The Curve Graph and Surface Construction in $S \times R$

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Ingrid Irmer
aus
Brisbane, Australien

Bonn Juli, 2010

Angefertigt mit Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhems-Universität Bonn

1. Gutachter: Frau Professor Dr. Hamenstädt
2. Gutachter: Herr Professor Dr. Ballmann
Tag der Promotion: 10.12.2010
Erscheinungsjahr: 2011

Danksagung

Ich möchte mich bei meiner Doktormutter, Frau Hamenstädt (auch bekannt als “die Chefin”) für ihre sehr gründliche und pflichtbewusste betreuung dieses Projekts, ganz besonders beim mehrmaligen korrekturlesen dieser Dissertation, bedanken. Ich habe viel von Ihrem pädagogischen Geschick und ausgeprägtem Fachwissen profitiert. Vielen Dank an Michael Joachim für geduldige topologische Nachhilfestunden und für seine freundlichen Ratschläge sowie an Juan Souto für die Einführung in die hyperbolische Geometrie. Joachim Vaerst hat mir geholfen, mich in diesem fremden Land zu integrieren und war jedesmal bereit, mich aus sämtliche Fettnäpfchen zu retten. Danke auch an Vivian Easson, Piotr Przytycki and Samuel Tapie für ihre arbeit bei der übersetzung von Bonahon.

Abstract

Suppose S is an oriented, compact surface with genus at least two. This thesis investigates the "homology curve complex" of S; a modification of the curve complex first studied by Harvey in which the verticies are required to be homologous multicurves. The relationship between arcs in the homology curve graph and surfaces with boundary in $S \times R$ is used to devise an algorithm for constructing efficient arcs in the homology curve graph. Alternatively, these arcs can be used to study oriented surfaces with boundary in $S \times R$. The intersection number of curves in $S \times R$ is defined by projecting curves into S. It is proven that the best possible bound on the distance between two curves c_0 and c_1 in the homology curve complex depends linearly on their intersection number, in contrast to the logarithmic bound obtained in the curve complex. The difference in these two results is shown to be partly due to the existence of what Masur and Minsky [19] refer to as large subsurface projections of c_0 and c_1 to annuli, and partly due to the small amount of ambiguity in defining this concept. A bound proportional to the square root of the intersection number is proven in the absence of a certain type of large subsurface projections of c_0 and c_1 to annuli.

Contents

Chapter 1

Introduction

Suppose that S is an oriented, compact, connected surface with genus g at least two. The complex of curves is an abstract, finite dimensional, locally infinite complex associated with a surface, originally introduced by Harvey in [13].

Definition 1 (Curve Complex)
The complex of curves is the simplicial complex whose vertex set $\mathcal{C}(S)$ is the set of all nontrivial free homotopy classes of simple closed curves on S. A collection $c_1, ...c_k \subset \mathcal{C}(S)$ spans a simplex if and only if $c_1, ...c_k$ can be realized disjointly. The curve graph is the one skeleton of the curve complex. Distance is defined by assigning each edge length one.

The curve graph has since proven to be a useful tool in studying Teichmüller Spaces, the mapping class group and the structure of 3-manifolds, for example [9], [12], [22] and [15]. In particular, it played an important role in the proof of Thurston's ending lamination conjecture. This thesis investigates the "homology multicurve complex" of S; a modification of the curve complex in which the verticies are required to be oriented multicurves in a fixed homology class.

Definition 2 (Mapping class group)
The mapping class group, Mod_g, is the group of homotopy classes of orientation preserving homeomorphisms of a closed, oriented surface S of genus g onto itself.

Definition 3 (Torelli group)
The Torelli group is the subgroup of the mapping class group that acts trivially on homology.

An element of the mapping class group therefore induces an isometry of the curve graph onto itself. It is well known, e.g. [9], Theorems 4.2 and 4.10, that for any two nonseparating curves in S, there is an element of Mod_g that maps one curve to the other. Similarly, if c is a separating curve in S such that one component of $S \setminus c$ has Euler characteristic χ_1 and the other component of $S \setminus c$ has Euler characteristic χ_2, there exists an element of the mapping class group that maps c into any other separating curve that separates S into two components, one with Euler characteristic χ_1 and the other with Euler characteristic χ_2. In other words, although the isometry group of the curve graph does not act transitively, there are only finitely many orbits. Similarly, the Torelli group induces an isometry of the homology multicurve complex onto itself. The action of the Torelli group preserves the number of connected components of a multicurve, and verticies in the homology multicurve graph can have arbitrarily many components, i.e. the homology multicurve complex is infinite dimensional. It follows that there are infinitely many orbits of verticies of the homology multicurve complex. In chapter three it is shown that there exist multicurves with arbitrarily many connected components that do not contain null homologous submulticurves and that are homologous to a fixed, oriented curve c_0. Infinite dimensionality of the homology multicurve complex is therefore not a property that can be made to disappear by requiring that the verticies do not contain null homologous submulticurves.

The main difficulty involved in working with the curve graph and its relatives is that it is not locally compact. In order to address this problem, the concept of a "tight geodesic" was introduced in [18] and modified slightly by Bowditch in [5]. Bowditch's definition of "tightness" can also be applied in the context of the homology multicurve graph, and all arcs constructed in this thesis will also be tight. It was shown in [19] that there are only finitely many tight geodesic arcs connecting any two verticies in the curve graph, and [17] and [26] independantly showed that distance in the curve graph is computable and developed an algorithm for calculating the distance between two verticies.

Two oriented curves c_0 and c_1 in a 3-manifold are homologous iff there exists an embedded surface H in $S \times R$ with $\partial H = c_1 - c_0$. (Lemma 1 of [27]).

It will be shown that surfaces in $S \times R$ with boundary $c_1 - c_0$ give considerable information about arcs in the homology curve complex with endpoints c_0 and c_1, and in reverse, the homology curve complex sheds light on the surfaces themselves. This makes the problem of calculating distances and constructing geodesic arcs much simpler in the homology multicurve complex than in the curve complex. The homological invariance of the intersection form on curves is used to define a locally constant function f on $S \setminus (c_0 \cup c_1)$, and this is shown to be related to the projection to $S \times 0$ of a surface in $S \times R$ with boundary curves $c_1 - c_0$. In particular, an algorithm is devised for constructing efficient arcs in the homology multicurve graph. Whenever c_0 and c_1 are homologous, simple curves in $S \times 0$, it is shown that the smallest genus surfaces in $S \times R$ with boundary curves freely homotopic to $c_1 - c_0$ can be constructed from an arc in the homology curve graph with endpoints c_0 and c_1, of the type constructed by the given algorithm. Alternatively, the Euler integral of f is related to the Euler characteristic of a smallest genus surface in $S \times R$ with boundary curves freely homotopic to $c_1 - c_0$. This is analogous to the situation in Euclidian three space, in which a projection of a link into a plane is used to construct an oriented surface (the "Seifert surface") with the given link as boundary.

The intersection number of curves in $S \times R$ is defined by projecting curves into S, and a family of examples is given to show that the best possible bound on the distance between two curves c_0 and c_1 in the homology curve graph depends linearly on their intersection number. This differs from the curve complex, in which an upper bound on the distance proportional to the logarithm of the intersection number is shown in [5]. The difference in these two results is shown to be partly due to the existence of what Masur and Minsky [19] refer to as large subsurface projections of c_0 and c_1 to annuli ("twisting"), and partly due to the small amount of ambiguity in defining this concept. Suppose two multicurves m_1 and m_2 both intersect an annulus A. Distance between two curves in the subsurface projection to A is related to the number of times a component of $m_1 \cap A$ is Dehn twisted in relation to a component of $m_2 \cap A$. In order to make this concept well defined, it is necessary to make use of properties of covering spaces of hyperbolic surfaces. A major source of difficulties is that most quantities dealt with here are only defined up to free homotopy, but without a metric on S, distance between two multicurves in the subsurface projection to an annulus is only defined up to plus or minus one. In the absence of a certain type of large subsurface projections of c_0 and c_1 to annuli, a bound on the distance between

c_0 and c_1 in the homology multicurve graph proportional to the square root of the intersection number of c_0 and c_1 is proven. This is done by using the concept of an interval exchange map to relate the function f, the absence of large subsurface projections to annuli and the Euler characteristic of S. The ambiguity in the definition of distance in the subsurface projection is used to construct an example of an interval exchange map that is self-similar on arbitrarily small subintervals. This interval exchange map is obtained from a limit of homologous curves without large subsurface projections to annuli, and shows that it is not possible to obtain better than a bound depending on the square root of the intersection number.

In [18] it was shown that the curve complex is δ-hyperbolic. It is known that the mapping class group is not hyperbolic, since it contains abelian subgroups generated by Dehn twists around disjoint curves, however it was shown in [18] that the mapping class group is relatively hyperbolic with respect to left cosets of a finite collection of stabilizers of loops. The discrepancy between distances in the homology multicurve graph and distances in the curve graph would seem to reflect the fact that there are abelian subgroups of the mapping class group that leave distances unchanged in the curve graph but not in the homology multicurve graph. As a result, the homology multicurve graph is not hyperbolic. A similar results along these lines is given in Theorem 1.1 of [8], in which it was shown that for a surface of genus at least 3, the distortion of the Torelli group as a subgroup of the mapping class group with respect to the word norm is exponential.

Chapter 2

Surfaces and the Curve Complex

2.1 The Function

Suppose $M \cong S \times R$, where S is a closed oriented connected surface with genus $g \geq 2$, and π is a choice of first factor projection function of M onto $S \times 0$. To simplify the notation, the submanifold "$S \times 0$" will often be referred to as S, not to be confused with the circle S^1. All curves, surfaces, and manifolds will be assumed to be piecewise smooth, except in section 2.2.

Definition 4 (Curve)
A curve c in M is a free homotopy class of piecewise linear maps of S^1 into M such that

1. c has a representative that is embedded in S
2. c is not contractible

A curve in S is defined similarly. In practice, whenever it is clear from the context what is meant, the term "curve" will also refer to the image in M or S of a particular representative of the homotopy class of maps.

Definition 5 (Multicurve)
A multicurve on S is a union of curves in S with representatives whose images can all be realised disjointly. In general, some of these curves might be freely

homotopic. A multicurve in M is a union of curves that projects onto a multicurve in S.

Definition 6 (Intersection Number)
If a and b are two multicurves in M, their intersection number $i(a,b)$ is the smallest possible number of points of intersection between a projection to S of a representative of the homotopy class a and a projection to S of a representative of the homotopy class b.

Definition 7 (Essential intersections)
Supose a and b are multicurves in S in general position. An essential point of intersection of a with b is a common boundary point of two arcs a_1 and a_2 of $a \cap (S \setminus b)$, such that neither a_1 nor a_2 is homotopic in S relative to its endpoints to an arc of $b \cap (S \setminus a)$.

The techniques for constructing "surfaces" in this thesis are quite general, and as a result, the "surfaces" constructed are not always embedded. In the definition of "surface in M" it is therefore convenient to allow self-intersections. If embeddedness is an issue, it will be shown that it is always possible to obtain an embedded surface with minimal genus.

Definition 8 (Surface (with boundary) in M)
A surface in M is a piecewise linear map ϕ from an orientable surface F into M whose image is locally embedded, and such that $\phi(\partial F)$ is embedded. In section 2.2, it is convenient to work in the smooth category, so surfaces will be assumed to be smooth.

Since surfaces are allowed to have boundaries, it is necessary to define what a homotopy is allowed to do to the boundary of a surface.

Definition 9 (Homotopy of Surfaces)
Let $\phi_1 : F \to M$ and $\phi_2 : F \to M$ be surfaces in M with images F_1 and F_2 respectively. F_1 will be said to be homotopic to F_2 if a homotopy $H(s,t)$ between ϕ_1 and ϕ_2 can be found such that $H(s,t*) : F \to M$ is locally embedded for any fixed $0 \leq t* \leq 1$. $H(s,t*) : F \to M$ is not required to have embedded boundary for $0 < t* < 1$.

