing is easily constructed: the boundary of a modified region consists of (oriented) segments g_i alternated with segments lying on the link (denoted by the same symbol p_i as the collapsed component they lie on); then we only need to shrink to points the segments of the latter type. An example of this is given in Figure 21; the construction implies that the boundary of the 2-cell in question is mapped onto the loop $g_2 \cdot g_3 \cdot g_1^{-1}$ in the 1-skeleton.

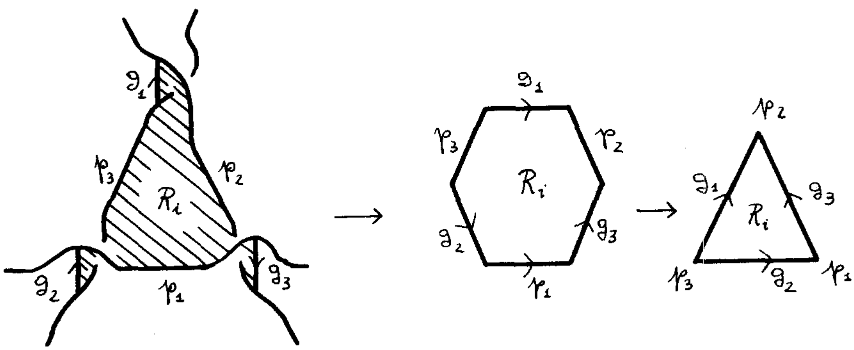

Fig. 21. How to obtain the gluing functions of the 2-cells.

(d) The 3-cells are two and they come from the balls B_+ and B_-; in order to describe the gluing functions we give explicit realizations, corresponding to the gluings of B_+ and B_-, of $S^2 = \partial B^3$ as union of the 2-cells. The point is to partially perform the collapsing of the components in order to make C become a horizontal plane again.

We start with B_+ and remark that in a situation as in Figure 22 we have that 'looking from above', i.e. from B_+, R_1 and R_2 are adjacent along g, and so are R_3 and R_4.

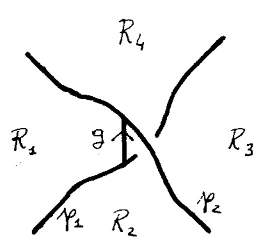

Fig. 22. Determination of the adjacencies of the 2-cells.

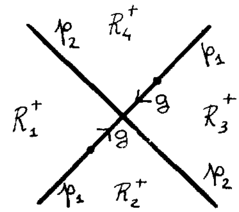

Fig. 23. Representation of the adjacencies obtained.

Then we replace the part of graph represented in Figure 22 by the part of graph given in Figure 23 (and lying on the horizontal plane); remark that we are again denoting by p_i both the components of the link and the 0-cells obtained by collapsing them.

We do this for all the g_i's and we get a graph (obtained by adding to G some vertices) in which some edges are labelled with a symbol p_i and some are labelled with a symbol g_i and an orientation; then we only have to shrink to points the edges labelled by p_i (of course we must allow the other edges to stretch). An example of this procedure is described in Figures 24 and 25.

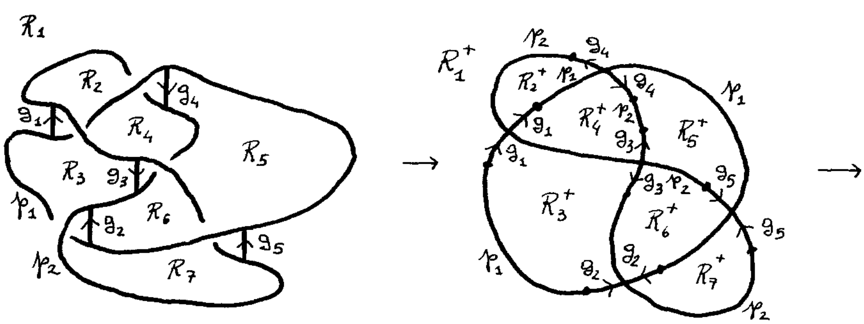

Fig. 24. An example of how to shrink the p_i edges.

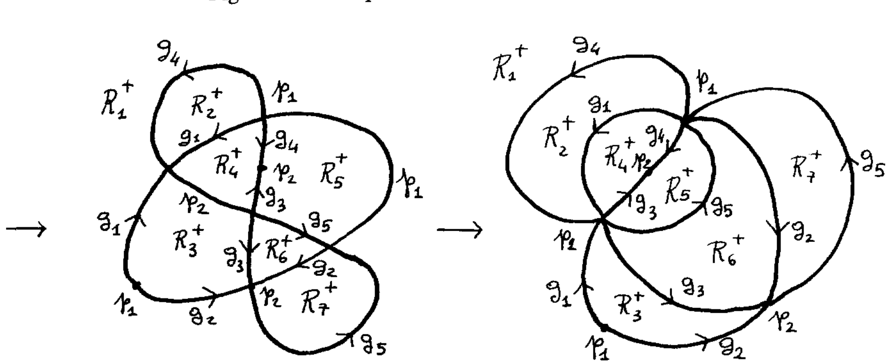

Fig. 25. Example of Figure 24 continued.

The object we obtain at last is a plane graph we denote by G_+, in which the edges are oriented and labelled by the symbols g_i (each appearing twice); as in the above example, the components of $S^2 \backslash G_+$ are denoted by $R_1^+, R_2^+, \ldots$ corresponding to $R_1 R_2, \ldots$. The gluing of the 3-cell coming from B_+ (denoted by B_+ again) is obtained by considering a copy $\mathbb{R}_+^2$ of the plane containing G_+, setting $S_+^2 = \mathbb{R}_+^2 \cup \{\infty\}$, considering S_+^2 as the boundary of the upper half-space B_+ and gluing S_+^2 to the 2-skeleton in the way prescribed by the labelling: each p_i must be mapped to the corresponding 0-cell, each g_i must be mapped to the corresponding 1-cell following the orientation, each R_i^+ must be mapped to the 2-cell R_i following the labelling of the boundary. Of course by the very construction these gluing instructions do not cause any contradiction.

A similar construction works for the other 3-cell B_-. We replace the left-hand side of Figure 26 by its right-hand side. Then we shrink to points the edges labelled by p_i and we call G_- the resulting labelled graph. As above we consider $\mathbb{R}_-^2 \supset G_-$, $S_-^2 = \mathbb{R}_-^2 \cup \{\infty\}$; the only difference now is that B_- is the lower half-space.

