ON THE CLASSIFICATION OF HOMEOMORPHISMS OF 2-MANIFOLDS AND THE CLASSIFICATION OF 3-MANIFOLDS

BY

GEOFFREY HEMION

University of Bielefeld, Federal Republic of Germany

Introduction

Given two topological spaces, is it possible to determine whether they are homeomorphic? This is the homeomorphism problem and most work in topology is directed toward some aspect of the homeomorphism problem. A plan for solving the homeomorphism problem for "most" 3-manifolds has been developed by Wolfgang Haken. However, a certain very special step in this plan has eluded proof. The problem of providing a proof for this special case amounts to the problem of classifying homeomorphisms of compact, orientable 2-manifolds. In this paper a method for classifying homeomorphisms of compact, orientable 2-manifolds will be given, and hence it will be possible to classify all compact, orientable, irreducible, boundary irreducible, sufficiently large 3-manifolds. Hence "most" 3-manifolds of interest can be classified, including all knot and link spaces.

Haken developed the theory in his series of papers: [1]–[5]. In [11], Schubert has explained the essential points of Haken's theory of normal surfaces. Waldhausen [12] has written a summary of the classification procedure, using the recent results of Johannson [6], [7].

The conjugacy problem for self-homeomorphisms of compact, orientable surfaces

Let the surface S be compact and orientable, and let f, g be two homeomorphisms of S onto itself. Assume that there exists a homeomorphism h of S onto itself such that $h^{-1}fh$ is isotopic to g. In this case, f and g are said to be conjugate. Given two homeomorphisms such as f and g, the conjugacy problem asks whether or not they are conjugate. In order to complete Haken's program for the classification of sufficiently large 3-manifolds, we need to prove a result which is slightly stronger than the conjugacy problem. Namely, if $\partial S \neq \emptyset$ and if f and g agree on ∂S, then our problem is to determine whether f and g

are conjugate by a homeomorphism h, such that $h^{-1}fh$ is isotopic to g by an isotopy which leaves ∂S fixed. As far as the classification of 3-manifolds is concerned, the interesting case is that in which $\partial S \neq \emptyset$. Hence we shall construct our proof for this case. It is possible to extend the proof to the case $\partial S = \emptyset$, but to do so requires many changes in the details of the arguments. However, these changes are in principle straightforward, and we will sketch very briefly in section 3.2 what they are. In order to solve the conjugacy problem we will first prove a certain theorem.

A statement of the main theorem

Let χ_S be the Euler characteristic of the surface S. (We choose the sign of the Euler characteristic so that it becomes larger as the surface becomes more complicated.) The solution of the conjugacy problem is well known when $\chi_S \leq 0$. The solution of the conjugacy problem is also well known when S is a disc with two holes. Hence we shall assume that $\chi_S \geq 1$, and that S is not a disc with two holes.

1. The *size* of a homeomorphism of S onto itself will be defined in section 1.3. Let f be a homeomorphism of S onto itself. The size of f, $d(f)_\Delta$, will be defined to be a certain positive integer. For any positive integer there are essentially only a finite, determined number of homeomorphisms of S onto itself whose size is less than that number.

2. Let d be a curve on S which is either closed, or is such that $d \cap \partial S = \partial d$. If there exists an integer, $m \neq 0$, such that $f^m(d)$ is properly homotopic to d, then d is a *periodic curve* under f. When using the term "periodic curve" we will generally ignore the trivial cases: curves or arcs which are null-homotopic or properly deformable into ∂S.

We will be concerned with the production of a certain positive integer, $N(f, g)$. The number $N(f, g)$ depends only upon $d(f)_\Delta$, $d(g)_\Delta$, χ_S, and r, where r is the number of components of the boundary of S. The number which will be found is certainly very large. If one were to be interested in solving the conjugacy problem *in practice* then this number would be unmanageable. However, with ingenuity it would undoubtedly be possible to bring it within reasonable limits.

The main theorem is:

THEOREM. *Suppose that in S there are no periodic curves under f, except those which are null-homotopic or properly deformable into ∂S. Let h be a homeomorphism of S onto itself such that fh is isotopic to hg. Then for some integer n, and some homeomorphism h' isotopic to $f^n h$, it is true that $d(h')_\Delta \leq N(f, g)$.*

Although this theorem may seem to solve the conjugacy problem in only a somewhat special case—S having no periodic curves, and the question of whether ∂S is left fixed

under the isotopy being disregarded—the theorem does actually solve it in general, when combined with known results. This will be shown in section 3.3.

The proof of the main theorem is divided into three parts. In the first part we prove that under the hypotheses of the theorem there exist "small" isotopies of f and g, and some isotopy of h, so that after the homeomorphisms have been altered by these isotopies, $fh = hg$, exactly. (Note that in general the homeomorphism which results from an isotopic deformation of the identity map will be called simply an "isotopy" for brevity.)

In the second part we produce a closed curve c on S, which is "small" in a certain sense. Proving the existence of the curve c is the main idea behind the entire proof. In constructing c we make use of a series of elementary geometric constructions. There are, however, a number of different cases to consider, which relate to each other in various ways, producing a certain amount of complexity. Hence the reader may be advised to first read the third part, where the theorem is proved quite simply, taking the existence of c for granted.

We will work in the piecewise linear category throughout.

1. Isotopies of f, g, and h

1.1 Some preliminary results and definitions

Let S be a compact, orientable surface with boundary, whose Euler characteristic χ_S is greater than zero, and which is not a disc with two holes. Let the homeomorphisms f, g, of S onto itself be given. Assume that there exists a homeomorphism, h, of S onto itself such that fh is isotopic to hg. We wish to prove that either there exists a "small" such h, (i.e. such that $d(h)_\Delta \leqslant N(f, g)$), or else there exists a periodic curve under f. (Clearly, if we produce a periodic curve under f, then since f and g are conjugate, there exists a periodic curve under g.)

The small h which we produce will be realized by a function of the form $f^m \cdot h_{\text{orig}}$, for some m. Here, h_{orig} is the original h which was given as part of our assumptions, and which may be "large". As we proceed through the proof we will change our homeomorphisms, "improving" them, and to make the notation easier we will simply call the new versions f, g, and h, again. When we wish to refer back to the unimproved homeomorphisms we will use notation like h_{orig}. Note that, trivially,

$$(f^m \cdot h_{\text{orig}})^{-1} \cdot f \cdot (f^m \cdot h_{\text{orig}}) = h_{\text{orig}} \cdot f \cdot h_{\text{orig}},$$

so that the new h will automatically conjugate f and g.

In order to prove our theorem we will work in the universal covering space of S. Since $\chi_S \geqslant 1$, the universal covering space is represented by the hyperbolic plane, H.

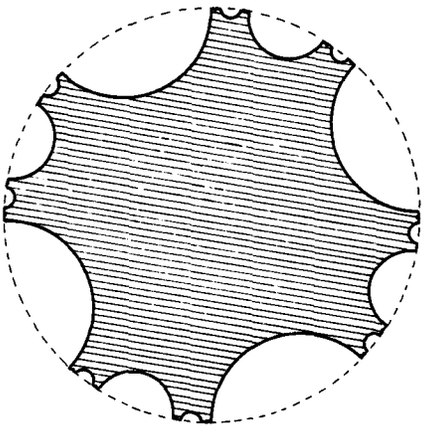


Fig. 1

The hyperbolic plane may be pictured in the following way: Imagine the surface S as being split into a disc, Δ, by a finite number of disjoint, simple arcs, whose ends are in ∂S. The disc is "flattened" to form a polygon, and then infinitely many of these polygons are glued together along corresponding edges to form the hyperbolic plane (Fig. 1).

The reader is referred to Nielsen's papers for a description of the hyperbolic plane. We shall use the following elementary results which appear in Nielsen's papers [8] and [10].

(i) If α is a simple closed curve on S which is not contractible, then α lifts to a system of simple, non-intersecting arcs on H. If $\bar{\alpha}$ is one such arc then each end of $\bar{\alpha}$ converges to a point of bd H, the boundary of the hyperbolic plane, and those points are different. If $\bar{\alpha}'$ is another arc in the lifting of α then neither of its endpoints coincide with an endpoint of $\bar{\alpha}$.

(ii) If k is a homeomorphism of S onto itself, then a lifting, $\bar{k}$, of k to H, is such that the ordering of the points of bd H is preserved when bd H is considered to be acted upon by k in a natural manner. The mapping of bd H is continuous. If k^2 is taken, hence $\bar{k}^2$, then the circular order of the points on bd H, under $\bar{k}^2$, is preserved (not reversed).

1.2 The definition of the fundamental region

We define a system of simple closed curves, $\{\alpha_i\}$, on S: Let b be one of the boundaries of S. Let x_0 be a point of b. Then the curves $\{\alpha_i\}$ are defined to be disjoint, modulo x_0, and are in the interior of S, modulo x_0. They are arranged according to a certain pattern, as in Nielsen's papers, and generate, with b, the fundamental group of S (Fig. 2).

We adjoin to $\{\alpha_i\}$, also, simple arcs joining x_0 to the other boundary components of S,

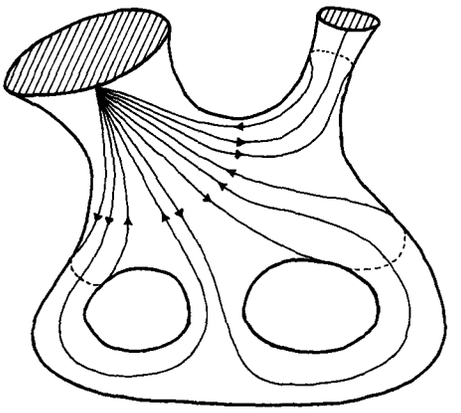


Fig. 2

which are disjoint from the other curves of $\{\alpha_i\}$, modulo x_0. Then for each boundary component, other than b, two curves from $\{\alpha_i\}$ are associated with it. One is parallel to the boundary, encircling it, and the other is an arc connecting it with x_0.

We may think of the curves from $\{\alpha_i\}$ as being in two classes; those with both endpoints in x_0 being "based loops", while those with only one endpoint in x_0 are "arcs".

Denote by Δ the disc gotten by splitting S along all the curves of $\{\alpha_i\}$ which are not boundary parallel. The lifting of Δ to H, $\tilde{\Delta}$, consists of discs with $2(1+\chi_S)+r$ sides. If we only count the sides which lift from curves of $\{\alpha_i\}$, (i.e. we do not count the liftings of boundary curves of S), then there are $2(1+\chi_S)$ sides for each disc of $\tilde{\Delta}$. Each such disc is called a "fundamental region of Δ".

Denote the lifting of x_0 by $\tilde{x}_0$, and the lifting of α_i by $\tilde{\alpha}_i$. Then each point of $\tilde{x}_0$ is a

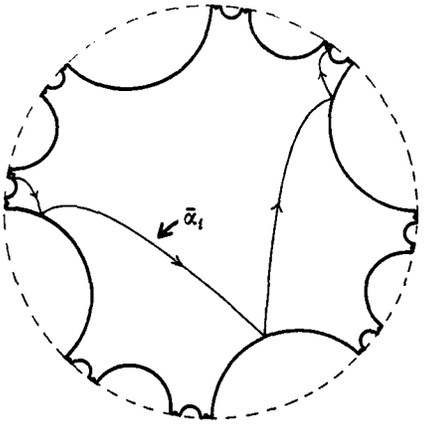


Fig. 3

"corner" of a number of fundamental regions of Δ. Note also that each point of $\bar{x}_0$ is on bd H. Further, if α_i is a based loop, having both ends in x_0, then $\bar{a}_i$ is a collection of arcs, each of which is a sequence of identical "segments" between adjacent intersections with $\bar{x}_0$. Each of these segments is a properly embedded arc on H, which splits H into two pieces (Fig. 3).