The definition of general position given in definition 1.11 of [14] will be used throughout this thesis. This definition assumes that all objects are in the piecewise linear category, which will also make sense in the smooth category used in section 2.2, since this is a special case of the piecewise linear category. In this work, the representative of the homotopy class of a surface is not usually important, so it is possible to assume without loss of generality that a surface is in general position, because every homotopy class contains representatives in general position. If a surface is in general position, it was shown in [14] that its self-intersections are all transverse and can consist only of arcs with endpoints on the boundary, closed curves and isolated triple points.

It will sometimes be convenient to work with representatives of homotopy classes of curves that are not in general position. Since intersection numbers are defined to be properties of homotopy classes, intersection numbers are still well defined in this case.

In this chapter, a sequence of homologous multicurves will be used to construct a surface H in M with oriented boundary curves c_1 and $-c_0$. In order to do this a choice of representatives of the homotopy classes c_0 and c_1 is used to define a function f on a subset of S. If these representatives of the homotopy classes only have essential points of intersection, the resulting function determines the minimum number of multicurves in a sequence needed to construct a surface with oriented boundary curves $c_1 - c_0$. f is then used to give an explicit construction of H, similar to a handle decomposition given by a Morse function.

Definition 10 (Homology Intersection Number)
The homology intersection number of two elements α and β of $H_1(S)$ is equal to the intersection product $\alpha \bullet \beta$ defined, for example, on page 367 of [7]. If a is an oriented curve in M whose projection into S belongs to the homology class α and b is an oriented curve in M whose projection into S belongs to the homology class β then the homology intersection number, $i_h(a, b)$, is defined to be equal to $\alpha \bullet \beta$.

For the arguments given here, it is more convenient to treat the homology intersection number of two curves as the signed intersection number. Suppose c and d are oriented, embedded representatives of a homotopy class of curves in S. Suppose also that a and b are in general position and intersect in at least one point p. By assumption, all curves are piecewise smooth, so it follows from the orientability of S that c has an annular neighbourhood $N(c)$, where

$N(c) \setminus c$ consists of two connected components. $N(c)$ can also be chosen such that any component of $d \cap N(c)$ is an arc with endpoints on the two distinct components of $\partial N(c)$. The orientations of S and c are sufficient to determine which component of $N(c) \setminus c$ is to the right of c and which component is to the left. The orientation of d provides an ordering of the points in each of the components of $d \cap N(c)$. d will be said to cross over c from left to right at p if the component of $d \cap N(c)$ passing through p enters the component of $N(c) \setminus c$ to the left of c before it enters the component of $N(c) \setminus c$ to the right of c. Right to left is defined analogously. If $d(t)$ is an oriented arc with starting point $p = d(0)$ on c, it will be said to leave c from the right if there is a $t*$ such that for $t < t*$, $d(t)$ is contained in the component of $N(c) \setminus c$ to the right of c. d approaches c from the right if there is a $t*$ such that for $t* < t$, $d(t)$ is contained in the component of $N(c) \setminus c$ to the right of c. Leaving and approaching c from the left is defined similarly.

The boundary of an oriented surface will be given the usual boundary orientation. In particular, all arcs in the surface with endpoints on the boundary leave or approach the boundary from the left.

Definition 11 (Homology Intersection Number (Alternative definition))
If a and b are oriented multicurves in S, choose embedded representatives a' and b' of the free homotopy classes in S, $[a]$ and $[b]$, that are in general position. $i_h(a, b)$ is equal to the number of points of intersection at which b' crosses over a' from right to left minus the number of points of intersection at which b' crosses over a' from left to right. If a and b are multicurves in M, define the homology intersection number by projecting them onto S.

In chapter 17 of [10], it was shown that the first definition is the same as the second definition for curves on a closed oriented surface. In particular, the second definition is independent of the representative of the homology class.

It is easy to check that at any point at which a crosses over b from right to left, b crosses over a from left to right, so $i_h(a, b) = -i_h(b, a)$.

It follows that any multicurve has zero homology intersection number with itself, and therefore with any multicurve homologous to itself. Also, a null homologous multicurve has zero homology intersection number with any other curve.

Definition 12 (Homology Multicurve Complex)

The homology multicurve complex is defined analogously to the curve complex. It is a simplicial complex whose verticies are oriented multicurves in a given homology class on S. A set of verticies bounds a simplex if each pair of curves has pairwise zero intersection number.

It has not yet been shown that the homology multicurve complex is connected. An algorithm for connecting any two homologous curves by an arc in the homology multicurve graph will be developed. It will become clear that this algorithm can be modified slightly to apply to multicurves, from which connectivity of the homology multicurve complex follows.

Definition 13 (Homology Multicurve Graph)
The homology multicurve graph is the one skeleton of the homology multicurve complex. Distance is defined by giving adjacent verticies distance one.

The reason for introducing the homology multicurve graph is that an arc with endpoints c_0 and c_1 in the homology multicurve graph of S will be used to construct a surface in M. A purpose of the next few lemmas is to outline an algorithm for constructing an arc in the homology multicurve graph, $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$. It will be shown that an arc constructed in this way is the most efficient arc with the desired properties.

Definition 14 (Boundary of a subset)
Given an oriented null homologous multicurve n, the surface $S \setminus n$ consists of two or more connected components. The boundary of each of the components of $S \setminus n$ is given the standard boundary orientation of a subset of S. If the boundary orientation of each component either agrees with the orientation of n or the orientation of $-n$, n will be said to bound a subset of S. The union of the components of $S \setminus n$ whose boundary orientation coincides with the orientation of n will be called the "subset of S bounded by n".

The figure 2.1 shows a null homologous multicurve that does not bound a subset of S.

Lemma 15
If a null homologous multicurve does not contain a nontrivial null homologous submulticurve, it has to bound a subset of S.

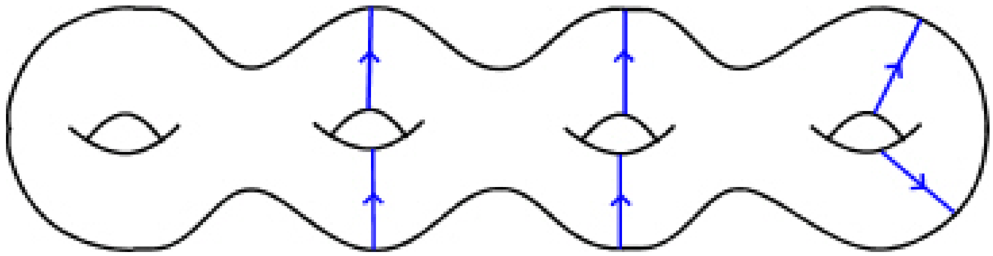

Figure 2.1: A null homologous multicurve that does not bound a subset of S

Proof. Suppose that a null homologous multicurve n does not contain any null homologous submulticurves. Since n is null homologous, $\backslash n$ is not connected. Since n does not contain null submulticurves, for any curve $d \in n$, $S \setminus (n \setminus d)$ is connected. It follows that $S \setminus n$ has two connected components. If n does not bound a subset of S, there are curves a and b in n such that one of the components of $S \setminus n$ is to the right of a and to the left b. Let x be a point on a and y be a point on b. Using connectivity, there is an oriented arc a_{xy} in one component of $S \setminus n$ with starting point x and endpoint y, and an oriented arc a_{yx} in the other component of $S \setminus n$ with starting point y and endpoint x. The curve obtained from connecting a_{xy} to a_{yx} has nonzero homology intersection number with n, which contradicts the assumption that n is null homologous.

□

If a null homologous multicurve n that bounds a subset of S doesn't contain null homologous submulticurves, it can't contain freely homotopic curves with opposite orientations (unless, of course, it only contains these two curves). It also can't contain freely homotopic curves with the same orientation, because then it wouldn't bound a subset of S. Therefore the number of curves in n is bounded above by $-\frac{3\chi(S)}{2}$.

Definition 16 (Surface Producing Sequence)
An arc $\{\gamma_i\}$ in the homology multicurve graph of S is "surface producing" if

1. for each i, $\gamma_{i+1} - \gamma_i$ bound a subset of S,

2. none of the γ_i contain freely homotopic curves with opposite orientation.

To see where the name "surface producing" comes from, suppose $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$ is surface producing. Let Γ_1 be the subset of $S \times 1$ that projects onto the subset of the surface $S \times 0$ bounded by $\gamma_1 - c_0$. γ_i was defined to be a multicurve in S, but a multicurve in S determines a free homotopy class in M, which can be projected into $S \times k$ for any k. Fixing a choice of projection function, there is a multicurve in $S \times k$ that projects onto γ_i in S, and to keep the notation simple, this multicurve will also be called γ_i. Similarly, let Γ_i be the subset of $S \times i$ that projects onto the subset of $S \times 0$ bounded by $\gamma_i - \gamma_{i-1}$. Let A_1^2 be the annulus in $S \times R$ given by $\pi^{-1}(\gamma_1) \cap S \times [1, 2]$ and let Δ_2 be the surface with boundary $\gamma_2 - c_0$ consisting of $\Gamma_1 \cup A_1^2 \cup \Gamma_2$. Δ_2 is clearly embedded. This process can be repeated to obtain Δ_{j+1} with boundary curves $c_1 - c_0$.

Without any further restrictions on the sequence γ_i, for $2 < j + 1$, Δ_{j+1} obtained in this way is not necessarily embedded and it is necessary to show that it is a surface. This has to do with the problem that there could exist a curve c such that several consecutive multicurves contain a curve freely homotopic to c. Suppose for example that c is contained in $\gamma_i, \gamma_{i+1}...\gamma_{i+n}$ but c is not a subset of γ_{i-1} or γ_{i+n+1}. There are two possibilities; either each of the multicurves $\gamma_{i+1}...\gamma_{i+n}$ contain only one curve freely homotopic to c, or there could be multicurves that contain several curves freely homotopic to c. In the first case, Δ_{j+1} contains a cylinder of the form $c \times [i, i + n + 1]$, i.e. there is a cylinder that climbs up many levels at once. This cylinder could transversely intersect one or several of Γ_j for $i < j < i + n + 1$. The second case is a bit more complicated because Δ_{j+1} contains several long cylinders of the form $c \times I$, where I is a closed interval in R. These cylinders do not intersect transversely. The construction of Δ_{j+1} can be altered slightly to remove this problem. For any curve c that gives rise to nontransversal self-intersections of Δ_{j+1} as described above it is possible to find $n + 1$ representatives of the free homotopy class $[c]$ that are all pairwise disjoint, call them $c^0, c^1, c^2...c^n$. Then let $\gamma_i^{'}$ be the multicurve γ_i with c replaced by c^0. If γ_{i+1} contains two curves freely homotopic to c, let $\gamma_{i+1}^{'}$ be the multicurve γ_{i+1} with one curve freely homotopic to c replaced by c^0 and the other curve freely homotopic to c replaced by c^1. If γ_{i+1} has only one curve freely homotopic to c then $\gamma_{i+1}^{'}$ is obtained by replacing this curve by c^0. Continuing in this way, the multicurves $c_0, \gamma_1, ...\gamma_i^{'}, \gamma_{i+1}^{'}...c_1$ can be used to construct the image in M of a representative of a homotopy class of Δ_{j+1} representing a surface with self-intersections. In chapter three it will be shown that the sequence $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$ can always be constructed so that this second problem doesn't occur.