Fig. 26. The rule for B_-.

The realization of S^3/L as a cell complex is now perfectly described. The construction seems to be quite involved, but a further assumption on the projection will make it much simpler. Before discussing this we prove another result, implying that the graphs G_- and G_+ together with their labelling are actually enough to reconstruct everything; even more: from G_+ and G_- one can remove the symbols p_i and the result remains true. Let us remark first of all that the labellings of ∂R_i^+ and ∂R_i^- suffice to define without ambiguity the identification between R_i^+ and R_i^-; in fact each g_i appears at most once on them.

LEMMA 4.1. *S^3/L is obtained by considering the disjoint union of two balls B_+ and B_- bounded by S_+^2 and S_-^2 as described above, and then identifying each R_i^+ to R_i^- as indicated by the labelling of ∂R_i^+ and ∂R_i^-.*

Proof. What we have to show is that from the gluing of the 2-cells one deduces the gluings of the 0-cells and the 1-cells.

We start with the 1-cells; a situation like the one represented in Figure 27 gives in G_+ and G_- respectively the left-hand side and the right-hand side of Figure 28.

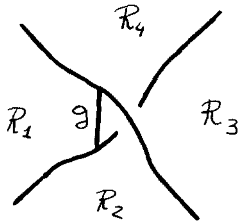

Fig. 27. Proof that the gluings of the 2-cells imply all the gluings of the 1-cells; a generic crossing.

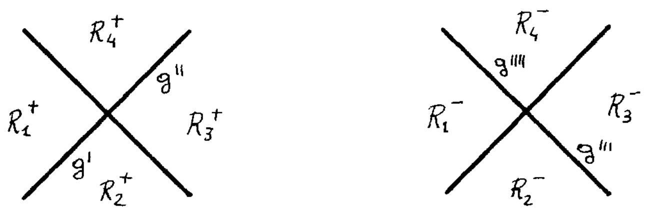

Fig. 28. These situations in G_+ and G_- come from Figure 27.

In Figure 28 we have added apices to the different copies of g in order to check that the gluing of the 2-cells actually allows to identify them all; in fact we have:

$$R_1^+ = R_1^- \Rightarrow g' = g'''', \; R_2^+ = R_2^- \Rightarrow g' = g''',$$
$$R_3^+ = R_3^- \Rightarrow g'' = g'''$$

and then the four copies of g are glued together.

Now, for 0-cells we confine ourselves to an example and leave the general case to the reader. Let $f(g)$ and $s(g)$ denote respectively the first and second endpoint of an oriented segment g. In a situation like that of Figure 29 we must check that $s(g_1) = f(g_2) = f(g_3) = s(g_3) = s(g_4) = s(g_5)$.

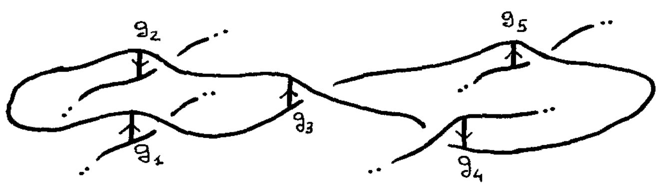

Fig. 29. Proof that the gluings of the 2-cells imply all the gluings of the 1-cells. An example.

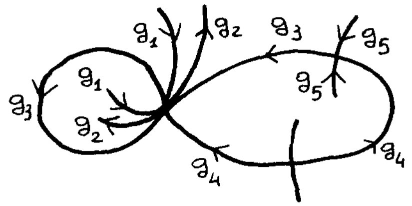

Fig. 30. Example of Figure 29 continued.

In G_+ we obtain the situation of Figure 30 (remark that we already know that the two copies of g_i are glued together).

Then we easily have that

$$s(g_1) = f(g_2) = f(g_3) = s(g_3)$$
$$f(g_3) = s(g_4) = s(g_5)$$

and the conclusion follows. □

According to the above result we remove from G_+ and G_- the symbols p_i and keep denoting the labelled graphs thus obtained by G_+ and G_-.

We discuss now the further assumption on the projection making the construction of G_+ and G_- much simpler. We say the projection of L is alternating if for all components K of L, given any orientation to K, each crossing at which K passes above is followed by one at which it passes below, and conversely.

Since the next result describes the first effective step in our algorithm we introduce some symbols in order to keep the pictures plane. To each vertex of G we attach a symbol ⊙ if we choose the corresponding vertical segment to point upwards, ⊗ otherwise.

LEMMA 4.2. *Assume the projection is alternating; then G_+ and G_- both coincide with G (the initial projection) as unlabelled graphs; moreover the labelling is obtained according to the rules given in Figure* 31 *and Figure* 32 (*for G_+ and G_- respectively*).

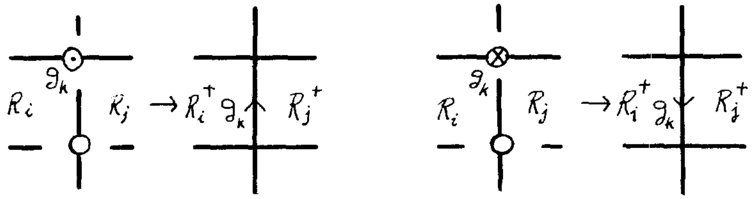

Fig. 31. Statement of Lemma 4.2: rules for G_+.

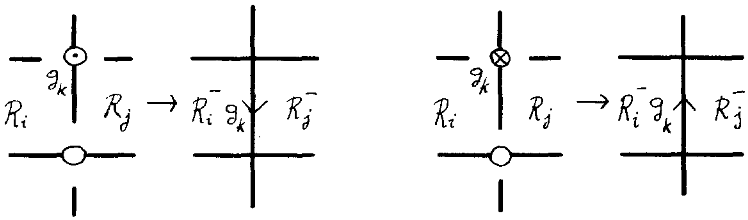

Fig. 32. Statement of Lemma 4.2: rules for G_-.