If α_i has one end in x_0 and one end in a boundary other than b, then $\bar{a}_i$ is a collection of disjoint, "short" arcs in H, each splitting H into two pieces.

1.3 The size of a homeomorphism

The size of a homeomorphism may now be defined. Let k be a homeomorphism of S onto itself. Let Δ be defined on S, and let δ be a fundamental region of $\bar{\Delta}$. Let $\bar{k}$ be a lifting of k to H. Then $\bar{k}(\delta)$ is a disc in H. Let $\delta_1, \delta_2, \ldots, \delta_N$ be a finite collection of fundamental regions of Δ, such that $\bar{k}(\delta) \subset \bigcup_{i=1}^{N} \delta_i$.

Definition. If the set $\{\delta_1, \delta_2, \ldots, \delta_N\}$ is chosen so that $\bar{k}(\delta) \subset \bigcup_{i=1}^{N} \delta_i$, and N is as small as possible, then N is the "positive size of k in Δ". The "negative size of k in Δ" is defined to be the positive size of k^{-1} in Δ. Then the "size of k in Δ", denoted $d(k)_\Delta$, is the greater of these two numbers.

(Note that in the sequel proofs will generally be given only for the "positive size" of homeomorphisms, it being understood that the proofs can be equally well carried out for the "negative size".)

Two important properties of the size of a homeomorphism can be immediately deduced.

LEMMA A. *Let k, l be homeomorphisms of S onto itself. Then $d(k \cdot l)_\Delta \leqslant d(k)_\Delta \cdot d(l)_\Delta$.*

LEMMA B. *Let N be a positive integer. Then there exist only finitely many isotopy classes of homeomorphisms of S onto itself which contain a homeomorphism of size no greater than N.*

Proof. Let δ be a fundamental region of Δ. Define the sequence $\{A_i\}$ of subsets of H as follows: $A_1 = \delta$, A_{i+1} = the union of the set of fundamental regions of Δ which meet fundamental regions in A_i.

If k is a homeomorphism of S onto itself then there is a lifting, $\bar{k}$, of k to H, such that $\bar{k}(\delta) \cap \delta \neq \emptyset$. If $d(k)_\Delta \leqslant N$ then we certainly have $\bar{k}(\delta) \subset A_N$.

By an isotopy involving rotations of the boundaries of S about themselves by less than one revolution, we may assume that the set of endpoints of arcs of $\{\alpha_i\}$ is taken onto itself by k. We may further assume that these isotopies do not increase the size of k in Δ, so that we still have $d(k)_\Delta \leqslant N$.

The set A_N contains only finitely many points which are liftings of endpoints of

$\{\alpha_i\}$. (These are the "corners" of fundamental regions of Δ.) But if two homeomorphisms have liftings which agree on the liftings of the set of endpoints of $\{\alpha_i\}$, then they are isotopic by an isotopy which leaves those endpoints fixed. Hence the lemma is true.

1.4 New choices for f, g, and h

We return to our homeomorphisms f, g, h, as given in 1.1. Take f^2, g^2, rather than f, g, to ensure that the circular order of the points of bd H remains fixed under liftings of f^2, g^2, respectively. Since $g^2 \cong (h^{-1}fh)^2 = h^{-1}f^2h$, our conjugating homeomorphism, h, remains unchanged. Having gotten the results for f^2, g^2, they imply the results for f, g. Namely: if there exists a periodic curve under f^2, then there exists one under f. If, on the other hand, $(f^2)^m \cdot h$ is "small" for some m, then it is still "small" with respect to f, g, and we still have g isotopic to $((f^2)^m)^{-1}f((f^2)^m h)$. (This will be made more precise later in the proof.)

It may be the case that f and g permute the boundaries of S. Let r be the number of boundary components of S. If we take the homeomorphisms $f^{r!}$, $g^{r!}$, rather than f, g, we will certainly have all boundaries of S being taken onto themselves by $f^{r!}$, $g^{r!}$. Therefore let us redefine f, g, to be the $2r!$-th powers of the origional f, g, respectively.

At this stage h may still permute the boundaries of S. That is to say that if we order the boundaries of S then an application of h represents a certain permutation of those boundaries. It may be proved, by induction on the number of boundaries, that there exists an orientation preserving homeomorphism, k, with $d(k)_\Delta \leqslant (2(1+\chi_S))^r$, and k permuting the boundaries in the same way as does h. (k may be constructed as a product of no more than r homeomorphisms, each of which exchange two boundaries.)

Let us, therefore, redefine f to be, $k^{-1}fk$. (Clearly f is conjugate to g if and only if $k^{-1}fk$ is conjugate to g.) We also redefine h to be, $k^{-1}h$. Then we will have:

$$g \cong h^{-1}fh = (k^{-1}h)^{-1}k^{-1}fk(k^{-1}h),$$

and so the conjugating relations hold for the new f, and h. Further, h now leaves the boundary curves of S invariant (and also f still leaves them invariant).

At the end of all these alterations we have:

$$d(f)_\Delta \leqslant (2(1+\chi_S))^r \cdot [d(f_{\text{orig}})_\Delta]^{2r!}, \quad \text{and} \quad d(g)_\Delta \leqslant [d(g_{\text{orig}})_\Delta]^{2r!}.$$

1.5 The elimination of trivial intersections

In this paragraph, f is altered by the application of a number of isotopies. At the end we have $\{f(\alpha_i)\}$ meeting $\{\alpha_i\}$ "nicely".

9 – 782904 *Acta mathematica* 142. Imprimé le 20 Février 1979

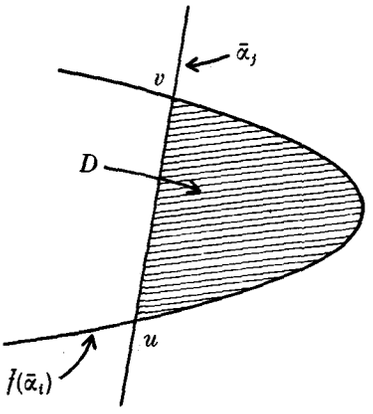


Fig. 4

Begin with an isotopy which is a rotation of the boundary, b, about itself by less than one complete revolution, so that, after this isotopy, f is the identity map on b. We do this for the other boundaries of S as well. It may be assumed that the size of this new f is no larger than that of the old f, in Δ.

Look at the covering space, H. Choose a curve, B, from the collection of curves, $\bar{b}$, which is the lifting of b. Since f is the identity on b, we may choose the lifting, $\bar{f}$, of f, which is the identity map on B.

By a general position argument we may assume that for all α_i, α_j, we have either $f(\alpha_i)$ coinciding with α_j, or else $f(\alpha_i)$ intersecting α_j in a finite collection of points. In the later case, if there is a "trivial intersection" between $f(\alpha_i)$ and α_j, then that intersection may be eliminated in the obvious way by an isotopy in S. (Fig. 4.)

Here, in H, there is an intersection between $\bar{\alpha}_j$ and $\bar{f}(\bar{\alpha}_i)$, such that a disc, D, is defined, whose boundary consists of one arc from each of $\bar{\alpha}_i$ and $\bar{f}(\bar{\alpha}_i)$. D does not meet $\bar{x}_0$, except possibly where the two arcs defining its boundary meet, at their ends (points u, v in the figure). By induction on the complexity of the intersections of $\{\alpha_j\}$ and $\{f(\alpha_i)\}$ it may be assumed that there are none of these intersections. At this stage we say that, "$\{f(\alpha_i)\}$ has no trivial intersections with $\{\alpha_j\}$". Note that if we multiply by f^{-1} we deduce immediately that $\{f^{-1}(\alpha_i)\}$ has no trivial intersections with $\{\alpha_i\}$.

If α_i is an arc from x_0 to a boundary, b', of S, other than b, then the situation illustrated below may occur, i.e. $f(\alpha_i)$ winds around the boundary b'. In this case a rotation of b' about itself produces an isotopy which, after an isotopy along the trivial intersections which may result, reduces the number of intersections of $\{f(\alpha_i)\}$ with $\{\alpha_i\}$. By induction on the complexity of the intersection of $\{f(\alpha_i)\}$ with $\{\alpha_i\}$, we may assume that both of the isotopies of this paragraph have been carried out as far as possible.

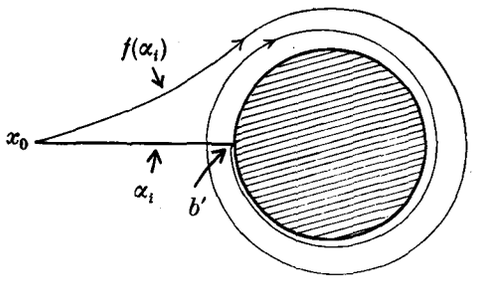


Fig. 5

1.6 Straightening the homeomorphisms

Choose α to be a curve of $\{\alpha_i\}$ which is not one of the arcs from x_0 to a boundary of S other than b. Looking at the covering space H, we choose a point, P, which is in $\bar{x}_0 \cap B$. Since B is left fixed by $\tilde{f}$, P is also left fixed.

In $\bar{\alpha}$, the lifting of α, let $\bar{\beta}$ be the curve which passes through P. Let β be one of the "segments" of $\bar{\beta}$ with an endpoint at P, i.e. β is a single covering of α on S. Let β_1, β_2, be the copies of β, adjacent to β, on B (Fig. 6).

Here, β_1, and β_2, are obtained from β by covering space transformations of H which take B onto itself and rotate H, one unit to the right and one unit to the left.

Let y be the endpoint of β that is not P; let y_i be the endpoint of β_i which is not on B, $i=1, 2$. It may be the case that the endpoint of $\tilde{f}(\beta)$, which is not P, namely $\tilde{f}(y)$, does not

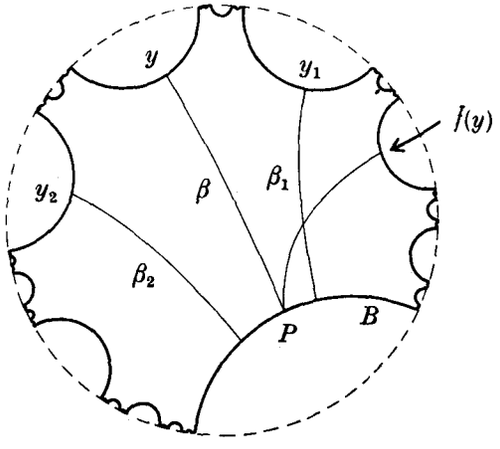


Fig. 6

fall between y_1 and y_2 on bd H, as illustrated above. In this case we alter f by an isotopy which is a rotation about b a certain number of times. At the end of this isotopy we still have f being the identity on b, and if $\bar{f}$ is redefined to be the old $\bar{f}$, acted upon by the lifting of the isotopy, and then acted upon by a covering space transformation taking the image of P back to P, we will have $\bar{f}(y)$ falling between y_1 and y_2 on bd H.

Having done this we go back to paragraph 1.5 and alter things so that $\{f(\alpha_i)\}$ meets $\{\alpha_i\}$ nicely once again. Note that the size of f cannot increase after all these operations.

All of the things which have been done to f in the last two paragraphs may be done, also, for g and h. Hence we assume that they have also been altered by isotopies and their sizes in Δ have not increased. The liftings of g, h, which leave B fixed are denoted by $\bar{g}$, $\bar{h}$, respectively.