Suppose ϕ is a parametrisation of Δ_{j+1}, i.e. $\phi : F \to M$ has image Δ_{j+1}. It can be assumed without loss of generality that Δ_{j+1} is in general position, and in this case it was seen that the self-intersections of Δ_{j+1} consists of a union of closed curves whose preimages in F each have two connected components. For each closed curve in Δ_{j+1} along which Δ_{j+1} intersects itself, it is possible to perform surgeries to remove the self-intersection. (To be more specific: suppose c is a curve in M along which Δ_{j+1} itersects itself, and let $\mathcal{N}$ be a neighbourhood of c in M such that $\mathcal{N} \cap \Delta_{j+1}$ consists of two annuli, each with boundary curves freely homotopic to c. $\partial\mathcal{N} \cap \Delta_{j+1}$ consists of four curves, each freely homotopic to c. These four curves bound a pair of disjoint oriented annuli, $\mathcal{A}_1$ and $\mathcal{A}_2$, in $\mathcal{N}$, such that $\partial\mathcal{A}_1$ and $\partial\mathcal{A}_2$ have the opposite orientations of the corresponding curves on the boundary of $\Delta_{j+1} \setminus \mathcal{N}$. Gluing $\mathcal{A}_1$ and $\mathcal{A}_2$ to $\Delta_{j+1} \setminus \mathcal{N}$ along the common boundary curves, an oriented cell complex without self-intersection along c is obtained.) After performing all such surgeries, a set of cell complexes is obtained, where each of these cell complexes can be embedded in M. Neither c_1 nor c_0 is null homologous, and Δ_{j+1} did not have any points of self-intersection on the boundary. Therefore, one of the cell complexes obtained after performing the surgeries has to have boundary $c_1 - c_0$. In this way, an embedded surface with boundary curves $c_1 - c_0$ and genus no larger than that of Δ_{j+1} is obtained. In the next section it will be shown that all surfaces with boundary $c_1 - c_0$ and smallest possible genus can be constructed from a surface producing sequence, and so there always exists a surface producing sequence that can be used to construct an embedded surface.

Definition 17
Denote the surface obtained as described in the previous paragraphs from a surface producing sequence γ as H_γ and call H_γ the surface passing through γ.

The second part of the definition of "surface producing" is not necessary to construct a surface, but it will be useful later to have this condition in order to obtain a bound on the number of curves in each multicurve.

Definition 18 (Horizontal and Vertical Arcs)
Given two multicurves a and b on an oriented surface S, a horizontal arc of a is a component of $a \cap (S \setminus b)$ that leaves and approaches b from the same

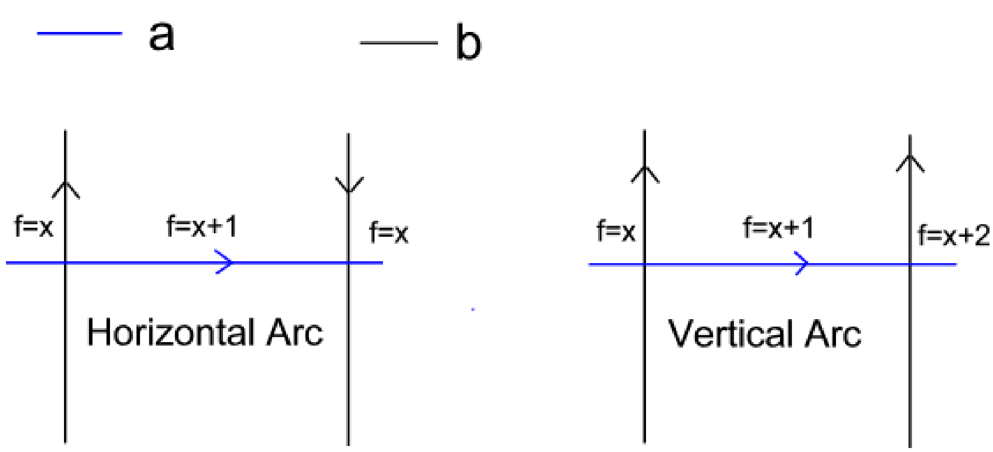

Figure 2.2: Horizontal and vertical arcs

side. A vertical arc of $a \cap (S \setminus b)$ leaves and approaches b from opposite sides.

If a horizontal arc of $a \cap (S \setminus b)$ leaves and approaches b from the right, then this arc is "to the right of b" and vice versa.

Definition 19 (Homotopic Arcs)
Suppose a and b are multicurves in S in general position. Two arcs of $a \cap (S \setminus b)$ will be called homotopic if they are homotopic relative to b. Two oriented arcs will be said to be homotopic and oriented in the same way if one can be homotoped into the other in such a way that the orientations coincide.

Definition 20 (Adding a handle to the multicurve b corresponding to a horizontal arc a_i of $a \cap (S \setminus b)$)
Let R be an oriented embedded rectangle in S whose interior is contained in $S \setminus (a \cup b)$. Suppose that one side of R lies along the arc a_i, the opposite side is homotopic to a_i with opposite orientation, and the two remaining sides are subarcs b_1 and b_2 of b. Since a_i is a horizontal arc, it is possible to choose R such that the orientation of R induces an orientation on the arcs b_1 and b_2 on its boundary opposite to the orientation of b_1 and b_2 as subsets of b. Adding a handle to the oriented multicurve b corresponding to a horizontal arc a_i of $a \cap (S \setminus b)$ involves adding ∂R to b as a chain. The arcs b_1 and b_2 on the boundary of R cancel out subarcs of b and are replaced by the arcs a_i and $-a_i$. Since ∂R is null homologous, the resulting multicurve is homologous to

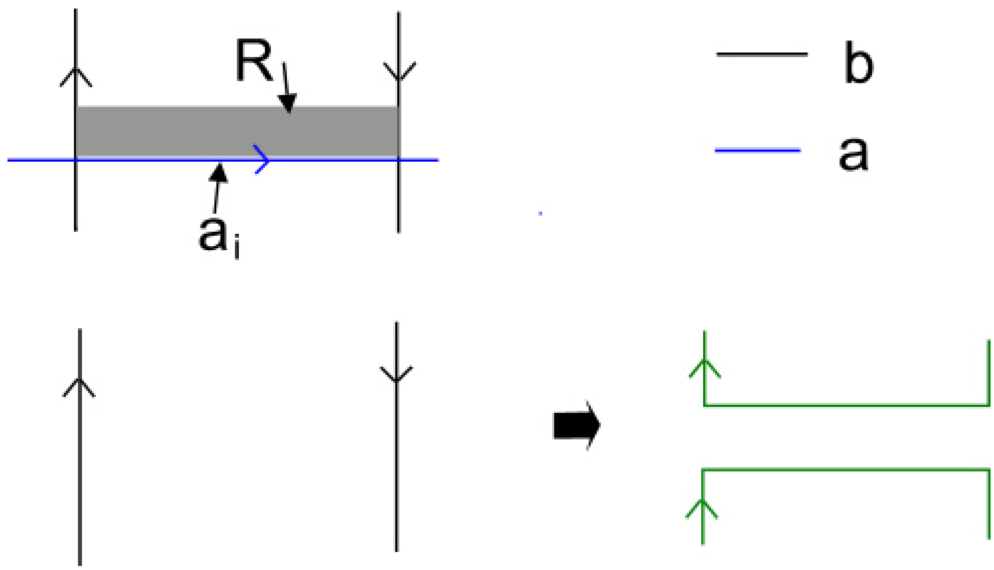

Figure 2.3: Adding a handle corresponding to a horizontal arc.

b.

Lemma 21
"verticalness" and "horizontalness" are properties of homotopy classes of arcs. Also, if a and b only have essential intersections, a horizontal arc of $a \cap (S \setminus b)$ to the right of b can't be homotopic to a horizontal arc of $a \cap (S \setminus b)$ to the left of b.

Proof. Suppose v is a vertical arc of $a \cap (S \setminus b)$ whose endpoints are both on the same component of the multicurve b. Let c_v be a curve formed by connecting up the endpoints of v on b. If v were homotopic to a horizontal arc, the homology intersection number of c_v with b could be changed by a homotopy, which is a contradiction. For the same reason, a horizontal arc with both endpoints on the same curve of b can't be homotopic to a vertical arc.

Now suppose that v is a vertical arc of $a \cap (S \setminus b)$ with endpoints on two distinct curves, b_1 and b_2 of b. If v is homotopic to a horizontal arc h, let d be the oriented curve homologous to $b_1 \cup b_2$ formed by adding a handle to $b_1 \cup b_2$ corresponding to h, and let c_v be the curve formed by connecting up the endpoints of v on d. The homotopy that takes v to h changes the homology intersection number of c_v with d, which is a contradiction. For the same reason, a horizontal arc of $a \cap (S \setminus b)$ with endpoints on two different curves in b can't be homotopic to a vertical arc.

Let h be a horizontal arc of $a \cap (S \setminus b)$ to the right of b. If h were homotopic to a horizontal arc of $a \cap (S \setminus b)$ to the left of b, this homotopy would decrease the intersection number of a with b, contradicting the assuption that a and b only have essential points of intersection. For the same reason, a horizontal arc of $a \cap (S \setminus b)$ to the left of b can't be homotopic to a horizontal arc to the right of b. □

Lemma 22
Let a and b be oriented multicurves in general position that only have essential points of intersection. An oriented arc of $a \cap (S \setminus b)$ is not homotopic to itself with the opposite orientation.

Proof. If an oriented arc a_1 of $a \cap (S \setminus b)$ has its endpoints on two different curves in the multicurve b, a_1 can't be homotopic to $-a_1$, because a homotopy of a_1 to $-a_1$ would change the component of b on which the arc has its starting point. Let v be a vertical arc of $a \cap (S \setminus b)$ with both endpoints on the component b_1 of b, and let c_v be the curve formed by connecting the endpoints of v by a subarc of b. v can't be homotopic to $-v$ because such a homotopy would change the homology intersection number of c_v with b_1. The only other possibility is that there could be a horizontal arc h with both endpoints on the curve b_1 in b. Since h is not homotopic with fixed endpoints to a subset of c (this follows from the assumption that a and b only have essential points of intersection), the tubular neighbourhod of $b_1 \cup h$ is a pant P with incompressible boundary. Any arc homotopic to h is homotopic to an arc contained inside P, so it is possible to assume without loss of generality that a homotopy that takes h to $-h$ only passes through arcs contained within P. P has a boundary curve freely homotopic to b_1, call the other two boundary curves of P p_1 and p_2. Let d be an oriented arc with starting point on p_1 and endpoint on p_2. P can be embedded inside a surface $S^{'}$ such that d is the intersection with P of an oriented curve $d^{'}$ in $S^{'}$ and h is the intersection of an oriented curve $h^{'}$ in $S^{'}$, where $d^{'}$ and $h^{'}$ only intersect at a single point inside P. h can't be homotopic to $-h$ in $S^{'}$, because this would change the homology intersection number of $d^{'}$ with $h^{'}$. It follows that h can't be homotopic to $-h$ within P.