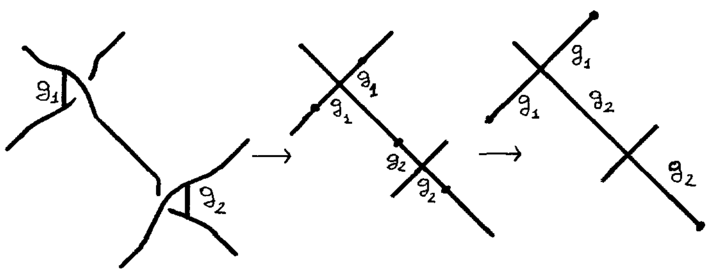

Fig. 33. Generic situation for an edge in the case of alternating projection.

Proof. We refer to G_+. All the edges in G are in the situation of Figure 33. Then we have that on each edge e of G there exists precisely one edge labelled by g_i; it follows that shrinking the other edges to points just corresponds to extending the label g_i to the whole e.

Orientation matters are easily settled. □

From now on we will assume that the projection is alternating (**Step 1(2)**), so that the construction of G_+ and G_- becomes much simpler. There are other reasons for assuming this, as we will point out later.

Before summarizing what we have we remark that for uniformity of notation it is convenient to reverse G_- (i.e. transform $\mathbb{R}^2_-$ by any orientation-reversing mapping, for instance the reflection relative to a line) so that S^2_- must be considered to bound the upper half-space, just like S^2_+. From now on upper half-spaces will be considered closed.

The following result summarizes the construction.

THEOREM 4.3. *Let G be a graph representing an alternating projection of a link L and satisfying the initial assumptions. Denote the components of $S^2 = \mathbb{R}^2 \cup \{\infty\}$ minus G by $R_1, R_2, \dots$ and assume they are homeomorphic to discs. Give the vertices of G names $g_1, g_2, \dots$ and attach to each of them a symbol $\odot$ or $\otimes$. Produce two new plane labelled graphs G_+ and G_- according to the rules given in* **Step 3(1)** *and* **Step 4(1)**. *Let S^2_+ and S^2_- be spheres (horizontal* *plane plus ∞ in distinct copies of $\mathbb{R}^3$) containing G_+ and G_- respectively; look at them as bounding the upper half-spaces B_+ and B_-. Then S^3/L is obtained by taking the disjoint union of B_+ and B_- and identifying each R_i^+ to R_i^- according to the labelling of the boundaries. Moreover the collapsed components of the link come precisely from the vertices of G_+ and G_-.*

Our next step will be to add edges and faces (but not vertices) to B_+ and B_- in order to decompose them into tetrahedra (i.e. triangulate them); of course this produces in an obvious way a realization in $\mathscr{T}_3$ of $S^3 \backslash L$.

Remark that S_+^2 and S_-^2 are now presented as polyhedra (nicely glued polygons); of course a necessary condition that by adding edges and faces one can decompose B_+ and B_- into tetrahedra is that no 'bigon' (polygon with two edges) appears. However, if we do have a bigon, we can eliminate it by identifying its two edges (i.e. by giving them the same name) and allowing the surrounding polygons to enlarge; in order to do this for all the bigons we need that:

(a) the two edges of the each bigon have 'the same orientation' (see Figure 34);

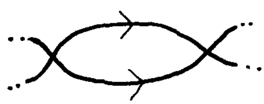

Fig. 34. A bigon whose edges have the same orientation.

(b) there does exist some region being not a bigon;

(c) the edges of each bigon keep having different names during the procedure of elimination (remark that this is true at the beginning, but when we eliminate a bigon with edges g_1 and g_2 and give them the name, say, g_1, then we must give name g_1 to all other existing copies of g_2, which may cause problems). The reason for this requirement is that if the edges of a bigon have the same name then the bigon projects onto a sphere in S^3/L, and there is no sense in switching it to a segment.

Conversely, these conditions are sufficient for eliminating the bigons.

The following result concerns the second and third of these conditions.

LEMMA 4.4. (1) *The only projections (satisfying the underlying assumptions) such that all the regions are bigons is the one consisting of two plain circles with two intersection points (being the projection of two trivially chained unknots)*;

(2) *the procedure of successive elimination of the bigons produces a bigon with edges having the same name if and only if one of the following situation occurs*:

(i) *in G we have two bigons having a common edge;*
(ii) *G is a projection of the type represented in Figure* 2 **(Step 1(4)).**

Proof. We leave it to the reader to check (1) and that situations (2(i)) and (2(ii)) actually lead to a bigon with edges having the same name.

As for the converse in (2), let us start with a bigon B with edges g_1 and g_2; if the other copy of g_1 does not lie on a bigon too there cannot be any problem: if we eliminate all other bigons but B we have that there keep being only two segments named g_1. Similarly if the other copy of g_2 does not lie on a bigon. So, let us assume that both the other copies of g_1 and g_2 lie on bigons too, and moreover that the situation is not as in (2(i)). It is quite easily checked that if proceeding from B in a direction (and hence in both directions) we find an edge not lying on a bigon then no problem arises. The only possibility left (using the hypothesis that the projection be alternating) is (2(ii)). □

From now on we will exclude the situations (2(i)) (of which (1) is a special case) and (2(ii)); this is contained respectively in **Step 1(3)** (which excludes even more: we will see later why) and **Step 1(4)**. Remark that excluding situations like (2(i)) implies in particular that at most two edges join two fixed vertices of G. This is used in the following result, in which otherwise 'two edges' should be replaced by 'more than one edge', and 'one is deleted' by 'all but one are deleted'; it deals with the first condition stated above for the elimination of the bigons. Remark that the proof depends on the hypothesis that the projection could not be trivially simplified, namely on the fact that the edges of the bigon are alternating.

LEMMA 4.5. *The following conditions are pairwise equivalent:*

(1) *an orientation of the segments g_i can be chosen in such a way that in G_+ and G_- the two edges of each bigon have the same orientation;*
(2) *it is possible to attach to each vertex of G a symbol ⊙ or ⊗ in such a way that different symbols are attached to the vertices of each bigon;*
(3) *let $\tilde{G}$ be the subgraph of G obtained in the following way: the vertices are the same; if two edges join two vertices, one of them is deleted; if only one edge joins two vertices then it is deleted; then $\tilde{G}$ contains no non-trivial simple simplicial loop with an odd number of vertices;*
(4) *if $\tilde{G}$ is as above then $\tilde{G}$ contains no non-trivial simplicial loop with an odd number of vertices (each counted the number of times it is touched).*

Proof. Equivalence of (1) and (2) follows immediately from the explicit way of constructing G_+ and G_- stated in Lemma 4.2 (G_- is now reversed, but of course this makes no difference).