1.7 Agreement between $\bar{f}\cdot\bar{h}$ and $\bar{h}\cdot\bar{g}$ on $\bar{\mathbf{x}}_0$

In this and the suceeding few paragraphs our goal is to find isotopies to f, g, h, such that, at the end of these isotopies, $g = h^{-1}fh$, exactly. Furthermore the size of g in Δ will not be increased by more than $12(\chi_S+1)$, (r = number of boundaries of S), while the size of f in Δ is not increased.

In order to do this we begin by examining $\bar{f}\cdot\bar{h}(\bar{\beta})$. It is a "broken arc" in H, as illustrated below. We may think of the ends of $\bar{f}\cdot\bar{h}(\bar{\beta})$ as "going off to infinity" in H, as indeed they do in the metric on H which is supplied by Nielsen. For us, the important thing about these endpoints is the following: Any isotopy of S lifts to an isotopy which does not move the endpoints of $\bar{f}\cdot\bar{h}(\bar{\beta})$ (Fig. 7).

Hence we examine the endpoint of $\bar{f}\cdot\bar{h}(\bar{\beta})$. Since $\bar{h}\cdot\bar{g}$ is isotopic to $\bar{f}\cdot\bar{h}$, there must exist a curve, let us call it $\bar{\beta}_n$, in $\bar{\alpha}$, the lifting of α, such that the endpoints of $\bar{h}\cdot\bar{g}(\bar{\beta}_n)$ coincide with those of $\bar{f}\cdot\bar{h}(\bar{\beta})$. Further, there exists an isotopy (the one which changes hg to fh in S) which lifts to an isotopy moving $\bar{h}\cdot\bar{g}(\bar{\beta}_n)$ to $\bar{f}\cdot\bar{h}(\bar{\beta})$. Since, also, isotopies cannot move boundary points of S off boundaries, we must conclude that $\bar{h}\cdot\bar{g}(\bar{\beta}_n)$ meets B at a point, Q, of $\bar{x}_0 \cap B$.

One may ask: How far is Q from P? Let us examine Figure 8 below, where β_3, β_4, are defined to be adjacent to β_1, β_2, along B, and β_5, β_6, to be adjacent to β_3, β_4, respectively. The endpoint of β_i which is not on B is denoted by y_i, $i = 1, \ldots, 6$.

Both endpoints of $\bar{f}\cdot\bar{h}(\bar{\beta})$ must fall between y_5 and y_6. This is because $fh(\alpha)$ is simple, and hence $\bar{f}\cdot\bar{h}(\bar{\beta})$ cannot meet the copies of itself which are produced by covering space transformations which rotate H about B, moving one unit to the right and one unit to the left. Because both f and h are adjusted as in paragraph 1.6, we must have the endpoints of $\bar{f}\cdot\bar{h}(\bar{\beta})$ lying between y_5 and y_6 on bd H.

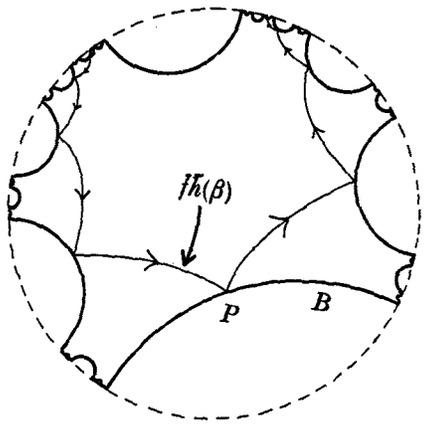


Fig. 7

A similar picture may be drawn for the arc $\bar{h}\cdot\bar{g}(\beta_n)$, so that its endpoints lie between points on bd H which are met by copies of β no more than three displacements along B from Q. But $\bar{h}\cdot\bar{g}(\beta_n)$ and $f\cdot\bar{h}(\beta)$ have the same endpoints. Hence the distance from P to Q, along B, is no greater than 6.

We may alter g by an isotopy which is a rotation about b. The rotation lifts to an isotopy in H which moves Q to P. We then redefine the lifting of the new g to be $\bar{g}$, where $\bar{g}$ leaves B fixed. With this new version of g, we have $\bar{h}\cdot\bar{g}(\beta)$ corresponding with $f\cdot\bar{h}(\beta)$ at P and at the endpoints. It then follows easily that $\bar{h}\cdot\bar{g}(\beta)$ must correspond with $f\cdot\bar{h}(\beta)$ at

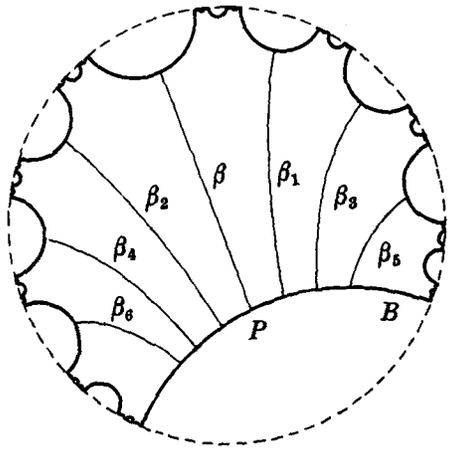


Fig. 8

all its intersections with $\bar{x}_0$. (This is because $\bar{h}\cdot\bar{g}(\bar{\beta})$ and $\bar{f}\cdot\bar{h}(\bar{\beta})$ are now isotopic, by an isotopy leaving the base point fixed.)

In fact still more can be concluded; namely that $\bar{h}\cdot\bar{g}$ corresponds with $\bar{f}\cdot\bar{h}$ on the whole set $\bar{b}$, the lifting of b. For let us choose α^* to be another based loop from $\{\alpha_i\}$. Let β^* be the curve of $\bar{\alpha}^*$ which passes through P. Then $\bar{h}\cdot\bar{g}(\bar{\beta}^*)$ must have the same endpoints as does $\bar{f}\cdot\bar{h}(\bar{\beta}^*)$, for if not we could not have hg being isotopic to fh. Hence, as before, it may be concluded that $\bar{h}\cdot\bar{g}(\bar{\beta}^*)$ corresponds with $\bar{f}\cdot\bar{h}(\bar{\beta}^*)$ on all its intersections with $\bar{x}_0$. Also, of course, $\bar{h}\cdot\bar{g}$ corresponds with $\bar{f}\cdot\bar{h}$ on B, since both maps are the identity there. Therefore, extending this correspondence as far as possible, we conclude that $\bar{h}\cdot\bar{g}$ corresponds with $\bar{f}\cdot\bar{h}$ on the whole set $\bar{x}_0$, and in fact, on the whole set $\bar{b}$.

It may still be the case that $\bar{h}\cdot\bar{g}$ does not correspond with $\bar{f}\cdot\bar{g}$ on the liftings of endpoints of arcs of $\{\alpha_i\}$ which connect x_0 with boundaries of S other than b. But then, using an argument similar to that above, and an isotopy as in paragraph 1.5, one may alter g by a rotation of size no greater than 6 around such a boundary. We do this for all boundaries other than b. These rotations lift to isotopies in H which leave $\bar{x}_0$ fixed. Hence after they have been preformed we have $\bar{h}\cdot\bar{g}$ corresponding with $\bar{f}\cdot\bar{h}$ on all points which are liftings of endpoints of $\{\alpha_i\}$, and further

$$d(g)_\Delta \leqslant [d(g_{\text{orig}})_\Delta]^{2r!} + 12(\chi_S + 1).$$

(Note that each point of $\bar{x}_0$ meets $2\chi_S + 4 - r$ fundamental regions.) Hence, if δ is a fundamental region of Δ, we know that $\bar{h}\cdot\bar{g}(\delta)$ agrees with $\bar{f}\cdot\bar{h}(\delta)$, at least on the "corners".

Without increasing the size of g we may ensure that $\{g(\alpha_i)\}$ has no trivial intersections with $\{\alpha_i\}$, as in 1.5.

1.8 Agreement between $\bar{f}\cdot\bar{h}$ and $\bar{h}\cdot\bar{g}$, everywhere

The task now is to make $\bar{f}\cdot\bar{h}(\delta)$ agree with $\bar{h}\cdot\bar{g}(\delta)$ everywhere, not just at the corners.

Begin by noting that although we have all the homeomorphisms: f, g, h, producing no trivial intersections with $\{\alpha_i\}$, it may still be the case that either $\{hg(\alpha_i)\}$ has trivial intersections with $\{\alpha_i\}$, or $\{fh(\alpha_i)\}$ has trivial intersections with $\{\alpha_i\}$.

Assume, first of all, that $\{fh(\alpha_i)\}$ has trivial intersections with $\{\alpha_i\}$. If we take f^{-1}, that would imply that $\{h(\alpha_i)\}$ has trivial intersections with $\{f^{-1}(\alpha_i)\}$. Now, $\{f^{-1}(\alpha_i)\}$ can have no trivial intersections with $\{\alpha_i\}$, as in 1.5. Hence any trivial intersections between $\{h(\alpha_i)\}$ and $\{f^{-1}(\alpha_i)\}$ may be eliminated by isotopies to h which leave $\bigcup_i \alpha_i$ setwise fixed (Fig. 9).

That is to say that if D is a disc produced by a trivial intersection between $h(\alpha_i)$ and $f^{-1}(\alpha_j)$, then the curves of $\{\alpha_i\}$ meet D in arcs having one end in $h(\alpha_i)$ and the other end in $f^{-1}(\alpha_j)$.

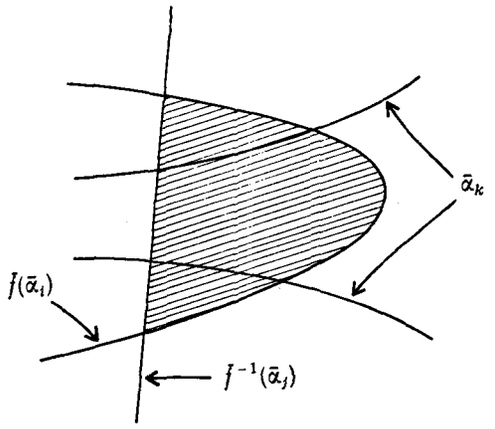


Fig. 9

Therefore we may alter h by an isotopy which does not alter its size in Δ, so that after the isotopy, $\{h(\alpha_i)\}$ has no trivial intersections with $\{f^{-1}(\alpha_i)\}$. But then we have $\{fh(\alpha_i)\}$ having no trivial intersections with $\{\alpha_i\}$. Note further that even after this alteration, we still have $\{h(\alpha_i)\}$ having no trivial intersections with $\{\alpha_i\}$.

Next look at the case that $\{hg(\alpha_i)\}$ has trivial intersections with $\{\alpha_i\}$. This time take h^{-1} and note that we must have $\{g(\alpha_i)\}$ having trivial intersections with $\{h^{-1}(\alpha_i)\}$. But then the same argument as before gives us an alteration to g which does not change the size of g in Δ, and after which $\{hg(\alpha_i)\}$ has no trivial intersections with $\{\alpha_i\}$. Furthermore, $\{g(\alpha_i)\}$ still has no trivial intersections with $\{\alpha_i\}$.

If, now, there are any trivial intersections of $\{hg(\alpha_i)\}$ with $\{fh(\alpha_i)\}$, then they are as illustrated in Fig. 10.