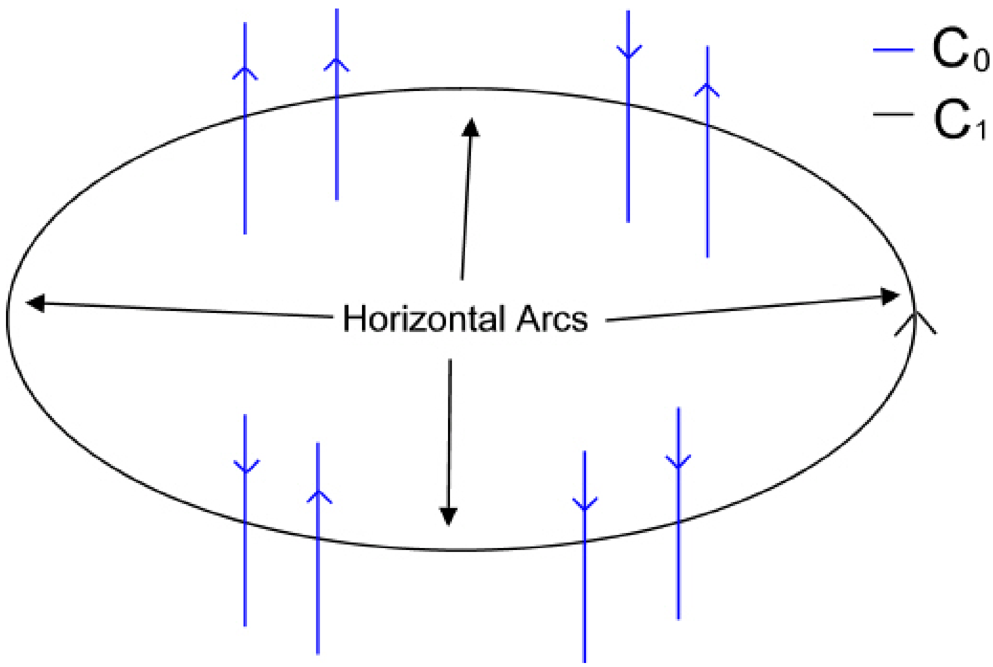

Figure 2.4: Existence of horizontal arcs

□

Since the homological intersection number of c_0 and c_1 is zero, there has to be as many points of intersection of c_0 with c_1 where c_0 crosses over c_1 from right to left as points of intersection where c_0 crosses over c_1 from left to right. This guarantees the existence of at least two horizontal arcs of $c_1 \cap (S - c_0)$, as shown in figure 2.4. It is being assumed here that c_0 and c_1 are representatives of their free homotopy classes that only have essential points of intersection, so whenever c_1 passes through an arc of $c_1 \cap (S - c_0)$ that leaves c_0 from the right and approaches c_0 from the left, it has to pass through a horizontal arc to the right of c_0 before passing through an arc that leaves c_0 from the left and approaches c_0 from the right. Therefore, if c_1 intersects c_0, there has to be a horizontal arc of $c_1 \cap (S - c_0)$ to the right of c_0. A symmetric argument shows that there also has to be a horizontal arc of $c_1 \cap (S - c_0)$ to the left of c_0.

For each horizontal arc of $c_1 \cap (S \setminus c_0)$ to the right of c_0 add a corresponding handle to c_0 as shown in figure 2.3, to obtain a multicurve disjoint from c_0 and with smaller intersection number with c_1. Discard all null homologous (multi)curves and denote the remaining multicurve as α_1. In particular, α_1 does not contain two freely homotopic curves with opposite orientation.

By the same argument as before, there is at least one horizontal arc, b_1, in $c_1 \cap (S \setminus \alpha_1)$ to the right of α_1. The only difference is that b_1 or another horizontal arc might have endpoints on two different curves. This process can therefore be repeated to obtain a multicurve α_2, etc, until a multicurve α_j is found that does not intersect c_1. Given the arc $c_0, \alpha_1, ...\alpha_j, c_1$ in the homology multicurve graph, if this sequence is not surface producing, a surface producing sequence is obtained as follows. If $c_0, \alpha_1, ...\alpha_j, c_1$ is not surface producing, then for some i, $\alpha_{i+1} - \alpha_i$ is a null homologous submulticurve that does not bound a subset of S. In lemma 15 it was shown that any null homologous multicurve can be decomposed into a union of null homologous multicurves, each of which bounds a subset of S. Let n be a submulticurve of $\alpha_{i+1} - \alpha_i$ that bounds a subset of S, $\delta_1^{'} := \alpha_i \cup n$, and δ_1 be the multicurve obtained from discarding all pairs of freely homotopic curves in $\delta_1^{'}$ with opposite orientation. $\delta_1 - \alpha_i$ bounds a subset of S consisting of the subset of S bounded by n plus a union of annuli. If $\alpha_{i+1} - \delta_1$ does not bound a subset of S, δ_2 is constructed in the same way as δ_1 only with δ_1 in place of α_i, etc. This process terminates after a finite number of steps, when a δ_j is obtained such that $\alpha_{i+1} - \delta_j$ bounds a subset of S.

At each step of the construction of the sequence $c_0, \alpha_1, ...\alpha_j, c_1$, it is possible to decrease the intersection number further by adding handles to α_i corresponding to horizontal arcs of $c_1 \cap (S \setminus \alpha_i)$ to the left of α_i as well as to the right, however, the resulting multicurve $\alpha_{i+1}^{'}$ will often intersect α_i. The multicurve $\alpha_{i+1}^{'} - \alpha_i$ will not bound a subset of S, because the handles are not all to the left of $-\alpha_i$.

A locally constant function f can be defined on $S \setminus (c_0 \cup c_1)$ as follows: suppose x and y are points on S not on either of the curves c_0 or c_1. Then let $f(x) - f(y)$ equal the homological intersection number of an oriented arc a_{yx} going from y to x with the oriented multicurve $c_1 - c_0$. Recall that the homological intersection number is defined such that if a_{yx} crosses $c_1 - c_0$ from right to left, this is counted as $+1$ and vice versa. This definition is independent of the choice of a_{yx} because the homological intersection number of any closed curve with $c_1 - c_0$ is 0 since $c_1 - c_0$ is null homologous. f is then defined by letting its minimum value equal 0. If different representatives of the free homotopy classes of c_0 and c_1 are chosen, the function obtained will be different. For the moment, it is enough to assume that the curves c_0 and c_1 have the least possible number of intersections, since

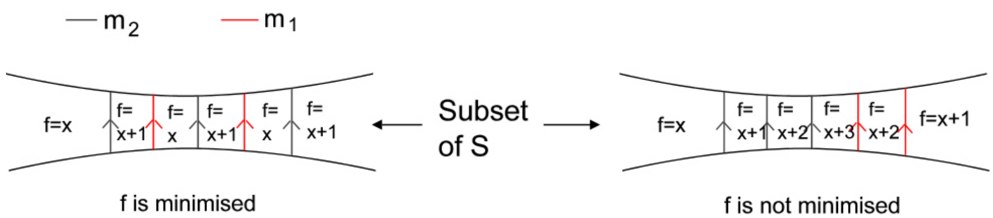

Figure 2.5: Calculating f for general multicurves

these functions will be used to obtain a surface producing sequence, and all the different functions obtained under these conditions will give the same arc in the homology multicurve graph. If the curves c_0 and c_1 are replaced by multicurves m_1 and m_2, the function f can be defined in the same way, but with the additional assumption that whenever $m_1 - m_2$ contains freely homotopic curves, the representatives of any free homotopy class are chosen to minimise f, as shown in figure 2.5. This is equivalent to the assumption that the freely homotopic curves are embedded in S in such a way that they bound a subset of S wherever possible. Without this assumption, f might have a maximum inside an annulus A that can be removed by choosing different representatives of the homotopy classes of m_1 and m_2. Whenever the maximum inside A is the only component of the maximum of f, attaching a handle to m_1 corresponding to f_{max} will only change m_1 up to homotopy, and the algorithm won't give the shortest surface producing sequence.

Definition 23 (The function obtained from $m_1 - m_2$)
Let m_1 and m_2 be homologous multicurves. The function f obtained from $m_1 - m_2$ is the locally constant function defined on $S \setminus (m_1 \cup m_2)$ with minimum value zero and such that, for any two points x and y in $S \setminus (m_1 \cup m_2)$, $f(x) - f(y)$ is the homology intersection number of $m_1 - m_2$ with an oriented arc with starting point y and endpoint x. If $m_1 - m_2$ contains any freely homotopic curves, it is assumed in addition that these freely homotopic curves are embedded in S in such a way as to minimise the maximum value of the function.

The function f can be thought of as a height function on $S \setminus (c_0 \cup c_1)$. Horizontal arcs of c_0 and c_1 are horizontal and vertical arcs of c_0 and c_1 are vertical with respect to this height function, as shown in figure 2.2.

Given c_0 and c_1, f is a bounded function on $S \setminus (c_0 \cup c_1)$ and has a

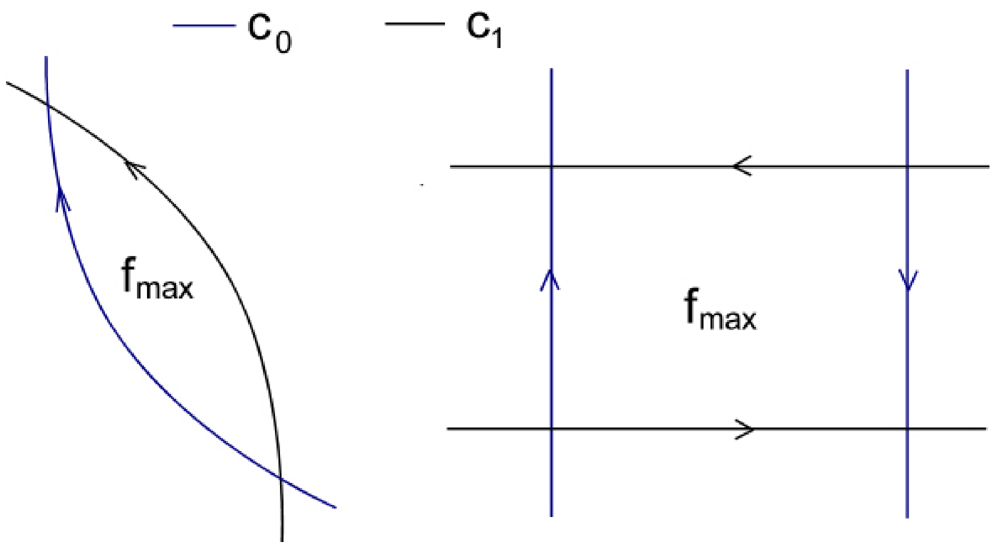

Figure 2.6: Examples of f_{max}.

maximum. Call the subset of S on which f takes on its maximum f_{max}. f_{max} has at least one connected component. The boundary of f_{max} consists of arcs of c_0 and c_1 such that f_{max} is to the right of any arc of c_0 on its boundary and to the left of any arc of c_1 on its boundary. In other words, the boundary of f_{max} is a null homologous multicurve made up of horizontal arcs of c_0 to the left of c_1 and horizontal arcs of c_1 to the right of c_0. This observation will be used to construct a shortest possible surface producing sequence $c_0, \gamma_1, ...\gamma_j, c_1$ with convenient properties.

Similarly, the subset of S, f_{min}, on which f takes on its minimal value is disjoint from f_{max} and is on the left of any arc of c_0 on its boundary and to the right of any arc of c_1 on its boundary.

Recall that the boundary of f_{max} is oriented in such a way that f_{max} is on its left, and let $a_1, a_2...$ be the arcs of c_1 on ∂f_{max}, $b_1, b_2, ...$ be the arcs of c_0 on ∂f_{max}. Then $\partial f_{max} = \sum_i a_i - \sum_j b_j$ (arcs are chains, and so they can be added and subtracted). Up to free homotopy on the boundary, f_{max} can be thought of as "that piece of S that is bounded by c_0 and γ_1". To make this more precise, consider the one dimensional CW complex $c_0 \cup c_1$ on S. Subtract the oriented arcs b_i from the oriented subcomplex c_0 and add the oriented arcs a_j. If this multicurve contains freely homotopic curves c and $-c$, cancel them out. This defines γ_1. Subtracting the arcs b_i from c_0,

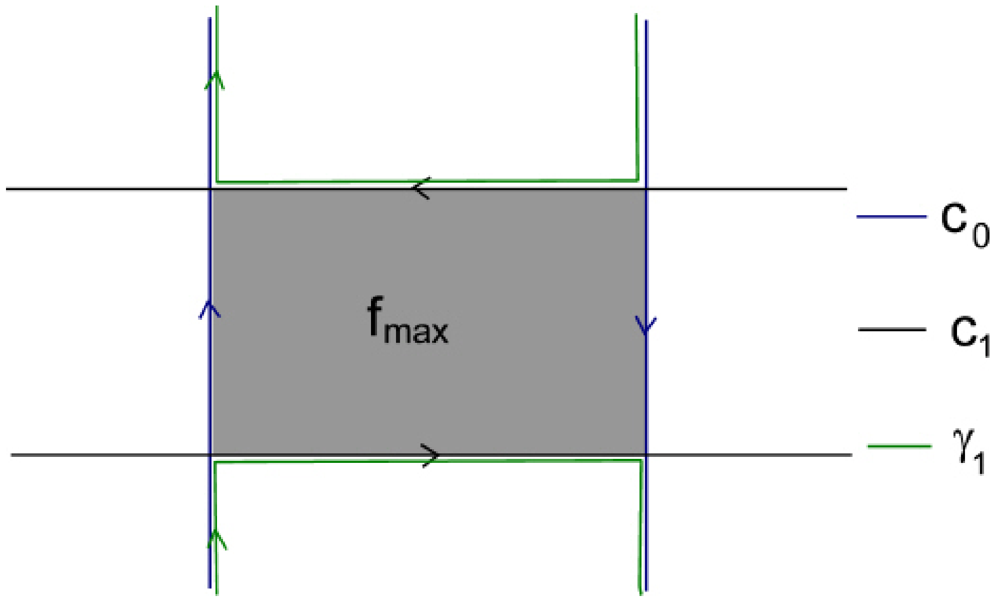

Figure 2.7: How to construct γ_1.

adding the arcs a_j and discarding the null homologous multicurve $-\partial f_{max}$ will be called "adding a handle" or "adding handles" corresponding to f_{max}, depending on the number of connected components of f_{max}.