(2) $\Rightarrow$ (3). Assume by absurd that in $\tilde{G}$ there exists a simple simplicial loop with vertices $v_1, v_2, \ldots, v_{2n+1}$. If for instance v_1 has $\odot$ attached to it (in G) then v_2 has $\otimes$ attached to it (by construction of $\tilde{G}$ there exists a bigon with vertices v_1 and v_2). Similarly we go on and we finally have that v_{2n+1} has $\odot$ attached to it: a contradiction.

(3)$\Rightarrow$(4) is a very easy general property (depending on the fact that if the sum of two numbers is odd then one of the numbers is odd).

(4)$\Rightarrow$(2). By construction of $\tilde{G}$ it suffices to attach symbols $\odot$ or $\otimes$ to its vertices in such a way that different symbols are attached to the endpoints of each segment. In each component of $\tilde{G}$ we choose a vertex v_1 and for any other vertex v we define $n(v)$ as the least number of vertices of a simplicial path joining v to v_1. Then we attach $\odot$ to v if $n(v)$ is odd and $\otimes$ otherwise. (4) easily implies that things match up. $\square$

It is easily seen that the assumptions of **Step 1(1, 2, 3, 4)** together imply the equivalent conditions of the previous result; furthermore we assume the unbounded region (defined by G) is not a bigon. (This is needed to make the elimination of the bigons automatic starting from G_+ and G_-; the verification is in **Step 1(5)**.) We keep denoting by G_+ and G_- the labelled graphs obtained by eliminating the bigons from the old G_+ and G_-, according to the procedure described in **Step 5(1, 3)**. We leave the remainder of the notation unchanged; in particular S^2_+ and S^2_- are spheres containing G_+ and G_- respectively.

Remark that in both S^2_+ and S^2_- each region keeps being homeomorphic to the closed disc (it is easily checked that elimination of bigons does not affect this property).

Since in the sequel we are going to use the identification between R_i^+ and R_i^- we must add something to our construction. In fact we previously remarked that (before eliminating the bigons) the labelling of ∂R_i^+ and ∂R_i^- suffices to determine this identification (each g_j appeared at most once on them); unfortunately this is not true any more after the elimination of the bigons. For instance from the situation of Figure 35 we get respectively in G_+ and G_- the situations at the left- and right-hand sides of Figure 36 and then there exists no sensible way to choose the identification between R_i^+ *and* R_i^-.

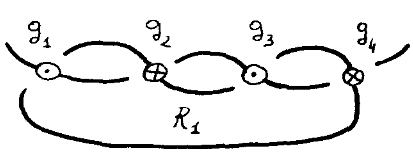

Fig. 35. After the elimination of the bigons the identification between R_i^+ and R_i^- must be specified. An example.

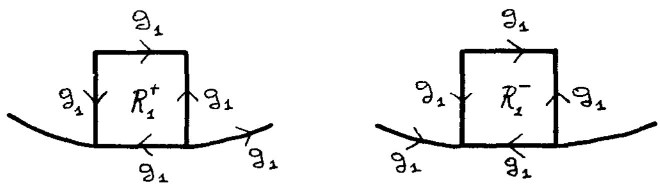

Fig. 36. Example of Figure 35 continued.

Then we go back to the situation before eliminating the bigons and we choose a vertex on each region R_i^+ and correspondingly one on R_i^-; we label these vertices by little arrows inside the region and pointing towards it, as shown in Figure 37.

Fig. 37. The little arrows we add to remove ambiguity.

A non-ambiguous way to make this choice is described in **Step 3(2)** and **Step 4(2)**. Then when eliminating the bigons we keep these little arrows (of course not inside the bigons) as stated in **Step 5(2)**. The identification between R_i^+ and R_i^- can now be described as follows: (i) the vertices of R_i^+ and R_i^- towards which the little arrows point are identified; (ii) ∂R_i^+ and ∂R_i^- are identified starting from the above vertices and following ∂R_i^+ in a counterclockwise sense and ∂R_i^- in a clockwise sense, in such a way that edges are identified to edges; (iii) the identification is extended to the interiors.

The fact that ∂R_i^+ and ∂R_i^- must be followed in opposite senses depends on the fact that G_- has been reversed. Remark that by the very example above the specification of the senses to be followed is necessary; the orientation of the g_i's may not be sufficient.

The next step in our construction (corresponding to **Step 6**) is to add some edges to G_+ and G_- in such a way that all the resulting regions (even the unbounded one) are triangles; we do this in such a way that the region R_i^- is divided in the same way as R_i^+ under the prescribed identification (i.e. we first divide R_i^+ and then we repeat for R_i^-; it is easily checked, using the description of the identification we just gave, that the method described in **Step 6** actually allows us to do this). Of course there are several ways for adding edges: if R_i^+ has $p \geqslant 4$ vertices we add inside R_i^+ edges $e_1, \dots, e_{p-3}$ in such a way that:

(a) $e_i \cap e_j$ consists of at most a common endpoint for $i \neq j$;
(b) e_i does not have both the endpoints in common with an edge of R_i^+;

(it is easily checked that these are necessary and sufficient conditions for a set of edges to divide R_i^+ into triangles without adding vertices); remark that a finite number of possibilities is given.

Let us denote once more by G_+ and G_- the graphs obtained by adding edges as described. We give the added edges names $g_{m+1}, g_{m+2}, \ldots$ (assuming m to be the greatest index previously appearing) both in G_+ and G_-: of course the edges to be identified under the identifications $R_i^+ = R_i^-$ will have the same name. One should also give new names to the triangles in order to keep in mind the way they must be glued together (the labelling of the edges may not be sufficient); however, since we are going to show that for writing down the equations we do not require this, we simplify the notation and give no names to the triangles.

Let us assume now that G_+ and G_- define triangulations of S^2_+ and S^2_-; remark once again that we keep having that the triangles are homeomorphic to the closed disc (adding edges as described does not affect the property that the regions be closed discs) so that there cannot be any forbidden self-adjacency on a triangle, but forbidden adjacencies between different triangles or edges may exist. It is easily checked that the condition that G_+ and G_- triangulate S^2_+ and S^2_- is equivalent to the fact that they do not contain pairs of edges having both the endpoints in common; this is assumed in **Step 6**. We will explain later why we make this assumption.