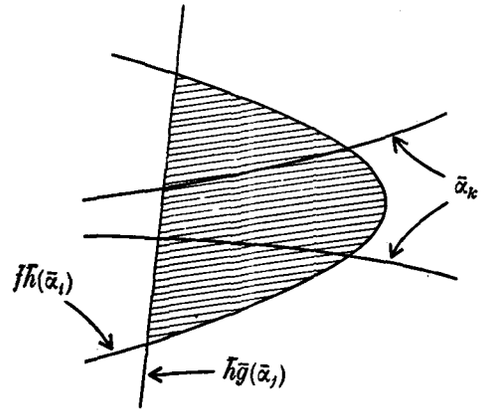


Fig. 10

That is, they may be eliminated by an isotopy applied to f which leaves $\bigcup_i \alpha_i$ setwise fixed (and hence does not alter the size of f in Δ). After thus altering f, we still have $\{fh(\alpha_i)\}$ having no trivial intersections with $\{\alpha_i\}$.

At this stage, since $\bar{f}\cdot\bar{h}$ and $\bar{h}\cdot\bar{g}$ agree on the liftings of all endpoints of arcs of $\{\alpha_i\}$, we must have $\bar{f}\cdot\bar{h}$ and $\bar{h}\cdot\bar{g}$ agreeing on the liftings of all the curves in $\{\alpha_i\}$. That is to say if δ is a fundamental region of Δ then $\bar{f}\cdot\bar{h}(\delta)=\bar{h}\cdot\bar{g}(\delta)$, although the agreement is not necessarily pointwise, except at the corners of δ.

Applying $\bar{h}^{-1}$, we have then $\bar{g}(\delta)=\bar{h}^{-1}\cdot\bar{f}\cdot\bar{h}(\delta)$, with agreement on the corners. But in this case we apply an isotopy to g which leaves $\bigcup_i \alpha_i$ setwise fixed, so that at all stages of the isotopy, the set $\bar{g}(\delta)$ is taken to itself. At the end of the isotopy there is pointwise equality between $\bar{g}$ and $\bar{h}^{-1}\cdot\bar{f}\cdot\bar{h}$ on δ, hence everywhere. Therefore $g=h^{-1}fh$, exactly.

So we have proved:

PROPOSITION. *If f', g', h', produce no trivial intersections in $\{\alpha_i\}$, and the liftings, $\bar{f}'$, $\bar{g}'$, $\bar{h}'$, are such that the maps, $\bar{f}'\bar{h}'$ and $\bar{h}'\bar{g}'$, agree on the corners of fundamental regions, then f', g', h', are isotopic, by isotopies which do not increase their size, to homeomorphisms f, g, h, which are such that $g=h^{-1}fh$.*

In the second and third sections we will often be working with $h(\Delta)$. If δ is a fundamental region of Δ then $\bar{h}(\delta)$ is a fundamental region of another system of curves on S, $h(\Delta)$. In general we expect h to be of large size in Δ. That means that $\bar{h}(\delta)$ crosses many fundamental regions of Δ. Conversely, $\bar{h}(\delta)$ is a fundamental region of $h(\Delta)$, and if δ^* is a disc of Δ, we will expect many fundamental regions of $h(\Delta)$ to meet it. Since $\{h(\alpha_i)\}$ has no trivial intersections with $\{\alpha_i\}$, we find that, in general, if $\bar{h}(\delta)\cap\delta^*$ is not empty then it is either a single disc, or if they happen to share a side, a single arc, or if they happen to share a corner, a single corner.

2. The small curve, c

In this second part our goal will be to produce a closed curve c which is "small" in both the systems Δ and $h(\Delta)$. In order to produce c we will assume that f has no periodic curves. This is where the condition about periodic curves in the main theorem arises. Once the curve c is produced, the theorem is easily proven in the third part.

2.1 The definition of the arcs γ_n

We examine the curves $(\bar{f})^n\bar{h}(\beta)$, for all integral values of n (notation as in 1.6). Since β is simple, all these curves are simple also. If $n=1$, we have proved in section 1.8 that $\bar{f}\bar{h}(\beta)$ has no trivial intersections with $\{\alpha_i\}$. (Hence the image of $\bar{f}\bar{h}(\beta)$ in S, namely $fh(\alpha)$, has

no trivial intersections with $\{\alpha_i\}$.) It is also true that $\tilde{f}\tilde{h}(\beta)$ has no trivial intersections with $\{\tilde{h}(\alpha_i)\}$. This is because $\{g(\alpha_i)\}$ has no trivial intersections with $\{\alpha_i\}$, and $g=h^{-1}fh$.

We wish to have all the arcs $\{(\tilde{f})^n\tilde{h}(\beta)\}$ enjoying these properties also. Hence we wish to find, for each n, an isotopy, $\tilde{k}_n$, lifting from an isotopy, k_n, of S, which is such that the arc, $\tilde{k}_n(\tilde{f})^n h(\beta)$, has no trivial intersections with either Δ or $h(\Delta)$. However such isotopies, $\tilde{k}_n$, may be easily found by simply eliminating the trivial intersections of $f^n h(\alpha)$, first with Δ, and then with $h(\Delta)$. (Remember that $\{h(\alpha_i)\}$ has no trivial intersections with $\{\alpha\}$.) Therefore, in the first step we need introduce no new trivial intersections with $\{h(\alpha_i)\}$, and in the second step we need introduce no new trivial intersections with $\{\alpha_i\}$.

Although no claim is made about the size of k_n in Δ, it is clear that the size of the curve, $k_n f^n h(\alpha)$, is no greater than the size of the curve, $f^n h(\alpha)$, in both Δ and $h(\Delta)$. (The size of a curve is the minimal number of fundamental regions which cover a single "segment" of its lifting to H.) So we have:

$$d(k_n f^n h(\alpha))_{h(\Delta)} \leqslant d(f^n h)_{h(\Delta)} = d(g^n)_\Delta \leqslant d(g)_\Delta^n.$$

The equality here follows from the fact that $g^n = h^{-1}f^n h$.

For ease of notation we will denote $\tilde{k}_n(\tilde{f})^n\tilde{h}(\beta)$ by the symbol γ_n, for all integers n. Then, relating the size of g to the size of g_{orig}, we have:

$$d(\gamma_n)_{h(\Delta)} \leqslant ([d(g_{\text{orig}})_\Delta]^{2r!} + 12(\chi_S + 1))^n.$$

In general, since h is large in Δ, the size of γ_n in Δ will be large. However, the important thing at this stage is that γ_n is limited in size in $h(\Delta)$, i.e. the limit on its size has nothing to do with h. Finally we may note that the set of endpoints which are not P, of the arcs, $\{\gamma_n\}$, occur in order along bd H.

2.2 The basic idea for constructing c

We have a certain intersection of $h(\Delta)$ with Δ. This intersection has been "straightened". Look at a typical fundamental region of Δ, call it δ, and look at the way the different fundamental regions of $h(\Delta)$ intersect δ (Fig. 11).

A certain number of the intersections are rectangles, not containing corners of δ. Such a rectangle is shaded in the figure. We call such a rectangle a "parallelity region". Two opposite sides of a parallelity region are in two different sides of δ, and the other two opposite sides of the parallelity region are in different sides of a fundamental region of $h(\Delta)$. The intersections of $h(\Delta)$ with Δ, which are not parallelity regions, are called "non-parallelity regions". Since h has been straightened, there is a limit to the number of non-parallelity regions in δ which are possible. That limit is $2n+2$, where n is the number of sides of a fundamental region. Substituting $n=2(1+\chi_S)+r$, we obtain $4\chi_S+2(r+3)$ as the greatest number of non-parallelity regions in δ which can be expected to occur.

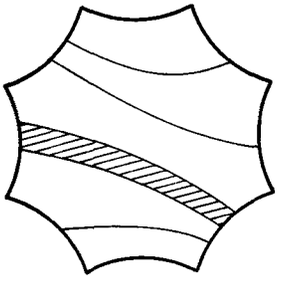


Fig. 11

Examine the set of arcs $\{\gamma_i\}$. We are interested in the relationship of γ_i to γ_{i+1}. In general, as one proceeds along γ_i, from P, one passes through a number of fundamental regions of Δ. At first, perhaps, γ_i and γ_{i+1} go through a certain set of fundamental regions of Δ together, before branching apart (Fig. 12).

We will be interested in the last fundamental region of Δ where they remain together. Of course, γ_i and γ_{i+1} may, perhaps, not go through any fundamental regions together. That is a difficulty with which we shall have to contend. On the other hand they may stay together so that γ_{i+1} meets the fundamental region in which γ_i has its endpoint away from P (or perhaps γ_i meets the fundamental region in which γ_{i+1} has its endpoint away from P). In this case we are also interested in this "last" fundamental region of Δ.

In section 2.5 we shall prove that γ_i is "near" to a non-parallelity region of this "last" fundamental region of Δ. That is to say that there exists an arc, drawn from γ_i to a non-parallelity region within this "last" fundamental region, which intersects only a "small" number of parallelity regions.

We do this for $\gamma_0, \gamma_1, \ldots, \gamma_N$, where $N=4\chi_S+2(3+r)$. Then, for some i, j, with

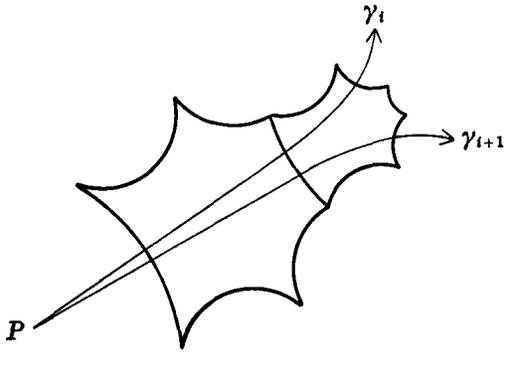


Fig. 12

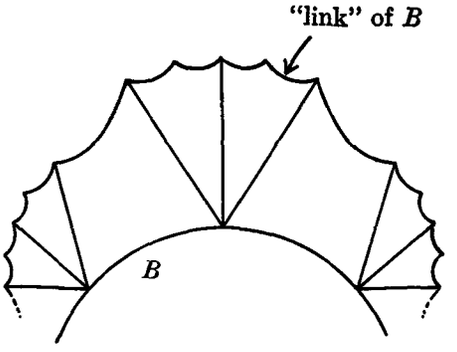


Fig. 13

$0 \leqslant i < j \leqslant N$, we will have the non-parallelity region linked to γ_i, corresponding with the non-parallelity region linked to γ_j, under a covering space transformation.

We may hope that the closed curve, c, which is the image on S of the arc in H which is generated by proceeding from the non-parallelity region near γ_i, across to γ_i, down γ_i to P, then up γ_j and across to the non-parallelity region near γ_j, will be of small size, both in Δ and in $h(\Delta)$. We must, however, be careful. Perhaps the two non-parallelity regions which generate c are the same on H, so that c is contractible on S. Perhaps c is parallel to a boundary. Perhaps, finally, γ_i and γ_j stay together for a long distance so that c will be large in Δ after all.

With an eye to these difficulties we return to the set of arcs, $\{\gamma_i\}$, and examine a certain special case.

2.3 The "link" of B and the definition of a particular case

We have our boundary, B, in the hyperbolic plane, H. Let us examine the set of sides of the fundamental regions of Δ, which might be termed the "link" of B in Δ. These sides are characterized by the property that they do not meet B, and yet they are sides of fundamental regions of Δ which do meet B (Fig. 13).

It is clear that by looking at covering space transformations which take B onto itself, we have transformations of the link of B onto itself. Hence we obtain equivalence classes of sides in the link of B. Two sides are equivalent if there is a covering space transformation taking one side to the other, which takes B to itself. The number of equivalence classes (if we do not count sides which are liftings of boundaries of S) is $4\chi_S^2+6\chi_S-2r\chi_S$.