∂f_{max} is disjoint from c_0 and each connected component of f_{max} intersects an annular neighbourhood of c_0 on the right side of c_0 (i.e. every component of f_{max} is "on the same side" of c_0). Therefore $i(\gamma_1, c_0) = 0$. The choice to use f_{max} instead of f_{min} was arbitrary, but it is not possible to simultaneously reduce the intersection number further by requiring that the subset of S bounded by γ_1 and c_0 be $f_{max} \cup f_{min}$ because f_{min} is to the left of c_0 and f_{max} is to the right of c_0, so this would not give a surface producing sequence.

The decision to cancel out freely homotopic curves with opposite orientation in γ_1 is arbitrary. If γ_1 contains some other null homologous submulticurve n, this could have been cancelled out also, however if the subset of S bounded by n is not disjoint from f_{max}, $c_0 - (\gamma_1 \setminus n)$ will not bound a subset of S, and so an extra multicurve would be needed in between c_0 and γ_1.

The multicurve γ_2 is constructed in the same way as γ_1 only with the curve c_0 replaced by γ_1. It is not difficult to see that the function f_1 obtained from γ_1 and c_1 has maximum one less than the maximum of f. Cutting out the arcs b_i make it possible to connect the subset of S, f_{1min}, on which f_1 takes

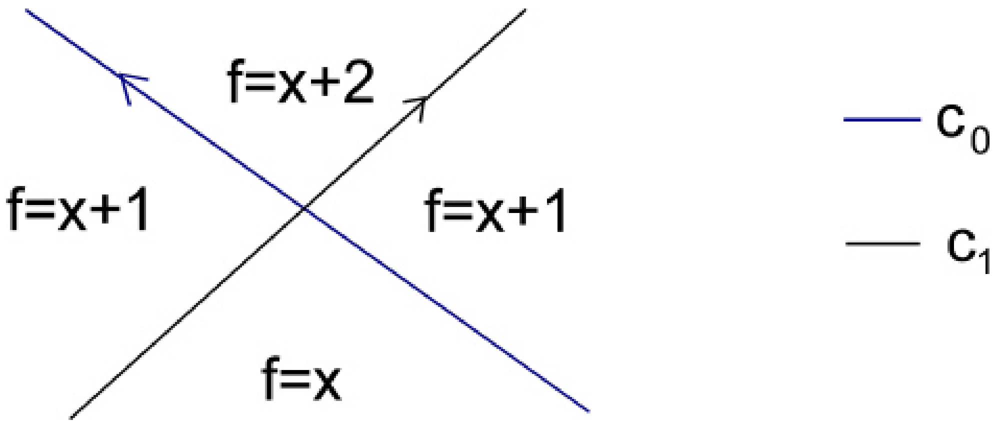

Figure 2.8: A point of intersection forces f to be at least two.

on its minimum, to f_{1max} (defined similarly), by an arc that crosses $c_1 - \gamma_1$ from right to left once less than any arc connecting f_{min} with f_{max}.

This process ends with the multicurve γ_j when the function f_j obtained from γ_j and c_1 has maximum value 1. The maximum of f_j can only be one if γ_j and c_1 don't intersect, because as shown in figure 2.8, an intersection forces the maximum of f_j to be at least two.

If the maximum of f_j is one, then the subset of S on which $f_j = 1$ is the subset bounded by $c_1 - \gamma_j$. This sequence of multicurves, $c_0, \gamma_1, ...\gamma_j, c_1$, is surface producing, so it is possible to construct a surface H_f with boundary $c_1 - c_0$ as described. The number of multicurves j in the sequence is equal to one less than the maximum of f.

With this algorithm for constructing surface producing sequences, it is convenient to work with multicurves in S that are not in general position. In this context, it is convenient to define a "point of intersection" as follows.

Definition 24 (Point of intersection for curves not in general position) The left and right side of an oriented curve in S has been defined. Suppose a is a multicurve in S and b is a second multicurve in S such that a and b are not in general position. If a and b coincide along some subarc or point, this subarc or point will be counted as a single point of intersection iff b crosses from one side of a to the other.

At each step of the algorithm, the intersection number with c_1 is decreased. Recall that the arcs of $c_1 \cap (S \setminus c_0)$ on ∂f_{max} were denoted $a_1...a_n$.

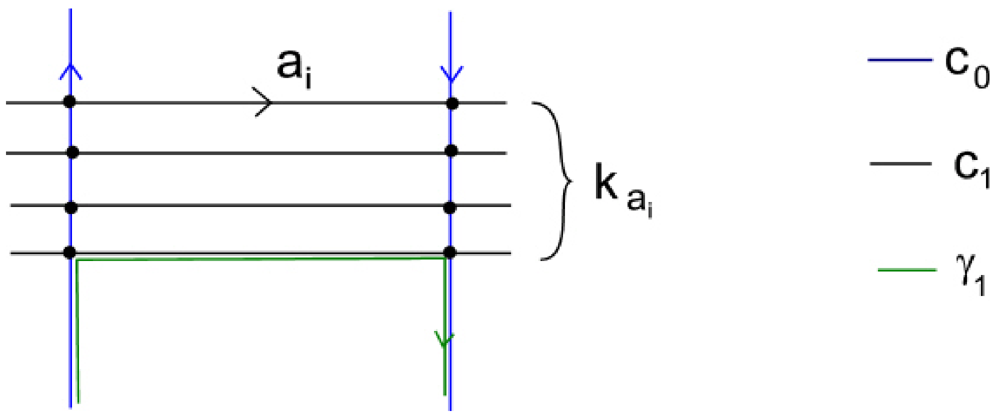

Figure 2.9: Calculating the reduction in intersection number.

Let k_{a_i} be the number of arcs of $c_1 \cap (S \setminus c_0)$ in the same homotopy class as a_i for $1 \leq i \leq n$. Then, as shown in figure 2.9, the intersection number of γ_1 with c_1 is at least $2\sum_i k_{a_i}$ less than the intersection number of c_0 with c_1.

Lemma 25
Suppose $[m_1]$ and $[m_2]$ are homologous multicurves and that m_1 and m_2 only have essential points of intersection. Let m_3 be obtained from m_1 by adding a handle corresponding to a horizontal arc a_1 of $m_2 \cap (S \setminus m_1)$. If k_{a_1} is equal to the number of arcs in the homotopy class a_1, then $i(m_1, m_2) = i(m_3, m_2) + 2k_{a_1}$.

Proof. By the way "points of intersection" were defined in definition 24, it is clear that a homotopy can only remove points of intersection pairwise, otherwise the homotopy would change the homology intersection number, which is impossible. It is also clear that the intersection number can be decreased by as much as $2k_{a_1}$ by adding the handle corresponding to a_1 to m_1; this is a consequence of the definition of homotopy class. It remains to show that there is no homotopy that decreases the intersection number by more than $2k_{a_1}$. Let R_{a_1} be the rectangle in S consisting of the closure of the union of rectangles in $S \setminus (m_1 \cup m_2)$, each of which have two opposite sides made up of arcs of $m_2 \cap (S \setminus m_1)$ in the homotopy class a_1. Let $m_3^{'}$ be the representative of the free homotopy class $[m_3]$ that coincides with the subcomplex m_1 of the CW complex $m_1 \cup m_2$ outside of R_{a_1} and with arcs of m_2 in the homotopy class a_1 on the boundary of R_{a_1}. Suppose also that

m_3' does not enter the interior of R_{a_1}. According to the definition 24, m_3' has intersection number $i(m_1, m_2) - 2k_{a_1}$ with m_2. Suppose that the intersection number of m_3' with m_2 could be further decreased. For this to happen, there has to be an arc of m_3' with endpoints p_1 and p_2 on m_2 that is homotopic with fixed endpoints to an arc of m_2. Since m_1 and m_2 only have essential points of intersection, this arc could not be an arc disjoint from the closure of R_{a_1}, so p_1 and p_2 have to be the endpoints of an arc of the form $l_1 \circ a_1 \circ l_2$, where l_1 and l_2 are arcs of $m_1 \cap (S \setminus m_2)$. However, this is a contradiction, because $l_1 \circ a_1 \circ l_2$ is homotopic to a_1 relative to m_1. □

Corollary 26
Whenever $a_1, a_2, ...a_n$ are homotopy classes of horizontal arcs of $c_1 \cap (S \setminus \gamma_k)$, $i(\gamma_k, c_1) \geq i(\gamma_{k+1}, c_1) + 2\sum_i k_{a_i}$.

Given a surface producing sequence $c_0, \gamma_1, ...\gamma_j, c_1$ and a choice of projection function π, in the discussion after the construction of Δ_{j+1} it was shown that it is possible to construct H_γ, where H_γ is a surface with boundary $c_1 - c_0$ in M. H_γ is constructed in a particular way that will be made use of in the next definition. In particular, H_γ is a finite union of subsets of $S \times i$, for $i = 1, 2, 3...j + 1$, each of which project one to one onto a subset of S, with a union of annuli, each of which projects onto a simple curve in S. ∂H_γ has boundary consisting of two simple curves c_0 and c_1. Since c_0 is on the boundary of a component of $S \times 1$ that projects one to one onto a subset of S and c_1 is on the boundary of a subset of $S \times j + 1$ that projects one to one onto a subset of S, each component of ∂H also projects one to one into S. Therefore, whenever H_γ is embedded, it is possible to define a function f_γ (or alternatively f_H) on $S \setminus \pi(c_0 \cup c_1)$ as follows:

$$f_\gamma : s \in S \rightarrow \text{ number of connected components of } \pi^{-1}(s) \cap H_\gamma \qquad (2.1)$$

From the way H_γ was constructed, it is clear that $f_\gamma \leq j + 1$, where $j+2$ is equal to the number of multicurves in the surface producing sequence $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$, including c_0 and c_1. If H_γ is not embedded, f_γ is defined in the same way as in the previous equation, only any point of $\pi^{-1}(s) \cap H_\gamma$ that is a point of self-intersection of H_γ is counted twice, or three times if it is a triple point.

Lemma 27
Given any two surface producing sequences $\{c_0, \beta_1, ...\beta_j, c_1\}$ and $\{c_0, \delta_1, ...\delta_k, c_1\}$, f_β and f_δ differ by at most a universal additive constant.