We are ready to describe the final step of our construction and to write down the equations. The point is to extend the triangulations of S^2_+ and S^2_- defined by G_+ and G_- to triangulations of B_+ and B_-, without adding vertices. This is done in the following way (we refer to B_+ for the notation): a vertex v^+ of the unbounded triangle is chosen and lifted over the horizontal plane, while all the triangles not containing v^+ are left on the horizontal plane. Then a straight edge is drawn from v^+ to any vertex on the plane, and a cone from v^+ is taken over the edges on the plane (the cone being then a triangle), as suggested by the 3-dimensional pictures of Figure 38.

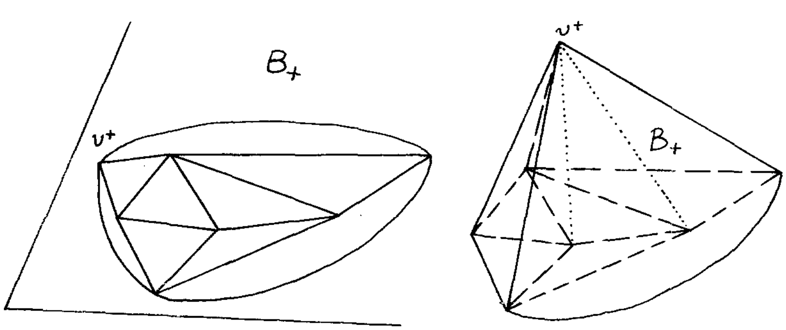

Fig. 38. The cone construction.

In Figure 38 the dotted edges are those we must add, while the others already existed; we did not draw the triangles. It follows that B_+ is decomposed into tetrahedra by adding:

(a) as many edges as the vertices of G_+ being not joined to v^+;
(b) as many triangles as the number of edges in G_+ not containing v^+.

The resulting tetrahedra are as many as the triangles in S^2_+ not containing v^+ (they are explicitly given by the cones from v^+ over these triangles).

Of course the same method works for B_- (and we denote by v^- the chosen vertex of the unbounded triangle). Let us recall that in an ideal tetrahedron the modulus is the same along opposite edges. Then in order to determine very easily all the moduli it suffices to number both in G_+ and G_- the triangles not containing the chosen vertex and write inside the ith one (starting from any edge and proceeding counterclockwise) the symbols given in Figure 39. As usual u' stands for $1/(1-u)$ and u'' for $1-1/u$. This construction corresponds to **Step 7** of the algorithm.

Fig. 39. The symbols to add inside the triangles; the left-hand side corresponds to G_+ and the right-hand side to G_-.

First compatibility equations. We want to write the compatibility equations for the lastly added edges (not appearing in the 2-dimensional pictures). Of course the procedure is the same in B_+ and B_-; we refer to B_+ for the notation; we know such an edge joins v^+ with a vertex v which, by construction, has around it triangles with symbols z_i, z_i', z_i'' written in. The corresponding 3-dimensional situation is described in Figure 40.

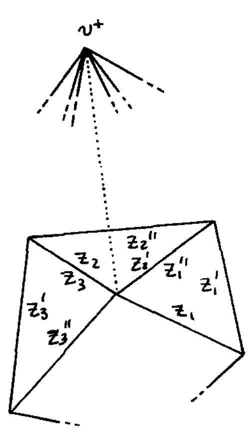

Fig. 40. First compatibility equations: the edges we do not see.

It follows that we must set to 1 the product of the symbols written inside the triangles having vertex v near the edge opposite to v.

Second compatibility equations. We want to write the compatibility equations for an edge g_i; if in G_+ a copy of g_i (i.e. an edge labelled with the symbol g_i) has not v^+ as endpoint but it belongs to a triangle containing v^+, then only one tetrahedron in B_+ contains it and the corresponding modulus is just the only symbol written near it; if it is not the edge of a triangle containing v^+ then the tetrahedra containing it in B^+ are two, and the corresponding moduli are the two symbols written near it; finally, if it has v^+ as endpoint and v is the other endpoint then it is contained in as many tetrahedra as the triangles containing v and not v^+, and the corresponding moduli are those written inside these triangles near the edge opposite to v. The situation in the three cases is represented in Figure 41.

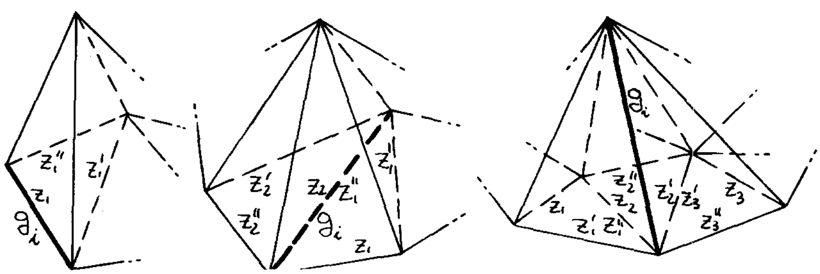

Fig. 41. Second compatibility equations.

These remarks lead quite naturally to the following conventions as in **Step 7**:

(a) if an edge in G_+ (or G_-) is not the edge of a triangle containing v^+ (resp. v^-) then when speaking of 'symbol written near it' we actually mean the product of the two symbols written near it;

(b) if an edge in G^+ (or G_-) has endpoints v^+ (resp. v^-) and v, we write near it the product of the symbols written inside the triangles containing v but not v^+ (resp. v^-) near the edge opposite to v.

Then the compatibility equation for g_i is obtained by setting to 1 the product of the symbols written near all the existing copies of g_i in G_+ and G_-.

The description of the construction and the compatibility equations is now complete.

We go back now to our assumption that G_+ and G_- triangulate S^2_+ and S^2_-. Let us assume for example that B_+ is not triangulated by G_+; this implies that in G_+ we have two vertices v_1, v_2 being joined by at least two edges. If v^+ is one of them then our cone construction does not even make sense. Assume

v^+ is not v_1 or v_2 and assume the compatibility equations (which make sense now) have a solution; then we would have in $\mathbb{H}^3$ two ideal tetrahedra having in common three vertices (v_1, v_2 and v^+) but not the face containing them, and of course this is absurd.