We look again at the set of arcs, $\{\gamma_i\}$. In particular we look at the arcs, $\{\gamma_i\}$, where $0 \leqslant i \leqslant N_1 - 1$. Here, N_1 is defined to be the following number:

$$N_1 = [4\chi_S^2 + 6\chi_S - 2r\chi_S]\{2[4\chi_S + 2(3+r)][2(1+\chi_S)+r][2(1+\chi_S)+2]+3\}.$$

Note that N_1 depends only upon the surface S, and not upon f, g, or h.

In this set, $\{\gamma_i\}$, $0 \leqslant i \leqslant N_1$, there must be a subset, $\{\xi_i^*\}$, $1 \leqslant i \leqslant 2[4\chi_S + 2(3+r)][2(1+\chi_S)+r][2(1+\chi_S)+2]+3$, such that all the ξ_i^* meet sides of the link of B which are equivalent. (We will call this number N_2.) The indexing of the set $\{\xi_i^*\}$, $1 \leqslant i \leqslant N_2$, is, for the moment, not specified as having any particular properties.

All the ξ_i^* pass through a single equivalence class of sides of the link of B. Further, each ξ_i^* passes through a side of the link of B which is no further than $d(f)_\Delta^{N_1}$ from P. (This is true because γ_0 passes through a side of the link of B which is a distance 1 from P. Then γ_1 must pass through a side of the link of B a distance no greater than $d(f)_\Delta$ from P, and so forth.)

Let σ be the side of the link of B which is in this equivalence class and which is as near to P as possible. Hence, σ must be a side of a fundamental region of Δ containing P.

Each ξ_i^* is equivalent, under a covering space transformation taking B to itself, to an arc, ξ_i, which passes through σ. Let us now specify the indexing of $\{\xi_i\}$, $1 \leqslant i \leqslant N_2$, so that the endpoints of arcs of $\{\xi_i\}$ which are not on B, occur in order on bd H. That is to say that if y_i is the endpoint of ξ_i which is not on B, then y_0 is furthest to the left, and for each i, y_i is to the left of y_{i+1}. This set of points, $\{y_i\}$, is contained in the set, $\bar{x}_0$, of liftings of the base point, x_0. Note further that the endpoints of $\{\xi_i\}$, on B, are all within $d(f)_\Delta^{N_1}$ from P. In fact they are all to one side of P, so they are all within $d(f)_\Delta^{N_1}$ of one another. (The distance between two points in H is defined as the least number of fundamental regions which can cover an arc connecting the two points.)

We may imagine the situation that y_i and y_j coincide, for some $i \neq j$. But it is not difficult to see that in this case there will be a periodic curve on S, namely the image of ξ_i in S. Hence it may be assumed that no y_i, y_j coincide, $i \neq j$. In fact, more generally we may assume, for the same reason, that no two different y_i, y_j are in the same component of $\bar{b}$, the lifting of b.

Therefore, between each pair of points, y_i, y_j, there are a great many points of $\bar{x}_0$ on bd H. Let us denote by z_i a point of $\bar{x}_0$ between y_i and y_{i+1} (including y_i, excluding y_{i+1}) which is as near as possible to P. (We shall be more specific about our choice of z_i in the next paragraph.)

The distance between z_i and P, in Δ, is denoted by $d(z_i, P)_\Delta$. The special case with which we shall first deal may be defined as follows:

$$d(z_i, P)_\Delta \leqslant d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1}, \quad \text{for all the } z_i,\ 1 \leqslant i \leqslant N_2 - 1.$$

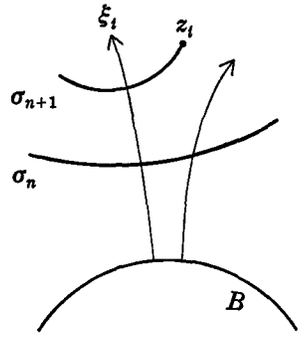


Fig. 14

2.4 The Points z_i

At this stage we need to make precise the definition of z_i. In the particular case which we are considering we know that all the ξ_i pass through σ. Let us fix our attention on a specific ξ_i and ξ_{i+1}. We know that in traversing the length of ξ_i, from its end at B to its other end at y_i, we pass through σ (or perhaps we end at σ, as the case may be). Similarly for ξ_{i+1}. If ξ_i does not end at σ, then denote by σ_1 the next side of a fundamental region of Δ which ξ_i meets after passing through σ. Similarly let σ_2, σ_3, and so forth, be the sides of fundamental regions of Δ which ξ_i meets after passing through σ_1. We know that ξ_{i+1} meets σ. It may be the case that ξ_{i+1} also meets σ_1. Perhaps it also meets σ_2. Let σ_n be the last such side which both ξ_i and ξ_{i+1} meet together before branching apart from one another.

It may be that ξ_i ends at σ_n, i.e. y_i is an endpoint of σ_n. In this case we define $z_i = y_i$. If ξ_i does not end at σ_n then let σ_{n+1} be the next side of a fundamental region of Δ through which ξ_i passes. Define z_i to be a point of $\tilde{x}_0$ which is between y_i and y_{i+1} on bd H and as close to σ_n as possible (Fig. 14).

Note that this definition of z_i is in no way in conflict with the requirement for z_i given in the previous paragraph.

2.5 Non-parallelity regions in the fundamental regions

To proceed further we need the following:

LEMMA. *Let σ_n, σ_{n+1}, be the sides of fundamental regions of Δ defined above. (We assume σ_{n+1} exists. If it does not, then the lemma is trivially true.) Let δ be the fundamental region of Δ containing σ_n and σ_{n+1}. Then there exists a non-parallelity region, Σ, of S (with respect to $h(\Delta)$, as defined in 2.2) such that ξ_i is "near" to Σ in the following sense: There exists an arc*

10 – 782904 *Acta mathematica* 142. Imprimé re 20 Févri 1979

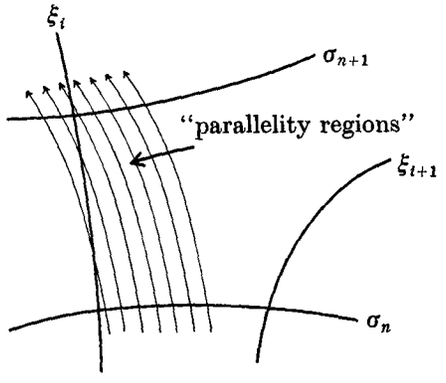


Fig. 15

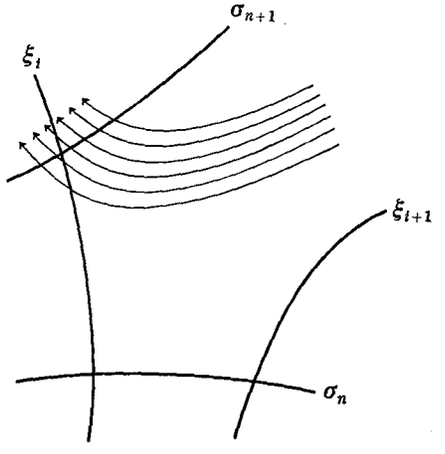


Fig. 16

in δ which connects ξ_i with Σ, and the arc passes through no more than $2d(g)_\Delta^{N_1}+d(f)_\Delta^{N_1}$ parallelity regions in δ.

Proof. In order to prove this let us begin by looking at the point $\xi_i \cap \sigma_{n+1}$ (Fig. 15). If $\xi_i \cap \sigma_{n+1}$ is in a non-parallelity region of δ, then we are finished. So let us assume it is in a parallelity region, and we travel along σ_{n+1}, from $\xi_i \cap \sigma_{n+1}$, in the direction of ξ_{i+1}. We see how many parallelity regions of δ are met before meeting the first non-parallelity region.

If we meet more than $2d(g)_\Delta^{N_1}+d(f)_\Delta^{N_1}$ parallelity regions before meeting a non-parallelity region, then let us see what the parallelity regions can do in δ. They all meet σ_{n+1} together, so they must all meet another side of δ together. If that side is σ_n then we would have too many fundamental regions of $h(\Delta)$ meeting ξ_i and ξ_{i+1}. Remember (paragraph 2.1) that both ξ_i and ξ_{i+1} can have size no greater than $d(g)_\Delta^{N_1}$ in $h(\Delta)$. Hence both ξ_i and ξ_{i+1} pass through no more than $d(g)_\Delta^{N_1}$ fundamental regions of $h(\Delta)$. But each parallelity region of δ is part of a fundamental region of $h(\Delta)$. Hence if all the parallelity regions under consideration pass through σ_n, then each fundamental region of $h(\Delta)$ which contains one of these parallelity regions of δ will meet either ξ_i, ξ_{i+1}, or B between ξ_i and ξ_{i+1}. The first two possibilities account for at most $d(g)_\Delta^{N_1}$ intersections each, and the third for $d(f)_\Delta^{N_1}$ intersections. Therefore if we have more than $2d(g)_\Delta^{N_1}+d(f)_\Delta^{N_1}$ parallelity regions we may assume that they do not go through σ_n. In that case they must all meet another side of δ, as illustrated (Fig. 16).

But then if we start at $\xi_i \cap \sigma_{n+1}$ and proceed along ξ_i, through δ, in the direction of σ_n, we must encounter a non-parallelity region which actually meets ξ_i.

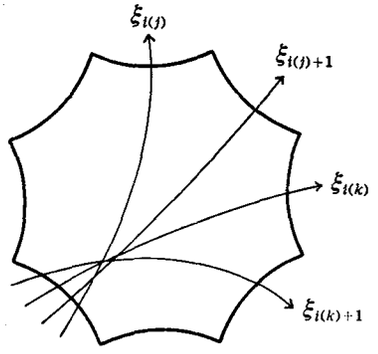


Fig. 17

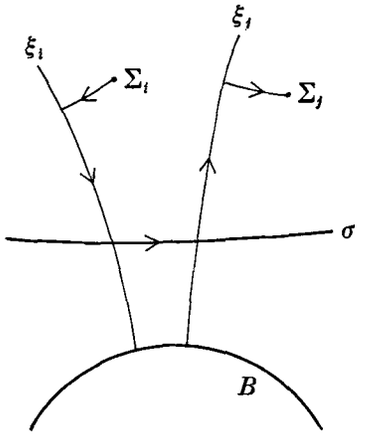


Fig. 18

In either case there is a non-parallelity region of δ within $2d(g)_{\Delta}^{N_1}+d(f)_{\Delta}^{N_1}$ parallelity regions from ξ_i.

2.6 Constructing the curve c in our special case

We have our sequence of arcs, $\{\xi_i\}$, $0\leqslant i\leqslant N_2$. Let us denote by N_3 the number:

$$N_3=[4\chi_S+2(3+r)][2(1+\chi_S)+r][2(1+\chi_S)+2]+1.$$

Then the sequence, $\{\xi_i\}$, $1\geqslant i\geqslant N_3$, is the "first half" of the original sequence. We shall work, for the time being, with this first half of the sequence.

In light of the preceeding paragraph we know that each ξ_i, $1\leqslant i\leqslant N_3-1$, is "close" to a non-parallelity region of the fundamental region in which ξ_i and ξ_{i+1} branch apart. There are no more than $4\chi_S+2(3-r)$ non-parallelity regions. Also there are $2(1+\chi_S)+r$ "corners" of each fundamental region. Therefore there must be at least $2(1+\chi_S)+1$ indices, [call them $i(j)$, $1\leqslant j\leqslant 2(1+\chi_S)+1$] such that the non-parallelity regions near to each $\xi_{i(j)}$ correspond under covering space transformations which move the corresponding points, $z_{i(j)}$, onto one another.

It may be that some of these non-parallelity regions are in the same fundamental region of Δ. Let us say that they are in the fundamental region, δ, of Δ. Then we would have a $\xi_{i(j)}$, and a $\xi_{i(k)}$, $i(j)\neq i(k)$, with $\xi_{i(j)}$ branching apart from $\xi_{i(k)+1}$ in δ (Fig. 17).