Proof. For any surface producing sequence $\{\beta_i\}$, f_β has to have the property that it increases by one when crossing over an arc of $c_1 - c_0$ from right to left and decreases when crossing from left to right. Suppose also that f_β changes when crossing over an oriented curve α in S that is not a subset of $c_1 - c_0$. Given a surface producing sequence $\{c_0, \beta_1, ...\beta_j, c_1\}$, recall that the surface H_β was constructed by attaching surfaces Γ_i in M, where Γ_i projects one to one onto the subset of S with boundary $\beta_i - \beta_{i-1}$. In addition, each Γ_i is oriented as a subset of S. Choose an arc α' of $\alpha \cap (S \setminus c_0 \cup c_1)$. If a given component of $\pi^{-1}(\alpha') \cap H_\beta$ is contained in Γ_i for some i, then this component, α_1 of $\pi^{-1}(\alpha') \cap H_\beta$ has a neighbourhood in Γ_i diffeomorphic to $\alpha_1 \times (-\epsilon, \epsilon)$. Since α' is oriented so is α_1. In this neighbourhood it therefore makes sense to talk about the left and right side of α_1. Since it is contained in Γ_i, this neighbourhood is projected one to one onto S. Alternatively, α_1 could be a subarc of $\partial\Gamma_i$ for some i. Each of the Γ_i are oriented in the same way as subsets of S, so each oriented curve c in $\partial\Gamma_i$ has a neighbourhood $N(c)$ in H_β such that $N(c) \setminus c$ has two components - one to the left and one to the right of c - and such that $\pi(N(c))$ is also of this form. π does not identify a point of $N(c)$ to the right of c with a point of $N(c)$ to the left of c. Similarly if the component α_1 of $\pi^{-1}(\alpha') \cap H_\beta$ intersects $\partial\Gamma_i$ for some i. Therefore if p and q are two points in S contained in the intersection of the projection of the neighbourhoods of each component of $\pi^{-1}(\alpha') \cap H_\beta$, then $f(p) = f(q)$, which contradicts the assumption that f changes when crossing over α. If f could change by more than one when crossing over an arc of $c_1 - c_0$ then the same argument applied to every component of $\pi^{-1}(\alpha') \cap H_\beta$ that is not on ∂H_β would also give a contradiction.

□

Lemma 28
Given the surface H_γ, the function f obtained from the projections of the boundary curves $c_1 - c_0$ of H_γ coincides with the function f_γ, where γ here is the arc in the homology multicurve graph constructed by successively adding handles corresponding to the extrema of the functions $f, f_1, ...$

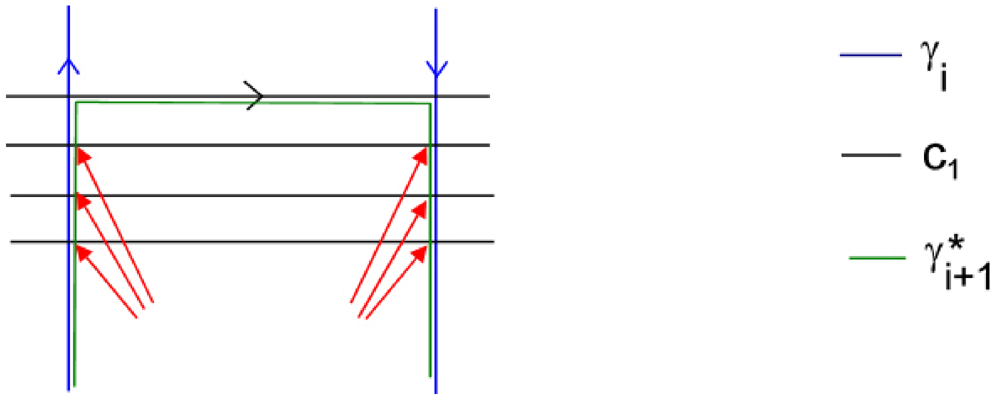

Figure 2.10: Points of intersection that could be removed by a homotopy.

Proof. The reason this is not immediately clear is that the multicurve γ_{i+1}, obtained from γ_i and c_1 by attaching a handle corresponding to f_{imax}, might have points of intersection with c_1 that can be removed by a homotopy for any i. The function f_{i+1} depends on the representative of the free homotopy class $[\gamma_{i+1}]$. In order to define the maximum value of f_{i+1}, it was assumed that the multicurves used to define the function have the smallest possible number of points of intersection.

Let R_{a_i} be the rectangle in S consisting of the closure of the union of rectangles in $S\backslash(c_1 \cup c_2)$, each of which have two opposite sides made up of arcs of $c_2 \cap (S\backslash c_1)$ in the homotopy class a_i, where each of the $a_1 ... a_n$ are homotopy classes of arcs with representatives on ∂f_{max}. Let $R := R_{a_1} \cup R_{a_2} \cup ... R_{a_n}$, and $\gamma_1^{'}$ be the multicurve homotopic to γ_1 constructed similarly to $m_3^{'}$ in lemma 25 i.e. $\gamma_1^{'}$ coincides with c_0 outside of R and is a representative of the homotopy class with the smallest possible number of points of intersection with c_1, according to definition 24. $\gamma_1^{'} - c_0$ therefore bounds the subset $f_{max} \cup R$ of S.

Let f_1 be the function obtained from γ_1 and c_1, and let $f_1^{'}$ be the function obtained from $\gamma_1^{'}$ and c_1. γ_1 and $-\gamma_1^{'}$ bound the subset R of S. For a point $s \in S$ for which both f_1 and $f_1^{'}$ are defined,

$$f_1(s) = \begin{cases} f_1^{'}(s) + 1 & \text{if } s \in R, \\ f_1^{'}(s) & \text{otherwise.} \end{cases} \tag{2.2}$$

In other words, the homotopy that takes γ_1 to $\gamma_1^{'}$ reduces the function by one on the subset R and enlarges the subset of S bounded by $\gamma_1 - c_0$ to obtain

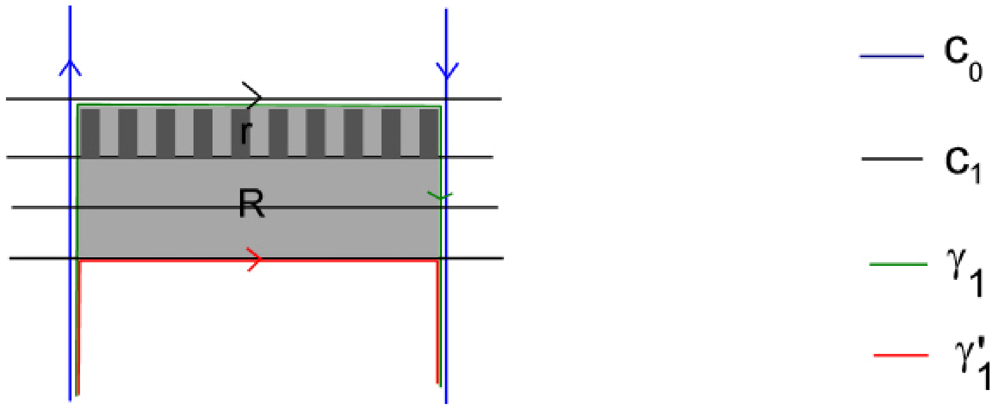

Figure 2.11: γ_1, γ_1', R and r.

the subset of S bounded by $\gamma_1' - c_0$, $f_{max} \cup R$. Any components of $S \setminus (c_0 \cup c_1)$ with one edge along f_{max} are contained in f_{1max}, and since these components aren't all contained in R, it follows that f_1 has the same maximum value as f_1'. f_{1max} is the union of f_{1max}' with a union r of rectangles of $S \setminus (c_1 \cup c_0)$ in R, as shown in figure 2.11. Attaching handles to γ_1 corresponding to rectangles in r reduces the number of points of intersection with c_1, and homotoping γ_1 to γ_1' has the same effect as attaching a handle to γ_1 corresponding to each rectangle in R and discarding contractible curves. If r is not the whole of R, when passing from γ_2 to γ_3, handles corresponding to further rectangles in R are attached. This is continued until for large enough i, f_{imax} contains all of R and γ_{i+1} has no points of intersection with c_1 on ∂R. If γ_1 is used in place of γ_1' to construct γ_2, the same multicurve will therefore be obtained up to free homotopy, despite the fact that γ_1 might have nonessential points of intersection with c_1. The same argument applies for all γ_i in place of γ_1, from which the lemma follows.

□

Lemma 29
The shortest surface producing sequence from c_0 to c_1 consists of j multicurves (not counting c_0 and c_1), where j is equal to one less than the maximum value of f.

Proof. The function f was defined for homologous multicurves m_1 and m_2, where it is assumed that m_1 and m_2 only have essential points of intersection and that if m_1 and m_2 contain freely homotopic curves, these curves are embedded in S in such a way that they bound a subset of S whenever possible. These assumptions are necessary here, because the properties that a sequence has to fulfil in order to be surface producing are independent of free homotopy of the multicurves in the sequence, but the maximum of the function is not. The assumptions on the multicurves involved ensure that the maximum of f is as small as possible.

The previous arguments have shown that every surface producing sequence defines a function on $S \setminus (c_0 \cup c_1)$. The maximum of the function is equal to the maximum number m of connected components of $\pi^{-1}(s) \cap H_\gamma$, for $s \in S$. The number of multicurves in the sequence $\{\gamma_i\}$ can't be less than $m-1$, because if it were, H_γ could have been constructed by connecting up $m-1$ or fewer pieces, each of which projects one to one onto S, which would contradict the fact that $\pi^{-1}(s) \cap H_\gamma$ has m connected components for some $s \in S \setminus (c_0 \cup c_1)$. The previous arguments have also shown that this minimum number of multicurves can always be achieved. For every sequence $\{\gamma_i\}$, the function f_γ is everywhere positive or zero. It follows from lemma 27 that f has the smallest possible maximum because its minimum is zero.

□

The function f_H might have been defined in a more general way, by taking orientation into account. If the surface H_γ had been constructed by attaching surfaces homotopic to subsets of S, where these surfaces are not all oriented as subsets of S, f_H could have been defined as

$$f_H : s \in S \setminus \pi(\partial H_f) \to x - y, \text{ where}$$

x := the number of connected components of the set $\pi^{-1}(s) \cap H_f$ for which $\pi^{-1}(s)$ intersects the surface H_f in a connected set contained in a neighbourhood in H_f that projects onto an open subset of S with the orientation induced by S, and

y := the number of connected components of the set $\pi^{-1}(s) \cap H_f$ for which $\pi^{-1}(s)$ intersects the surface H_f in a connected set contained in a neighbourhood in H_f that projects onto an open subset of S with the opposite orientation to that induced by S.

A connected component of $\pi^{-1}(s) \cap H_f$ that does not have a neighbourhood in H_f that projects onto an open set of S is not counted.

With this definition, it is no longer necessary to require that the surfaces homotopic to subsets of S used to construct H_f are all oriented in the same way, however, for surfaces constructed from surface producing sequences, this doesn't really provide any new information. Instead of using a surface Γ_i homotopic to f_{imax} in the construction of H, a surface homotopic to $S \setminus f_{imax}$ with the opposite orientation could be used instead. The surface with boundary curves $c_1 - c_0$ constructed in this way would give rise to a new function everywhere equal to $f_H - 1$. The difference between the maximum and minimum values, which is what determines the number of multicurves needed, remains unchanged. The advantage of definition 16 over the original definition is that it gives reasonable results for surfaces that are not constructed from surface producing sequences. For example, suppose $c_1 - c_0$ is a multicurve and H is any orientable surface in M with boundary $c_1 - c_0$. The same argument as in lemma 27 shows that f_H, as defined in definition 16, has to be constant on any component of $S \setminus (c_1 - c_0)$. From this it follows that if H has smallest possible genus, it has to be homotopic to a subset of S.

Corollary 30 (Corollary of lemma 27)
Suppose that $c_1 - c_0$ is a multicurve, where c_0 and c_1 are simple, homologous curves in S. Then any orientable surface in M with smallest possible genus whose boundary is freely homotopic to $c_1 - c_0$ has to be homotopic to a subset of S.