A necessary condition for having the possibility of adding edges to G_+ and G_- in such a way that they triangulate S^2_+ and S^2_- is that in G_+ and G_- we do not already have pairs of edges having both the endpoints in common. This means that if at the beginning in G one had a pair of edges with common endpoints then he has eliminated one of them, i.e. that these edges were the boundary of a bigon; this is assumed in **Step 1(3)**.

Such condition does not seem to be sufficient, as in G_+ and G_- (before adding edges) one could have several multiple adjacencies between regions and then be forced to add edges to G_+ in some particular way; the hard point is that even if one succeeds in adding edges to G_+ in order to triangulate S^2_+, these edges copied in G_- may not work, as adjacencies are different.

As a conclusion of the paragraph we deduce from the above discussion another good reason for assuming the projection to be alternating: if this were not the case several consecutive multiple adjacencies between regions along edges would have appeared since the very first steps.

5. Explanation of the algorithm: completeness equations

We want to describe now how to obtain the equations implying the completeness of the hyperbolic structure associated to the moduli $z_1, \ldots, w_1, \ldots$ solving the equations previously obtained. According to what we stated in Section 3 we must compute the derivative of the holonomy of the meridians of the toric links of the removed vertices in $S^3 \backslash L$, the holonomy being referred to the similarity structure induced on this toric links by $z_1, \ldots, w_1 \ldots$. The calculation is performed using the rule mentioned in Section 3.

We number the components of the link and for the ith one we fix in G (the initial projection of the link, a graph) an edge belonging to it (like in **Step 2**); we consider around such an edge a standard meridian m_i of a tubular neighborhood of the component in question, as represented in Figure 42.

We want to describe (a loop isotopic to) m_i in S^2_+ and S^2_- (we obtain this by adding something to G_+ and G_-); by the very construction such a loop will be a simple simplicial loop with respect to the triangulation of the toric link of the removed vertex. According to the rule mentioned in Section 3 we must keep track not only of the arcs of m_i (contained in the 2-cells) but also of the

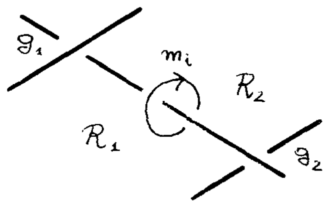

Fig. 42. The standard meridian around a component of the link.

way it touches the 1-cells. Thus we introduce the symbolic conventions of Figure 43: its left- and right-hand sides mean respectively that:

Fig. 43. Conventions about m_i.

(a) an arc of m_i is contained in the 2-cell R in the toric link of the vertex v and it goes from e_1 to e_2;
(b) m_i touches the edge e in the toric link of the vertex v, though in the neighborhood of the touching point it is not contained in R_1 or R_2; moreover it passes first close to R_1 and then close to R_2.

We go back to the situation considered above (Figure 42). While performing the construction described in Section 4 we can modify m_i to the loop represented in Figure 44 (this loop is chosen to be contained in the modified regions).

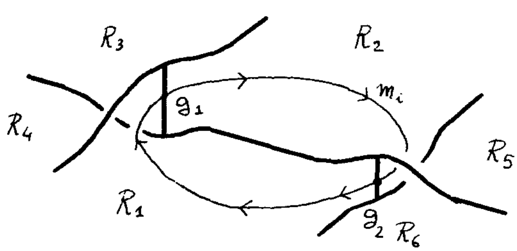

Fig. 44. How to represent the meridian by a loop in the 2-skeleton.

It easily follows that, with the above symbolism, in S^2_+ and S^2_- this loop is represented respectively by the left- and right-hand sides of Figure 45.

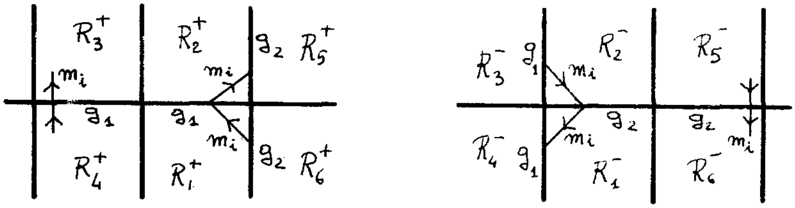

Fig. 45. Representation in S_+^2 and S_-^2 of the loop of Figure 44.

In Figure 45 the representation of S_-^2 refers to the situation before reversing; this implies that the rules to be followed are those described in **Step 3(3)** and **Step 4(3)**.

Now, we must discuss what happens when eliminating the bigons; from now on we refer to S_+^2 for the notations, the rules being identical for S_-^2. It is quite easily checked that if an arc of m_i is contained in the bigon then we are left a simple passage, as shown in Figure 46.

Fig. 46. What is left when eliminating a bigon containing an arc.

Remark that, according to our symbolism, if there is another arc of m_i adjacent to the one in question we do not need to add new symbols, we only delete the bigon; then the resulting rules are those described in the upper-left and upper-right of Figure 9 of **Step 5(4)**. Now, if m_i has only a passage on one of the edges e_1 of a bigon we must consider the fact that when identifying to e_1 the other edge e_2 of the bigon another passage appears on the other existing copy of e_2; moreover if m_i passes from the left to the right of e_1 (with respect to its orientation) then the same must hold for e_2; the resulting rules are those described in the lower-left and lower-right parts of Figure 9 (**Step 5(4)**).

These procedures must be followed with some care, since we must keep track of all the passages appearing (i.e. as we will explain soon, we must not allow the order of elimination of the bigons make us lose passages which could have appeared with a different choice of order). Assume for instance that we have the situation of Figure 47.

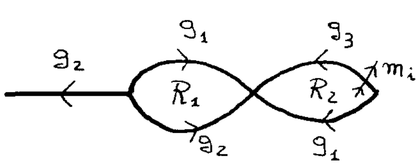

Fig. 47. When eliminating the bigons we must take care not to lose passages of m_i. An example.

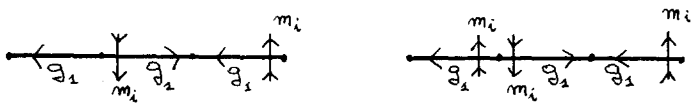

Fig. 48. Example of Figure 47 continued.

Then if we use the above procedures and eliminate: (a) first R_1 and then R_2; (b) first R_2 and then R_1; we get respectively the left- and right-hand side of Figure 48.