But this can happen in δ for no more than $2(1+\chi_S)$ arcs, since that is the number of sides of δ. Hence there must exist ξ_i, ξ_j, $i\neq j$, with $1\leqslant i<j\leqslant N_3-1$, with ξ_i close to the non-parallelity region Σ_i, say, and ξ_j close to the non-parallelity region Σ_j, say. Σ_i corresponds to Σ_j under a covering space transformation taking z_i to z_j, and $\Sigma_i\neq\Sigma_j$.

We may now construct the arc, C_1, as illustrated in Fig. 18. Begin in Σ_i, go over

to ξ_i, passing through as few fundamental regions as possible (as in paragraph 2.5). Then travel down ξ_i to σ, across σ to ξ_j, then up ξ_j to the fundamental region containing Σ_j, and finally across to Σ_j, passing through as few fundamental regions of $h(\Delta)$ as possible. We may assume that the starting and finishing points correspond, in that the images in S are the same point. The image of C_1 in S, namely c_1, is a closed curve. It is small in the following sense:

$$d(c_1)_\Delta \leqslant 2(d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1})$$
$$d(c_1)_{h(\Delta)} \leqslant 3(2d(g)_\Delta^{N_1} + d(f)_\Delta^{N_1}).$$

These numbers are calculated by adding up the sizes of each of the "segments" of C_1 in Δ and in $h(\Delta)$. (The limits on the sizes of c_1 in Δ and in $h(\Delta)$ may be expressed in terms of f_{orig} and g_{orig} by means of the inequalities in part 1.)

Hence we have succeeded in finding a "small" closed curve, c_1, in this special case which we are considering. Since $\Sigma_i \neq \Sigma_j$, we know that c_1 cannot be contractible on S. It may, however, be deformable to a boundary of S. If this is the case then take the set of arcs, $\{\xi_i\}$, $N_3+1 \leqslant i \leqslant N_2$, and call this the "second half" of the sequence. Then we find a curve, c_2, constructed from the second half of the sequence, as was done for the first half. The curve c_2 is generated by a curve, C_2, in H, between two non-parallelity regions, as was the case with C_1.

If c_2 is not deformable to a boundary then we are finished, since the size of c_2 is subject to the same limitations as was the size of c_1. Hence we assume that c_2 is also parallel to a boundary. In this case we generate the curve, c_3, on S, by going once around c_1 on S, starting and finishing on the image of σ, then going along the image of σ to c_2, going once around it, then back to the start. Clearly the size of c_3 is limited by twice the limits for c_1 and c_2, i.e.:

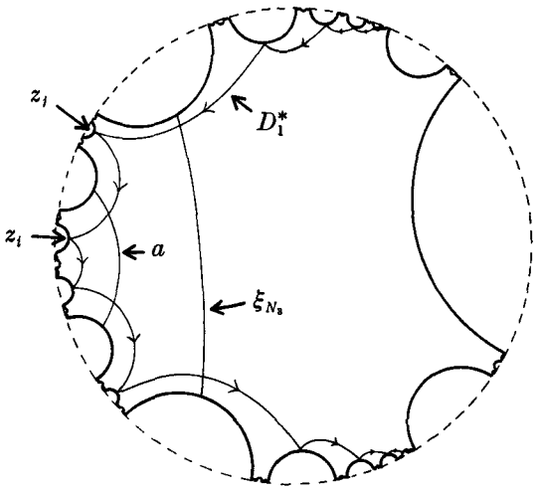
$$d(c_3)_\Delta \leqslant 4(d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1})$$
$$d(c_3)_{h(\Delta)} \leqslant 6(2d(g)_\Delta^{N_1} + d(f)_\Delta^{N_1})$$

We will prove in the next paragraph that c_3 cannot be deformable to a boundary on S.

2.7 Proving that c is not deformable to a boundary in our special case

LEMMA. *The curve, c_3, cannot be deformable to a boundary of S.*

Proof. Let us begin by looking at the curve, c_1, on S. In general, c_1 does not pass through x_0. We deform c_1, using a homotopy, so that c_1 is deformed to the curve, d_1,



Fig, 19

which does pass through x_0. We may assume that this homotopy lifts to a homotopy in H which deforms C_1 to a curve, D_1, which passes through z_i. (Remember that the curve, C_1, has been constructed from the arcs ξ_i and ξ_j.)

The curve, d_1, is obtained from c_1 by a homotopy on S, hence, in H, the endpoints of the curve containing C_1 do not move on bd H when C_1 is changed to D_1. Denote by D_1^* the curve of $\tilde{d}_1$ which contains D_1.

We prove that at least one of the endpoints of D_1^* is on the same side of ξ_{N_s} as is ξ_i, in H (Fig. 19).

Note, to begin with, that z_i and z_j are on the same side of ξ_{N_s}. Can we have *both* ends of D_1^* on the other side of ξ_{N_s}, away from z_i and z_j, as illustrated in Figure 19? If we had this situation then we may assume that a "segment" of D_1^*, which is not the one between z_i and z_j, is crossed by ξ_{N_s}. (A "segment" is one covering of d_1, starting and finishing in $\tilde{x}_0$.) Let the copy, a, of ξ_{N_s}, be the copy corresponding to the segment of D_1^* between z_i and z_j under a covering space transformation given by some power of the element of the fundamental group of S represented by d_1. Since the image of ξ_{N_s}, on S, is simple, we cannot have a meeting ξ_{N_s}. But this is impossible because there are fewer segments of D_1^* to one side of a than there are to one side of ξ_{N_s}, which violates the invariance of H under covering space transformations. Therefore we must conclude that one end of D_1^1, hence one end of C_1^* (the arc of $\tilde{c}_1$ containing C_1), is on the same side of ξ_{N_s} as are z_i and z_j.

By similar reasoning we may prove that one end of C_2^* is on the opposite side of ξ_{N_s+1} to z_i and z_j.

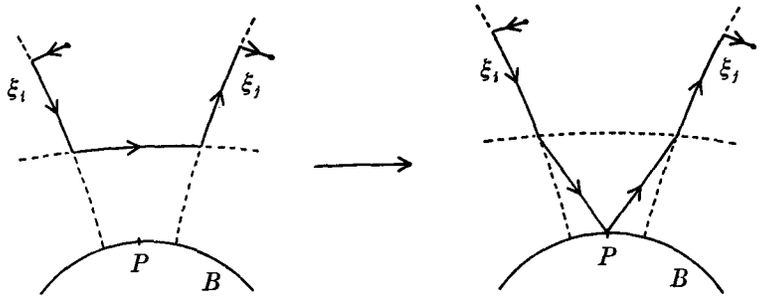


Fig. 20

We next produce the curve, G_1, from C_1, as illustrated in Fig. 20. G_1 is produced from C_1 by proceeding down ξ_i to σ, then to P, then from P up to σ, where it joins with ξ_j and then continues as before. Do the same to C_2 to get G_2. The images on S are g_1 and g_2. Then g_3 is obtained by going once around g_1 on S, starting and finishing at x_0, then once around g_2, starting and finishing at x_0. Clearly, g_3 is homotopic to c_3. Since S is not a disc with two holes, the curve, g_3, can only be homotopic to a boundary of S if g_1 and g_2 are homotopic to one another, at least up to multiples, by a homotopy which leaves x_0 fixed. For this to be true we would have to have G_1^* and G_2^* having the same endpoints on bd H. But since they have endpoints on opposite sides of ξ_{N_3} and ξ_{N_3+1}, and g_1, g_2 are homotopic to boundaries of S, we must have both G_1^* and G_2^* being homotopic to B, in H. But this is impossible. For example if G_1^* were homotopic to B then z_i could be moved to z_j by a covering space transformation taking B to itself. Since z_i and z_j lie between the endpoints of σ on bd H, this is impossible. Hence we conclude that c_3 is not homotopic to a boundary of S.

Let the curve c, on S, be defined to be either c_1, c_2, or c_3, whichever is not homotopic to a boundary.

2.8 The second special case: "many" γ_j^* pass through σ

In the special case which was defined in paragraph 2.3, we are finished. A small curve, c, has been produced. What of the other cases which are possible? That is to say, what if $d(z_i, P)_\Delta > d(f)_\Delta^{N_1 N_2} + d(f)_\Delta^{N_1}$, for some $1 \leqslant i \leqslant N_2 - 1$?

In this case we look at ξ_i and ξ_{i+1}. Remember that $\xi_i^* = \gamma_n$, $\xi_{i+1}^* = \gamma_m$, for some, $0 \leqslant m$, $n \leqslant N_1$. Also, γ_j was defined to be $k_j f^j \bar{h}(\beta)$, for each j. Let us, therefore, define f^* to be, $f^* = f^{m-n}$. Then we shall have, up to an isotopy leaving $\tilde{x}_0$ fixed, $f^*(\gamma_n) = \gamma_m$.

Define also, $g^* = g^{m-n}$. From now on we will work with f^* and g^*. We will need to alter them as f and g were altered in Part 1. Note that the sizes of f^* and g^* in Δ and $h(\Delta)$ will

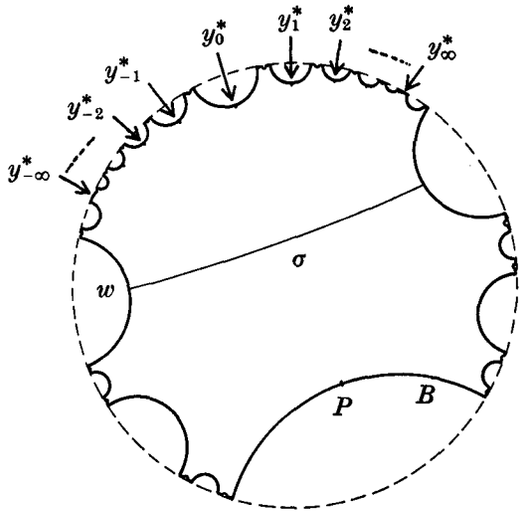


Fig. 21

not be increased. Also we will need to find a suitable conjugating homeomorphism, h^*, for f^* and g^*.

We return to ξ_i and ξ_{i+1}. In general, $\xi_i \cap B$, and $\xi_{i+1} \cap B$, are different points. As shown in paragraph 2.3, they can be no further than $d(f)_\Delta^{N'}$ apart. Let us alter f^* by an isotopy rotating the boundary, b, of S, about itself so that the lifting of the isotopy moves the point $\xi_{i+1} \cap B$ to the point $\xi_i \cap B$. Then take the lifting, $\tilde{f}^*$, of f^*, which leaves B fixed. This insures that f^* is "straight" (as in paragraph 1.6). We may also alter f^* by the isotopies of paragraph 1.5.

Let the sequence, $\{\gamma_j^*\}$, of arcs in H, be defined by: $\gamma_j^* = (\tilde{f}^*)^j(\xi_i)$, for all integers, j. Let y_j^* be the endpoint of γ_j^* which is not at P. Then certainly y_0^* is between the endpoints of σ on bd H (Fig. 21).

To fix ideas let y_1^* be to the right, say, of y_0^*. Then for all j, y_{j+1}^* will be to the right of y_j^*. Clearly there will be an accumulation point to the left and also an accumulation point to the right. The question is: do the accumulation points fall between the endpoints of σ on bd H, or not?

Hence we have two cases to consider: (i) the accumulation points do not both fall between the endpoints of σ on bd H, and (ii) they do. Let us tackle case (i) first. Assume that the left hand side accumulation point (i.e. $\lim_{j\to-\infty} y_j^*$) is not between the endpoints of σ. The argument for the right hand side accumulation point is similar.