A notational complication that has become apparent in the previous two lemmas is due to the fact that, when talking about a sequence in the homology multicurve graph, it is only of interest to know the multicurve γ_i up to free homotopy. To keep the notation as simple as possible, the same notation will sometimes be used when referring to a curve or the free homotopy class containing the curve. If it is necessary to choose the representative of the free homotopy class in a particular way, this will be explicitly stated. It will often be useful to choose the representatives of the homotopy classes to make the intuitive picture of f_{max} clearer. Let c_0 and c_1 be representatives of their homotopy classes that only have essential points of intersection, according to definition 24. Fix these representatives, and choose the representatives of the homotopy classes such that $\gamma_{i+1} - \gamma_i$ is the boundary of the subset of f on which f is no less than its maximum value minus i, i.e. f_{imax}. (c_0 is understood to be γ_0 and c_1 γ_{j+1}). That this choice makes sense, despite the fact that this choice of the representatives of the homotopy classes $[\gamma_i]$ might have nonessential points of intersection with c_1, has been discussed in lemma 28. The boundary of f_{imax} is an embedded subcomplex of the one dimensional CW complex $c_0 \cup c_1$ for every i, and has zero intersection number with γ_i and γ_{i+1}. γ_{i+1} is obtained from γ_i by subtracting the arcs of $\gamma_i \cap (S \setminus c_1)$ on ∂f_{imax} and adding the arcs of $c_1 \cap (S \setminus \gamma_i)$ on ∂f_{imax}. Also, no arc of $c_1 \cap c_0$ will be on the boundary of f_{imax} for more than one value of i, so each arc can only be added or subtracted at most once. Each of the multicurves γ_i is therefore a subcomplex of $c_0 \cup c_1$, and is oriented as a subset of $c_0 \cup c_1$. From figure 2.8, it is easy to verify that f_{imax} can not meet itself at a vertex, because if four components of $S \setminus (\gamma_i \cup \gamma_k)$ come together at a point and the function is equal on two of them, it must be larger on a third component and smaller on the fourth. Therefore, if γ_i doesn't meet or cross over itself at a vertex, neither will γ_{i+1}. It follows that the γ_i chosen in this way are embedded, oriented subcomplexes of $c_0 \cup c_1$. The main advantage of doing this is that the functions $f, f_1, f_2...$ are related in an obvious way. The disadvantage of this choice is that, as already mentioned, these choices of representatives of the multicurves γ_i might have nonessential points of intersection with c_1, and that as subcomplexes, these representatives aren't all pairwise in general position. Confusion can arise because the homotopy class of a boundary does not determine the topology of the surface that it bounds. For example, if f_{imax} is a rectangle, as shown in figure 2.12. The representative of the homotopy class of $[\gamma_{i+1}]$ can also be chosen such that $\gamma_{i+1} - \gamma_i$ bounds a pair of pants.

Lemma 31
If c_0 and c_1 had been interchanged in the algorithm for constructing the surface producing sequence $c_0, \gamma_1, \gamma_2 ... \gamma_{j-1}, \gamma_j, c_1$, the sequence $c_1, \gamma_j, \gamma_{j-1}, ...\gamma_2, \gamma_1, c_0$ would have been obtained.

Proof. Suppose the representatives of the multicurves $c_0, \gamma_1, ...\gamma_j, c_1$ are chosen as outlined in the previous paragraph, in particular, each of the γ_i are oriented subcomplexes of the CW complex $c_0 \cup c_1$ such that $\gamma_{i+1} - \gamma_i$ is the boundary of the subset of f on which f is no less than its maximum value minus i. Let h be the function on $S \setminus (c_0 \cup c_1)$ obtained from $c_0 - c_1$. It is easy to check that h has its maximum where the function f obtained from $c_1 - c_0$ has its minimum, and vice versa. By definition, γ_j is the multicurve chosen such that $c_1 - \gamma_j$ bounds the subset of S given by $S \setminus f_{min}$. In other

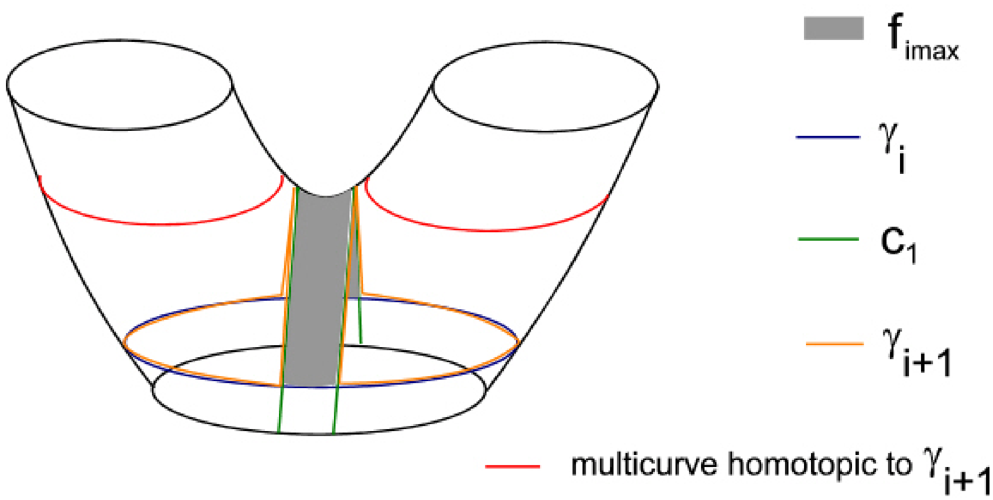

Figure 2.12: The topology of the subset of S bounded by $\gamma_{i+1} - \gamma_i$ depends on the choice of representatives of the free homotopy classes.

words, $\gamma_j - c_1$ is the boundary of f_{min} or h_{max}, i.e. γ_j satisfies the definition of the first multicurve in the sequence $c_1, ...c_0$. Similarly for γ_{j-1}, γ_{j-2}, etc.

□

The genus of the subset Γ_i of S is bounded from above by the genus g of S, so lemma 29 shows that it is always possible to construct a surface with boundary $c_1 - c_0$ with genus less than or equal to gm. If the subsets of S being connected up to form the surface are not all required to be oriented in the same way, at each step it is possible to choose between Γ_i and $S \setminus \Gamma_i$, so a bound of $\frac{gm}{2}$ is obtained.

2.2 Minimal Genus Surfaces

In this section it is convenient to work in the smooth category. All maps will also be assumed to be smooth. Let c_0 and c_1 be simple, homologous curves as defined in the previous section.

It is shown that every surface in M with boundary curves $c_1 - c_0$ with smallest possible genus can be constructed by the algorithm outlined in section 2.1. In particular, the following theorem will be proven:

Theorem 32
Recall the definitions of surface and homotopy of surfaces given in definitions 8 and 9. In particular, suppose H is a smooth, oriented surface locally embedded in M with boundary curves $c_1 - c_0$ and smallest possible genus. Then there exists a surface producing sequence γ such that H is homotopic to H_γ.

M is given a product metric $ds_M^2 = ds_S^2 + dR^2$ where ds_S is a choice of metric on $S \times 0$, and R is, as usual, the coordinate obtained by projecting onto the second component of $S \times R$. Similarly, H and all surfaces in M homotopic to H are assumed to be covered by coordinate charts $(U_1, s_1, R), ..., (U_k, s_k, R)$, where the s_i are coordinates obtained by projecting onto $S \times 0$. Theorem 32 is proven by treating the restriction to H of the R coordinate as a Morse function, which is shown to be possible in lemma 38. The reason for requiring this function to be Morse is that, when H is embedded, the level sets of R have controlled intersection properties, which will be used for constructing a surface producing sequence.

The fact that H has nonempty boundary makes it necessary to give a definition of critical point that could include boundary points of H.

Definition 33 (Critical Point of R)
In the interior of H, a critical point of the restriction to H of the R coordinate of M is a point x of H at which the derivative of the restriction of R to H vanishes. The boundary of H is a union of embedded, one dimensional submanifolds of M, so a critical point on the boundary of H is a point where the restriction of H to the boundary has zero derivative.

Definition 34 (Degenerate Critical Point)
A degenerate critical point on the interior of the surface is a critical point at

which the Hessian matrix has zero determinant. A degenerate critical point on the boundary is any critical point that is not an isolated local extremum.

Definition 35 (Morse function)
A Morse function on a surface H is a $\mathcal{C}^\infty$ function from H into $\Re$ for which all critical points are nondegenerate.

Definition 36 (Hausdorff Topology (From [20]))
Given the topology on M induced by the product metric, a sequence of closed sets $\{\Lambda_n\}$ in M is said to converge to Λ in the Hausdorff topology if the following two conditions are satisfied:

1. Any accumulation point of a sequence $\{x_n \in \Lambda_n\}$ belongs to Λ
2. Every $x \in \Lambda$ is the limit point of a sequence $\{x_n \in \Lambda_n\}$

Definition 37 ($\mathcal{C}^2$ topology)
Fix a set of coordinate charts, $(U_1, s_1, R), ..., (U_k, s_k, R)$ on H, where the s_i are coordinates obtained by projecting into $S \times 0$. Let $\mathcal{C}^2(H)$ be the set of all $\mathcal{C}^2$ maps of H into the real line. The $\mathcal{C}^2$ topology on $\mathcal{C}^2(H)$ is the topology with the neighbourhood basis given by sets of the form $\mathcal{N}^2(f, \epsilon)$, where $\mathcal{N}^2(f, \epsilon)$ consists of all functions g in $\mathcal{C}^2(H)$ such that, within every coordinate chart

$$|f(x) - g(x)| < \epsilon \text{ , } |Df(x) - Dg(x)| < \epsilon \text{ and } \left|D^2 f(x) - D^2 g(x)\right| < \epsilon$$

for all $x \in H$.

Lemma 38
Suppose H is a compact embedded surface in M with boundary curves c_1 and c_0. Then there is an embedded surface in M, call it H', with the following properties:

1. H' is homotopic to H according to definition 9
2. The restriction of the R coordinate to H' is a Morse function
3. No two critical points of the Morse function from 2 have the same value of the R coordinate.

Proof. It is a standard result, e.g. [21] Theorem 2.7, that on a compact manifold without boundary, the Morse functions form an open, dense (in the $\mathcal{C}^2$ topology) subset of the set of all smooth functions of the manifold into $\Re$. This and similar standard results in Morse theory are proven by altering a given function by adding arbitrarily small functions with small derivatives. Similar arguments are used here; the main difference is that the function R is treated as fixed while the subset of M to which R is restricted is altered by a homotopy. To start off with, the existence of an embedded surface H^1 homotopic to H to which the restriction of R to a neighbourhood of the boundary is a Morse function will be shown. The standard Morse theory arguments that assume empty boundary will then be shown to apply to H^1.