Of course the right result is the latter; in fact if at the beginning we have a passage of a certain m_i on an edge g_j, such a passage must cause a passage of m_i to appear on all edges identified to g_j during the elimination of the bigons. It is quite easily checked that the recursive method described in **Step 5(4)** actually allows us not to lose passages of m_i.

Remark that according to the rules described in order to keep the construction as simple as possible it is convenient (as stated in **Step 2**):

(a) to choose the edges of G where to put the meridians as edges of bigons;
(b) not to choose these edges as having only one vertex on a bigon.

It is worth remarking that near each vertex of G_+ and G_- we have arcs and passages of at most one m_i: in fact if $i \neq j$ we have that m_i and m_j lie on the toric links of different vertices. This guarantees that the picture does not become too complicated.

The next step is to discuss what happens when adding edges to G_+ in order to make the polygons become triangles; of course simple passages are not affected, while an arc may be cut into several arcs and one gets the rule described in Figure 10, referring to **Step 6.**

Thus in the final pictures of G_+ and G_- (with all the moduli written in) we have now also these symbols of passages and arcs of the m_i's. In order to compute the derivative of the holonomy of m_i we first remark that it has as many vertices as arcs (it is a loop); the number of arcs equals the number of arcs labelled by m_i we have either in G_+ or in G_- (of course the same numbers of arcs labelled by m_i appear in G_+ and G_-, as the triangles in S^2_+ and S^2_- are identified in pairs). This allows us to calculate the factor $+1$ or -1 in the derivative of the holonomy of m_i, as mentioned in Section 3.

Then we must establish what moduli m_i leaves on the right (we give the toric links their natural orientation). Remark that we are now only interested in the vertices of m_i (its intersection with the 1-cells) so that we can view the situations of Figure 49 as identical.

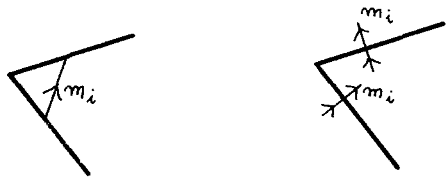

Fig. 49. These two situations can be identified when calculating the derivative of the holonomy.

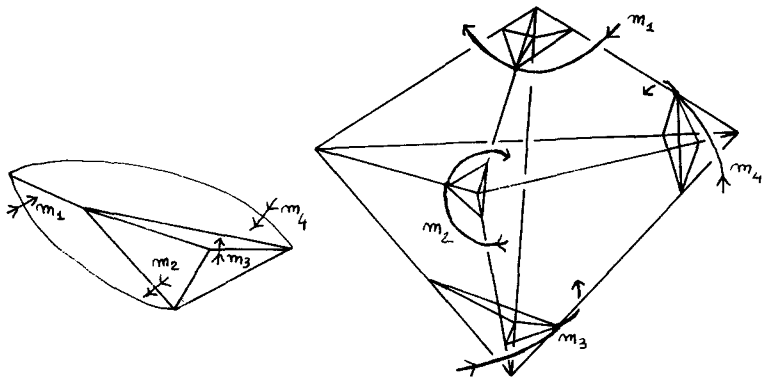

Fig. 50. The possible positions of the passages of the meridian.

Figure 50 summarizes all possible positions of the passage.

Remark that in the 2-dimensional part of Figure 50 we always have that m_i leaves on the left the vertices to which the passages are referred; in the 3-dimensional part it always leaves the depicted triangles on the right (with respect to the natural orientation). It follows that (up to the factor ± 1 already discussed) the derivative of the holonomy of m_i is obtained in the following way: in G_+ and G_- we select all the edges e touched by m_i such that the passage of m_i on e leaves the vertex it is referred to on the left; then we take the product of the symbols written near all these edges e. Completeness equations are obtained by setting the derivative of the holonomy equal to 1 and the sum of the arguments of the moduli appearing in it equal to π times the number of vertices.

Appendix

In this appendix we are going to prove the following result.

THEOREM A.1. *Given any regular projection of an arbitrary link it is possible to add components to the link in order to get a link represented by a projection to which the algorithm applies.*

We prove this result constructively, i.e. we describe explicitly how to add the components. So, let us start with a regular projection of a link (a graph with above–below specifications at the crossings); we assume the link has been isotoped in order to satisfy the initial conditions; we also assume the projection is not a plain circle or that of Step 1(4) (otherwise we add something else, no matter what). During the construction we will leave it to the reader to check that, once a property has been satisfied, the further steps we make do not waste such a property. In the pictures the new components added are boldfaced.

The first step is to make the regions become discs; this is done very easily and we do not loiter on this point.

The next step is to make the projection become alternating. We say an edge is alternating if it passes once above and once below at its endpoints; we say it is above (below) if it passes above (below) at both the endpoints. The idea is just to cut each edge being not alternating into two edges being alternating; the following lemma implies that it is possible to do this.

LEMMA A.2. *Let $e_1, \ldots, e_n$ be the edges of a region, consecutively numbered, and assume for instance e_1 is below. Then there exist integers*

$$1 = p_1 < q_1 < p_2 < q_2 < p_3 < \cdots < p_m < q_m \leqslant n$$

such that the e_{p_j}'s are below, the e_{q_j}'s are above and the other e_j's are alternating.

Proof. We confine ourselves to checking that the first non-alternating edge we find after e_1 is above, and moreover that there exists one. We omit pictures as the situation is easily figured out. Since e_2 passes above e_1 we must have that e_2 is either alternating or above. In case it is alternating the same remark applies to e_3. In order to conclude we only need to remark that we cannot close up and find only alternating edges: if $e_2, \ldots, e_{n-1}$ are alternating then e_n is above. □

Now, inside each region having a non-alternating edge, we add (with the symbolism of the lemma) an edge with endpoints on the middle points of e_{p_j} and e_{q_j}, passing below the former and above the latter; we do this for $j = 1, \ldots, m$ in such a way that the added edges do not meet. For instance with $n = 8$, $p_1 = 1$, $q_1 = 2$, $p_2 = 4$ and $q_2 = 6$ we obtain Figure 51.

Of course if we do the same for all the regions things match up; remark that we are adding plain unlinked circles (but obviously these circles are not unlinked from the previous link).