Denote by w the left hand endpoint of σ. There must exist an index, u, such that y_u^* is between the endpoints of σ, and y_{u-1}^* is not between them.

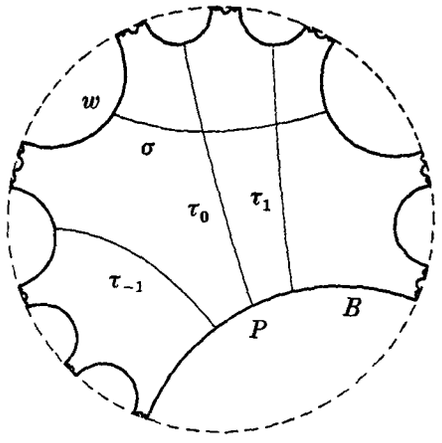


Fig. 22

We define the homeomorphism, h^*, to be: $h^* = f^{(u(m-n)+n)} \cdot h$. Let $\tilde{h}^*$ be the lifting of h^* which leaves B fixed. Then $\tilde{h}^*(\beta) = \gamma_{(u(m-n)+n)}$ (at least up to an isotopy leaving the base point fixed). Apply the straightening isotopies of paragraphs 1.5 and 1.6 to h^*. The isotopies will be of size no greater than $d(f)_\Delta^{N_1}$ because the distance from $\xi_i \cap B$ to P is no greater than $d(f)_\Delta^{N_1}$.

Let us define the sequence of arcs, $\{\tau_j\}$, in H, by: $\tau_j = (\tilde{f}^*)^j \tilde{h}^*(\beta)$. We also define the set of points, $\{z_j^*\}$, for the arcs, $\{\tau_j\}$, as was done in paragraph 2.4 for the arcs $\{\xi_j\}$. The definition given there did not include the possibility that a z_j^* is at distance 1 from P. In this case define z_j^* to be the point of $\tilde{x}_0$ which is at distance 1 from P and is the closest possible to the endpoint of τ_j which is not P.

Because of the way we have chosen f^* and $\{\tau_j\}$, we must have $d(z_{-u}^*, P)_\Delta > d(f)_\Delta^{N_1 N_2}$.

Concerning the position of w (the left hand endpoint of σ) with respect to the sequence $\{\tau_j\}$, we cannot quite assert that w falls between τ_{-1} and τ_0. The reason is that the endpoint of $\gamma_{u(m-n)+n}$ which is not on B may be on the component of $\tilde{b}$ which contains w. Certainly we can assert that w falls between τ_{-2} and τ_1. For the sake of definiteness, and without affecting the argument, we may assume that w is between τ_{-1} and τ_0 (Fig. 22).

Since w is between τ_{-1} and τ_0, then $\tilde{f}^*(w)$ must be between τ_0 and τ_1. Further, $d(\tilde{f}^*(w), P)_\Delta \leqslant d(f^*)_\Delta \leqslant d(f)_\Delta^{N_1}$. In fact, in general, $(\tilde{f}^*)^j(w)$ lies between τ_{j-1} and τ_j on bd H, and $d((\tilde{f}^*)^j(w), P)_\Delta \leqslant d((f^*)^j)_\Delta \leqslant d(f)_\Delta^{jN_1}$. If we take, then, the set, $\{\tau_j\}$, $0 \leqslant j \leqslant N_2 - 1$, we must have all the τ_j in this set passing through σ. The reason is that it takes a long time for w to get far away from P under the repeated action of f^*. In detail, all points of $\tilde{x}_0$ between ξ_i and ξ_{i+1} are more than the distance $d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1}$ from P. On the other hand, the isotopies to f^* and h^* are each of size no greater than $d(f)_\Delta^{N_1}$; hence we certainly

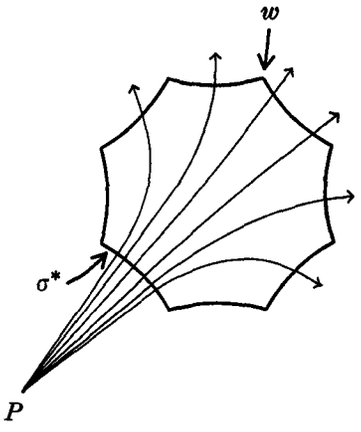


Fig. 23

have a τ_k and a τ_{k+1}, both of which pass through σ, such that all points of $\tilde{x}_0$ between τ_k and τ_{k+1} are further than $d(f)_\Delta^{N_1 N_2}$ from P. Hence all arcs of $\{\tau_j\}$, $0 \leqslant j \leqslant N_2$, must pass through σ.

But suddenly we find ourselves in the special case of paragraph 2.3. Hence there exists a small curve, c, which is not contractible, and not contractible into a boundary, such that:

$$d(c)_\Delta \leqslant 4(d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1})$$

$$d(c)_{h(\Delta)} \leqslant 6(2d(g)_\Delta^{(N_1)^2} + d(f)_\Delta^{(N_1)^2})$$

(Note that the expression for $d(c)_{h(\Delta)}$ contains the exponential $(N_1)^2$, rather than simply N_1. This is because we are working with f^*, rather than f, and f^* is, essentially, the product of f with itself, up to N_1 times. These expressions for the size of c can be related to the sizes of f_{orig} by using the inequalities in part 1.)

2.9 The third, and last, special case: all γ_j^* pass through σ

The only remaining problem is to deal with case (ii) of the preceeding paragraph, i.e. all the y_j^* fall between the endpoints of σ on bd H.

This time, let w be a point of $\tilde{x}_0$ with the property that it is as near as possible to P, i.e. $w \in \tilde{x}_0$, and $d(w, P)_\Delta \leqslant d(w', P)_\Delta$, for all points, w', of $\tilde{x}_0$, between the left and right accumulation points of $\{y_j^*\}$, on bd H. Then w falls between γ_v^* and γ_{v-1}^*, say, for some integer, v. We define the homeomorphism, h^*, to be: $h^* = f^{(v(m-n)+n} \cdot h$, in this case. Next, straighten h^*, and find g^*, as in paragraph 2.8.

We must produce a small curve, c. To begin with, we know that every τ_j goes through σ. As in 2.4, we let σ^* be the last side of a fundamental region of Δ such that *all* the τ_j meet σ^* (Fig. 23).

We may assume that w is a corner of a fundamental region of Δ which contains σ^*. We construct the curve, c, as was done in 2.6, but this time using σ^* in place of σ. We shall have the same limitations on the size of c, as were found in 2.8, if we can establish that $(\tilde{f}^*)^j(w)$ is "near" to σ^* for small j, (i.e. the distance from $(\tilde{f}^*)^j(w)$ to σ^* is no greater than $d(f)_\Delta^{jN_1}$). However this is easily proven to be the case, as follows:

Denote by δ the fundamental region of Δ containing both σ^* and w. Then $\tilde{f}^*(\delta)$ is covered by $d(f^*)_\Delta \leqslant d(f)_\Delta^{N_1}$ fundamental regions of Δ. If w is between τ_{-1} and τ_0, then $\tilde{f}^*(w)$ is between τ_0 and τ_1, in H. In fact, $\tilde{f}^*(w)$ is between τ_0 and τ_1, and is on the far side of σ^* from P. (Remember that σ^* divides H into two pieces.)

Let us say that $\tilde{f}^*(w)$ is further than $d(f^*)_\Delta$ from σ^*. In this case we would have the entire fundamental region, $\tilde{f}^*(\delta)$, being on the far side of σ^* from P, and in fact $\tilde{f}^*(\delta)$ would not meet δ, except possibly at a corner or a side.

Where does the point, w, lie with respect to $\tilde{f}^*(\sigma^*)$? It must be either an endpoint of $\tilde{f}^*(\sigma^*)$, or else it lies on the same side of $\tilde{f}^*(\sigma^*)$, in H, as P. (Remember that $\tilde{f}^*(\sigma^*)$ divides H into two pieces.) However, this is impossible since it would imply that $(\tilde{f}^*)^{-1}(w)$ would lie either on σ^*, or on the same side of σ^*, in H, as P. This contradicts the definition of w. We must conclude that $d(\tilde{f}^*(w), \sigma^*)_\Delta \leqslant d(f^*)_\Delta$. In general,

$$d((\tilde{f}^*)^j(w), \sigma^*)_\Delta \leqslant d((f^*)^j)_\Delta \leqslant d(f^*)_\Delta^j \leqslant d(f)_\Delta^{jN_1}.$$

Hence it may be assumed that the curve, c, which is produced in this case has the same size limitations as were found in 2.8.

We conclude that in all cases there exists a closed curve, c, with c being neither contractible nor contractible to a boundary on S, and,

$$d(c)_\Delta \leqslant 4(d(f)_\Delta^{N_1 N_2} + 2d(f)_\Delta^{N_1}) = N_4,$$
$$d(c)_{h(\Delta)} \leqslant 6(2d(g)_\Delta^{(N_1)^2} + d(f)_\Delta^{(N_1)^2}) = N_5.$$

We call these numbers; N_4, N_5 for convenience.

3. The proof of the theorem and the solution of the conjugacy problem

3.1. We are given our curve, c, with $d(c)_\Delta \leqslant N_4$, and $d(c)_{h(\Delta)} \leqslant N_5$. Take a regular neighbourhood of c in S. In general this will not be simply an annulus, since c may be singular. Hence the regular neighbourhood is a surface on S with a number of boundaries. Some of the boundary curves may bound discs on S. If there are any such curves, then "fill them in" with discs of S. Hence we end up with a certain bounded surface, $T_0 \subset S$. Let us look at $f(T_0)$. It may be the case that $f(T_0)$ is isotopic to T_0, and then again it may not be the case. If it is not, then take a regular neighbourhood of $c \cup f(c)$ on S and "fill in the

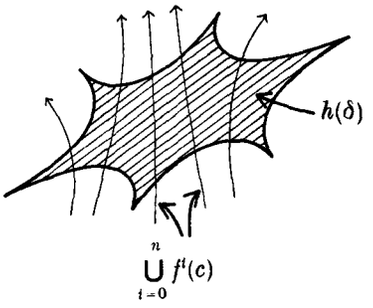


Fig. 24

discs" as before. We get the bounded surface, T_1, on S. Again, $f(T_1)$ may or may not be isotopic to T_1. If not, then form T_2, and so forth.

We continue this process, forming T_i as a regular neighbourhood of $\bigcup_{j=0}^{i} f^j(c)$ on S with the discs filled in. There must exist an n, with $n \leqslant \chi_S + 1$, such that $f(T_n)$ is isotopic to T_n. We consider two cases:

(i) There exists a boundary curve of T_n which is not parallel to a boundary of S, or

(ii) There exists no such curve.

Since c is not parallel to a boundary of S, we may think of these two cases as: (i) T_n is not "all" of S, and (ii) T_n is "all" of S.

If case (i) holds then the boundary curve of T_n which is not parallel to a boundary of S is clearly a periodic curve under f. Hence we may assume that case (ii) holds. In this case we can estimate the size of h in terms of the sizes of f and g in Δ. The idea is as follows:

To begin with, it is clear that under our assumptions, $S - \bigcup_{i=0}^{n} f^i(c)$ is a collection of discs (apart from the components containing the boundaries of S). The question is, how "big" are these discs in Δ, and in $h(\Delta)$? Since $f^i(c)$ is limited in size in Δ, for each i, we must be able to establish an upper limit for the size, in Δ, of a disc of $S - \bigcup_{i=0}^{n} f^i(c)$. The same is true in $h(\Delta)$, so that any disc of $S - \bigcup_{i=0}^{n} f^i(c)$ can cross only a limited number of fundamental regions of $h(\Delta)$.