The preliminary conditions and those of Step 1(1, 2, 4) are now satisfied. The next step is to remove situations in which one has a pair of edges with

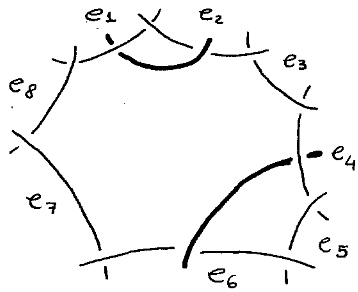

Fig. 51. How to make a projection alternating.

common endpoints (in particular, bigons). This allows the conditions of Step 1(3, 5, 6) to be satisfied. This is done by adding on each edge having both the endpoints in common with another edge two plain unlinked circles as described in Figure 52.

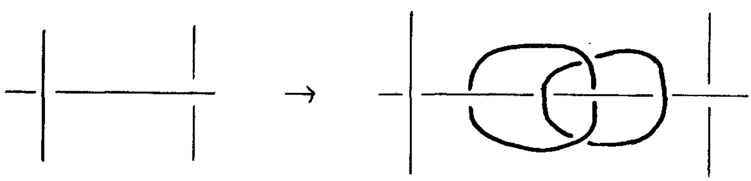

Fig. 52. How to avoid bigons.

In order to conclude our construction we prove the following elementary fact.

LEMMA A.3. *Assume the projection satisfies the preliminary conditions and those of Step* 1. *Moreover assume that*:

(i) *there exists no bigon*;
(ii) *two regions have at most two common vertices and if they do these vertices are the endpoints of a common edge.*

Then Step 6 *can be performed.*

Proof. Since we have no bigons to eliminate, the graphs G_+ and G_- we have before Step 6 are just the same as the initial graph (straight and reversed). It easily follows from (ii) that no matter how we triangulate the regions we always get a triangulation of both S^2_+ and S^2_-. □

Since condition (i) of the lemma is already satisfied by the previous steps we made we only need to satisfy (ii). We begin to eliminate multiple adjacencies of regions along edges; assume R_1 and R_2 are adjacent along more than one edge, and call e one of them; then we add two plain unlinked circles as represented in Figure 53.

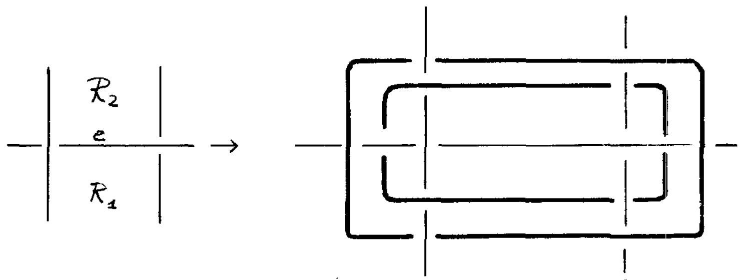

Fig. 53. How to remove multiple adjacencies along edges.

Since it is easily checked that such a construction does not introduce new multiple adjacencies an induction argument allows us to prove that one can always add components in such a way that the regions have at most one common edge. We are left to remove the situations in which two regions have a common vertex being not the endpoint of a common edge. This is easily done as suggested by Figure 54 (once again we are adding two plain unlinked circles).

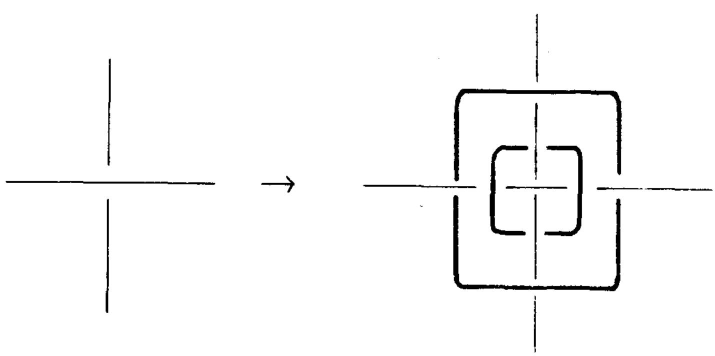

Fig. 54. How to remove multiple adjacencies at vertices.

REFERENCES

[Ad1] Adams, C. C., 'Thrice-punctured spheres in hyperbolic 3-manifolds', *Trans. Amer. Math. Soc.* **287** (1985), 645–656.

[Ad2] Adams, C. C., 'Augmented alternated link complements are hyperbolic', in London Math. Soc. Lecture Notes Series, No. 112 (1986) (D. B. A. Epstein ed.), pp. 115–130.

[Be-Pe] Benedetti, R. and Petronio, C., 'Lectures on hyperbolic geometry' (to appear in the series Universitext, Springer-Verlag).

[Ep-Pe] Epstein, D. B. A. and Penner, R. C., 'Euclidean decomposition of non-compact hyperbolic manifolds', *J. Diff. Geom.* **27**(1) (1988), 67–80.

[Ma-Fo] Matveev, S. V. and Fomenko, A. T., 'Constant energy surfaces of Hamiltonian systems, enumerating of three-dimensional manifolds in increasing order of complexity, and computation of volumes of closed hyperbolic manifolds', *Uspekhi Mat. Nauk* **43**(1) (1988), 5–22; *Russian Math. Surveys* **43**(1) (1988), 3–24.

[Me1] Menasco, W. W., 'Polyhedra representation of links complements', *Contemp. Math.* **20** (1983), 303–325.

[Me2] Menasco, W. W., 'Closed incompressible surfaces in alternating link and knot complements', *Topology* **23**(1) (1984), 37–44.

[Ne-Za] Neumann, W. D. and Zagier, D., 'Volumes of hyperbolic 3-manifolds', *Topology* **24**(3) (1985), 307–332.

[Th1] Thurston, W. P., *The Geometry and Topology of Three-Manifolds.* Princeton University Mathematics Department, 1979.

[Th2] Thurston, W. P., 'Three dimensional manifolds, Kleinian groups and hyperbolic geometry', *Bull. Amer. Math. Soc.* **6**(3) (1982), 357–381.

[Th3] Thurston, W. P., 'Incompressible surfaces in 2-bridge knot complements', *Invent. Math.* **79** (1986), 225–246.

[Th4] Thurston, W. P., 'Hyperbolic structures on 3-manifolds', *Ann. Math.* **124** (1986), 203–246.

Author's address:

Carlo Petronio,
Scuola Normale Superiore,
Piazza dei Cavalieri, 7,
56126 Pisa,
Italy.

(Received, October 15, 1991)