Let us then take a fundamental region, δ, of Δ, and let it be acted upon by h. We are interested in knowing how "big" the disc $h(\delta)$ can become in Δ. An upper limit for the size of $h(\delta)$ in Δ can be calculated in the following way: Since the size of $f^i(c)$ is limited in $h(\Delta)$, for each i, we must have only a limited number of crossings of the disc $h(\delta)$ by each curve $f^i(c)$. Hence these crossings split the disc $h(\delta)$ into only a limited number of smaller discs (Fig. 24).

However each of these smaller discs is part of a disc of $S - \bigcup_{i=0}^{n} f^i(c)$, which is limited in

size in Δ. Therefore the size of $h(\delta)$ in Δ can be no more than the number of discs of $h(\delta) - \bigcup_{i=0}^{n} f^i(c)$, multiplied by the maximum possible size of a disc of $S - \bigcup_{i=0}^{n} f^i(c)$ in Δ.

To be more specific, since the length of $f^i(c)$ in Δ is no greater than $d(c)_\Delta d(f)_\Delta^i$, for each i, we must certainly have no disc of $S - \bigcup_{i=0}^{n} f^i(c)$ having size greater than $(\chi_S+1)d(c)_\Delta d(f)_\Delta^{\chi_S+1}$ in Δ. (Here the size of a disc is defined in the obvious way.)

But for each i, the size of $f^i(c)$ in $h(\Delta)$ is no greater than $d(c)_{h(\Delta)} d(g)_\Delta^i$. Here, again, the size of a disc of $S - \bigcup_{i=0}^{n} f^i(c)$ in $h(\Delta)$ is no more than $(\chi_S+1)d(c)_{h(\Delta)} d(g)_\Delta^{\chi_S+1}$.

Hence, if we take a fundamental region, δ, of Δ and let it be acted upon by h, then $h(\delta)$ meets, in its interior, no more than $(\chi_S+1)d(c)_{h(\Delta)} d(g)_\Delta^{\chi_S+1}$ arcs of $\bigcup_{i=0}^{n} f^i(c)$. Hence the size of $h(\delta)$ in Δ can certainly be no more than the maximum size of a disc of $S - \bigcup_{i=0}^{n} f^i(c)$, multiplied by the number of times $h(\delta)$ meets $\bigcup_{i=0}^{n} f^i(c)$. Therefore we obtain the limit:

$$d(h)_\Delta < (\chi_S+1)^2 (N_4)(N_5) d(f)_\Delta^{\chi_S+1} d(g)_\Delta^{\chi_S+1}$$

This limit can be expressed in terms of f_{orig} and g_{orig} by means of the inequalities:

$$d(f)_\Delta \leqslant (2(1+\chi_S))^r [d(f_{\text{orig}})_\Delta]^{2r!},$$

and

$$d(g)_\Delta \leqslant [d(g_{\text{orig}})_\Delta]^{2r!} + 12(\chi_S+1),$$

as found in Part 1.

3.2 The generalization to surfaces without boundary

In this section we will sketch the alterations which are necessary in order to construct a proof in the special case that $\partial S = \emptyset$. Clearly many changes in detail will be necessary since we have made extensive use of the fact that isotopies of surfaces with boundaries cannot permute those boundaries. In reality, though, the proof is simpler when $\partial S = \emptyset$ since most of the difficulties which we have encountered resulted from complications involving rotations about the boundaries, and curves which may have been boundary parallel. All of these problems disappear when $\partial S = \emptyset$.

Again we take H to be the universal covering space of S, the hyperbolic plane. This time, however, x_0 will be an interior point of S, and so the lifting of x_0 to H will be an infinite collection of disjoint points in int H, all equivalent to one another under covering space transformations. With this in mind, sections 1.1 to 1.3 remain unchanged. Section 1.4 is no longer necessary. In section 1.5 we include a further operation into the process of "removing trivial intersections". Namely, any intersection of the form illustrated in Figure 25 is "trivial" and should be removed by an alteration to f.

(Of course the problems involving rotations about the boundaries, in section 1.5, cannot occur.) We then choose a specific point, P, of $\hat{x}_0$ in H, and choose the lifting $\hat{f}$ of f which leaves P fixed. Then section 1.6 is no longer needed.

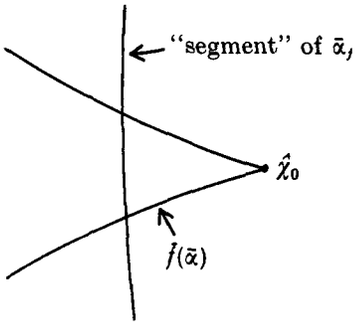


Fig. 25

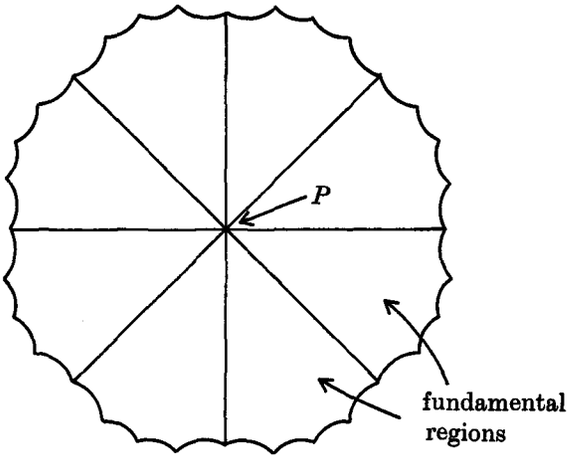


Fig. 26

At this stage all the lines of $\bar{f}(\bar{\alpha}_i)$ meet the lines of $\bar{\alpha}_j$ in at most one point each, for all possible i, j. Hence the proof may be continued in analogy to sections 1.7 and 1.8. (It is assumed that all liftings, $\bar{f}$, $\bar{g}$, $\bar{h}$, leave the same point P of $\hat{x}_0$ fixed.)

Part 2 is technically easier when $\partial S = \emptyset$. We do not need the special case of section 2.3 (this is because the problem of rotations about the boundaries cannot occur). In fact we only need the second and third cases, in sections 2.8 and 2.9. To see this, consider a certain neighbourhood of P in H. Specifically, the set of fundamental regions which meet P (Fig. 26).

We look at the boundary of this neighbourhood of P. It consists of a number of sides of fundamental regions. We know that $\bar{f}$ leaves P fixed. Hence if we examine the repeated action of $\bar{f}$ on $\bar{h}(\beta)$, it may be considered as representing repeted small rotations about P in the same direction, approaching a certain limit point. (This follows from Nielsens work.) If $\bar{h}(\beta)$ intersects the same side of the boundary of the neigbourhood of P after arbitrarily many applications of both $\bar{f}$, and of $\bar{f}^{-1}$, then we are in the case of section 2.9. If not, then we are in the case of 2.8.

3.3 The proof of the conjugacy theorem

Let S be a compact, orientable surface, and let f, g be homeomorphisms $S \to S$ which agree on ∂S. We wish to determine whether or not f and g commute; that is, does there exist a homeomorphism $h\colon S \to S$ which is such that $h^{-1}fh$ is isotopic to g, by an isotopy which leaves ∂S fixed?

In order to answer this question, we must first determine if there are any non-singular periodic curves in S under either f or g. If there are, they can be found directly,

using the procedure described by Nielsen in [10]. Alternatively, they may be found by transforming the problem into the realm of 3-manifolds and then using Haken's theory. (First take the product $S \times I$, where I is the unit interval, then glue the ends together, using either f or g. The resulting space is a Stallings fibration, and the problem is to find incompressible annuli or tori in this space.)

We may split S along all such non-singular periodic curves under f, and g, obtaining the surfaces S_f, and S_g, respectively, If S_f and S_g are not homeomorphic, by a homeomorphism which is the identity on ∂S, then f and g cannot be conjugate. Hence we may assume that there exists a homeomorphism $k: S \to S$, which is the identity on ∂S. (It is an easy matter to see whether or not such a homeomorphism exists, and if one does, to construct it.)

Let us now denote S_g by the symbol S^*, and denote $k^{-1}f|_{S_f}k$ by f^*, and $g|_{S_g}$ by g^*. The problem then is to determine whether or not f^* and g^* are conjugate on S^*. In this new setting we are guaranteed of having no non-singular periodic curves under either f^* or g^*. In fact it is possible to say even more. One can show, by an easy argument involving Nielsen's constructions [9], that the existence of a singular periodic curve implies the existence of a non-singular one (see Johannson [6]). Therefore we may assert that there are no periodic curves under either f^* or g^*. But then we may apply our theorem:

First determine $N(f^*, g^*)$, and then check all homeomorphisms of size no greater than this number on S^* to see if they conjugate f^* and g^*. If a conjugating homeomorphism h^* exists which is such that $h^{*-1}f^*h^*$ is isotopic to g^*, by an isotopy which is constant on ∂S, then for some integer n, $f^{*n}h^*$ will be of size no greater than $N(f^*, g^*)$ after being straightened, as in sections 1.5 and 1.6. That is, $f^{*n}h^*t$ is of size no greater than $N(f^*, g^*)$, where t is a homeomorphism which is isotopic to the identity on S, and involves only rotations of the boundaries about themselves. But clearly if $h^{*-1}f^*h^*$ is isotopic to g^*, by an isotopy which is constant on ∂S, then also the same is true for $(f^{*n}h^*t)^{-1}f^*(f^{*n}h^*t) = (h^*t)^{-1}f^*(h^*t)$.

Hence this conjugating homeomorphism will be found, and so the conjugacy problem will be solved for f^* and g^* in S^*. But then, by extending h^* across the periodic curves in S to a homeomorphism $h: S \to S$, we obtain a solution to the conjugacy problem for f and g in S.

References

[1]. HAKEN, W., Theorie der Normalflächen, ein Isotopiekriterium für den Kreisknoten. *Acta Math.*, 105 (1961), 245–375.

[2]. —— Ein Verfahren zur Aufspaltung einer 3-Mannigfaltigkeit in irreduzible 3-Mannigfaltigkeiten. *Math. Z.*, 76 (1961), 427–467.

[3]. —— Über das Homöomorphieproblem der 3-Mannigfaltigkeiten I. *Math. Z.*, 80 (1962), 89–120.

[4]. —— Some results on surfaces in 3-manifolds. In P. J. Hilton (ed.): *Studies in modern topology*, MAA Studies.

[5]. —— Connections between topological and group theoretical decision problems. In W. W. Boone et al. (eds.): *Word problems*, North-Holland, Amsterdam 1973, 427–441.

[6]. JOHANNSON, K., Homotopy equivalences of knot spaces. Univ. Bielefeld, Feb. 1976.

[7]. —— Homotopy equivalences of 3-manifolds with boundary. Univ. Bielefeld, Aug. 1976.

[8]. Nielsen, J., Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen I. *Acta Math.*, 50 (1927), 189–358.

[9]. —— Untersuchungen zur Topologie der geschlossenen zweiseitigen Flächen II. *Acta Math.*, 53 (1929), 1–76.

[10]. —— Surface transformation classes of algebraically finite type. *Danske Vid. Selsk. Math.-Phys. Medd.*, 21, no. 2 (1944).

[11]. SCHUBERT, H., Bestimmung der Primfaktorzerlegung von Verkettungen. *Math. Z.*, 76 (1961), 116–148.

[12]. WALDHAUSEN, F., Recent results on sufficiently large 3-manifolds. *Proceedings of Symposia in Pure Math.*, 32 (2), 21–38, August 1976.

Received March 29, 1978