Let N be a collar of the boundary of H; the existence of which is guaranteed by theorem 6.1, chapter 4 of [16]. ∂H is a compact manifold without boundary, so by theorem 2.7 of [21], if the restriction of R to ∂H is not a Morse function, there is a Morse function R_m on ∂H arbitrarily close to R in the $\mathcal{C}^2$ topology. N is diffeomorphic to two copies of $S^1 \times [0, \iota]$, which defines coordinates (t, r) on each component of N, where t is the parameter on S^1 and r is defined on the interval $[o, \iota]$ and is equal to zero on the boundary curves c_0 and c_1. Let $\phi(t, r)$ be a smooth function on N, $0 \leq \phi \leq 1$, $\phi\,|_{\partial H} = 1$, and let $\eta(t)$ be the function $R_m(t) - R$ on ∂H. The function $R + \phi(t, r)\eta(t)$ is therefore a Morse function when restricted to ∂H. To construct a function without degenerate critical points on a neighbourhood of the boundary, it is enough to show that $\phi(t, r)$ can be chosen such that $\frac{d(R+\phi(t,r)\eta(t))}{dr}$ and $\frac{d(R+\phi(t,r)\eta(t))}{dt}$ are not simultaneously zero on a neighbourhood N_1 of ∂H contained in N. As a consequence of smoothness, $\frac{d(R+\phi(t,r)\eta(t))}{dt}\Big|_{r=\kappa} - \frac{d(R+\phi(t,r)\eta(t))}{dt}\,|_{r=0}$ can be made arbitrarily small by choosing κ sufficiently small. Since $R + \phi(t, r)\eta(t)$ is a Morse function on ∂H, when restricted to ∂H, $\frac{d(R+\phi(t,r)\eta(t))}{dt}$ is only zero at (isolated) critical points $p_1 = (t_1, 0)$, $p_2 = (t_2, 0)...p_n = (t_n, 0)$. Therefore, $N_1 \subset N$ can be chosen such that in N_1, $\frac{d(R+\phi(r,t)\eta(t))}{dt}$ can only pass through zero in a neighbourhood of the form $P_i := (p_i - \epsilon, p_i + \epsilon) \times (0, \epsilon)$, for $i = 1, 2, ...n$. Inside each of the P_i, ϕ can be chosen such that $\frac{d(R+\phi(r,t)\eta(t))}{dr}$ is nonzero. This is possible because ϵ can be chosen such that R, η and their derivatives do not vary much in the ϵ neighbourhoods. It follows that N_1 and ϕ can be chosen such that $R + \phi(r, t)\eta(t)$ is a Morse function on N_1.

Let H^1 be the subset of M that coincides with H outside of N and is given by the graph $(s, R + \phi(r, t)\eta(t))$ in any coordinate chart (U_i, s_i, R) over N. Since H is smoothly embedded in M as a submanifold with boundary, it follows from theorems 6.1 and 6.3 of [16] that H has an embedded neighbourhood $\mathcal{E}(H)$ in M. As R_m approaches R in the $\mathcal{C}^2$ topology on ∂H, $R+\phi(r,t)\eta(t)$ also approaches R in the $\mathcal{C}^2$ topology on N. If R_m was chosen to be sufficiently close to R in the $\mathcal{C}^2$ topology, it follows that H^1 is contained in $\mathcal{E}(H)$ and is also embedded. By construction, the restriction of the R coordinate to H^1 is a Morse function on a neighbourhood of the boundary.

Lemma 39 (Lemma B of [21])
Let K be a compact subset of an open set U in $\Re^3$. If $g : U \to \Re$ is smooth and has only nondegenerate critical points in K, then there is a number $\delta > 0$ such that if $h : U \to \Re$ is smooth and at all points of K satisfies

$$(1) \left| \frac{\partial g}{\partial x_i} - \frac{\partial h}{\partial x_i} \right| < \delta, \ (2) \left| \frac{\partial^2 g}{\partial x_i x_j} - \frac{\partial^2 h}{\partial x_i x_j} \right| < \delta$$

for $i, j = 1, ..., n$, then h also only has nondegenerate critical points in K.

Let $\mathcal{F}$ be the set of all smooth functions f from H^1 into $\Re$ such that $f\,|_{N_1} = R\,|_{N_1}$. By the definition of $\mathcal{F}$ and H^1, no element of $\mathcal{F}$ will have degenerate critical points in N_1. Recall that on the interior of H^1, degenerate critical points of f are points at which both the gradient of f and the determinant of the Hessian matrix are zero. Both the gradient of f and the determinant of the Hessian matrix are continuous quantities in the $\mathcal{C}^2$ topology, from which it follows that the set of Morse functions is open in $\mathcal{F}$.

To show denseness, let $(U_1', s_1, R), ..., (U_k', s_k, R)$ be a finite covering of H^1 by coordinate neighbourhoods $\{U_i'\}$ with coordinates (s_i, R), where s_i and R are coordinates obtained by projecting U_i' onto $S \times 0$ and R respectively. Suppose also that $(U_1', s_1, R), ..., (U_k', s_k, R)$ are chosen such that $(U_1, s_1, R), ..., (U_k, s_k, R)$ is a finite covering of $H^1 \setminus N_1$, where $U_i := U_i' \cap (H^1 \setminus N_1)$. It is possible to find compact sets $C_i \subset U_i$ such that $C_1, C_2, ...C_k$ cover $H^1 \setminus N_1$. Let $\mathcal{N}$ be a neighbourhood of a function f in $\mathcal{F}$. Degenerate critical points are removed in stages. Let η be a smooth function from H^1 into $[0, 1]$ such that $0 < \eta$ in a neighbourhood of C_1 and $\eta = 0$ in a neighbourhood of $H^1 \setminus U_1$. The function $f_1 := f + \epsilon\eta : H^1 \to \Re$ belongs to $\mathcal{F}$. By lemma A, page 11 of [21] (this is a corollary of Sard's Theorem) it follows that for almost all choices of $\epsilon\eta$, f_1 has no degenerate critical points on C_1. If ϵ is chosen small enough, f_1 will be contained in the neighbourhood $\mathcal{N}$ of f.

Now that a function f_1 has been obtained that does not have degenerate critical points in C_1, by Lemma B it is possible to find a neighbourhood $\mathcal{N}_1$ of f_1, where $\mathcal{N}_1 \subset \mathcal{N}$, so that any function in N_1 also has the property that is does not have degenerate critical points in N_1.

The next part of the proof involves repeating this process with f_1 and $\mathcal{N}_1$, to obtain a function f_2 in $\mathcal{N}_1$ that does not have degenerate critical points in C_2, and a neighbourhood $\mathcal{N}_2$ of f_2, $\mathcal{N}_2 \subset \mathcal{N}_1$, such that no function in $\mathcal{N}_2$ has degenerate critical points in C_2. It is automatically the case that f_2 does not contain degenerate critical points in C_1, since it is in $\mathcal{N}_1$. Finally, a function $f_k \in \mathcal{N}_k \subset \mathcal{N}_{k-1} \subset ... \subset \mathcal{N}_1 \subset \mathcal{N}$ is obtained, where f_k does not have degenerate critical points anywhere on $C_1 \cup ... \cup C_k := H^1 \setminus N_1$.

Now it is known that there is a Morse function, call it R_m, arbitrarily close to R on H^1 in the $\mathcal{C}^2$ topology, let $\eta(x) := R_m(x) - R(x)$. Let H^2 be the surface given by the graph $(s, R(s) + \eta(s))$ over each coordinate patch (U_i, s_i, R). As before, if η and its first derivatives are small enough, H^2 is embedded.

It is a direct consequence of the definitions that as η approaches zero in the C^2 topology, H^2 approaches H in the Hausdorff topology.

The proof that the surface can be chosen such that all critical points occur at different values of R is exactly the same as the standard result in the literature, for example, lemma 2.8 of [21], only once again, instead of altering the function by adding arbitrarily small correction functions to it, a correction function η is interpreted as a recipe for moving the surface H^2 up or down in the R direction by an amount determined by the value of η at that point, to obtain a surface $H^{'}$ homotopic to H^2. By the denseness result just proven, η can also be chosen to be close enough to zero in the $\mathcal{C}^2$ topology to ensure that $H^{'}$ is embedded.

□

The next lemma will be useful in the proof of theorem 32. Let $H_b^a := H \cap (S \times [b, a])$, $H_b := H \cap (S \times [b, \infty))$, $H^a := H \cap ((-\infty, a])$ and $H(a) := H \cap (S \times a)$.

Lemma 40

Suppose H is embedded. Then there is an embedded representative $H^{'}$ of the homotopy class of H such that $H_0^{'}$ does not have a component consisting of an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0. If a representative $H^{"}$ of the homotopy class of H is sufficiently close to $H^{'}$ in the Hausdorff topology,

$H_0^{"}$ does not have a component consisting of an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0, either.

Proof. If H_0 contains a component consisting of an annulus A with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0, there is a homotopy of H that takes the interior of A into a subset $A^{'}$ of $S \times [0, -\infty)$ and fixes $H \setminus A$. The image, H_1, of this homotopy might not be embedded, but can be assumed to be in general position. Since H_1 is in general position, as discussed in [14], its self-intersections can only consist of a union of curve, arcs with endpoints on the boundary and a discrete union of triple points. Since H is embedded, and the homotopy only moved annuli or certain types of bordered spheres, it follows that the self-intersections of H_1 consist of a union of:

1. curves homotopic to c_0
2. curves contractible in H_1
3. arcs with endpoints on the boundary of H_1 that are homotopic relative to their endpoints to arcs on ∂H_1
4. isolated triple points

The lemma will be proven by showing that all of the self-intersections of H_1 can be removed by a homotopy without creating new points of intersection with S.

If H_1 contains a self-intersection along a curve c_0 that is freely homotopic in H_1 to the boundary curve c_0, this self-intersection can be removed by cutting an annulus off the boundary of H_1 to obtain a new surface, homotopic to H_1, without this self-intersection. Similarly if H_1 intersects itself along an arc homotopic relative to its endpoints to a sub arc of the boundary of H_1. It might be the case that there is more than one free homotopy class of curves in H that is freely homotopic to c_0 in M i.e. there will be more than one distinct annulus in H with core curve freely homotopic to c_0 in M. If H_0 contains a component consisting of an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0, as shown in figure 2.13, each distinct annulus in H with core curve freely homotopic to c_0 in M has to intersect $S \times [0, \infty)$, otherwise H couldn't be embedded. For sufficiently small ϵ, the intersection of $S \times -\epsilon$ with H is freely homotopic in H to the intersection of $S \times 0$ with H. Every component A_0 of H_0 that is either an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0 is a subset of a component $A_{-\epsilon}$ of $H_{-\epsilon}$ that is either an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0. Consider the homotopy that fixes $H \setminus A_{-\epsilon}$ and maps the R coordinate of each component A_ϵ to R', where $R' = k(R + \epsilon) - \epsilon$, and k is a constant chosen to be small enough such that every component of A_ϵ is mapped into $S \times [-\epsilon, 0)$. Call the image of H under this homotopy H_2. This homotopy might create self-intersections along curves that are contractible in the surface, but by construction it can't create self-intersections along curves freely homotopic to c_0.

It remains to show that the intersections of H_2 along curves that are contractible in H_2 can be removed by a homotopy. Let $\phi : F \to H$ be a parametrisation of H_2, where F is a surface with boundary, and let c be a curve in H_2 along which H_2 has a self-intersection. Since H_2 is orientable, $\phi^{-1}(c)$ consists of two curves, c^1 and c^2. $\phi(c^1)$ and $\phi(c^2)$ are compressible in H_2, so they each bound a disc in H_2. The disc d_1 in H_2 bounded by c^1 could contain further contractible curves along which H_2 intersects itself, similarly for the disc d_2 in H^2 bounded by c^2. By the Jordan curve theorem, it makes sense to talk of the "innermost" curves, in d_1 and d_2 freely homotopic to c. Suppose c was chosen to be this innermost, contractible curve in d_1. The union of the two discs in H_2 bounded by $\phi(c^1)$ and $\phi(c^2)$ is an embedded 2-sphere in M. Since any 2-sphere in M bounds a ball, (Proposition D.3.17 of [2]), the points of intersection along c can be removed by a homotopy without creating new points of intersection. This can be repeated until an embedded surface is obtained, which can be smoothed off to obtain H'.

It follows from compactness of $S \times 0$ and H' that if a surface H'' homotopic to H' has a component of H''_0 consisting of an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0, H'' can't be arbitrarily close to H' in the Hausdorff topology. □

Proof of Theorem 32. It can be assumed without loss of generality that H is in general position. To start off with, suppose also that H is embedded.

If $c_1 - c_0$ is a multicurve, by lemma 15 it has to bound a subset of S. In this case, H has to be homotopic to a subset of S, as shown in corollary 30.

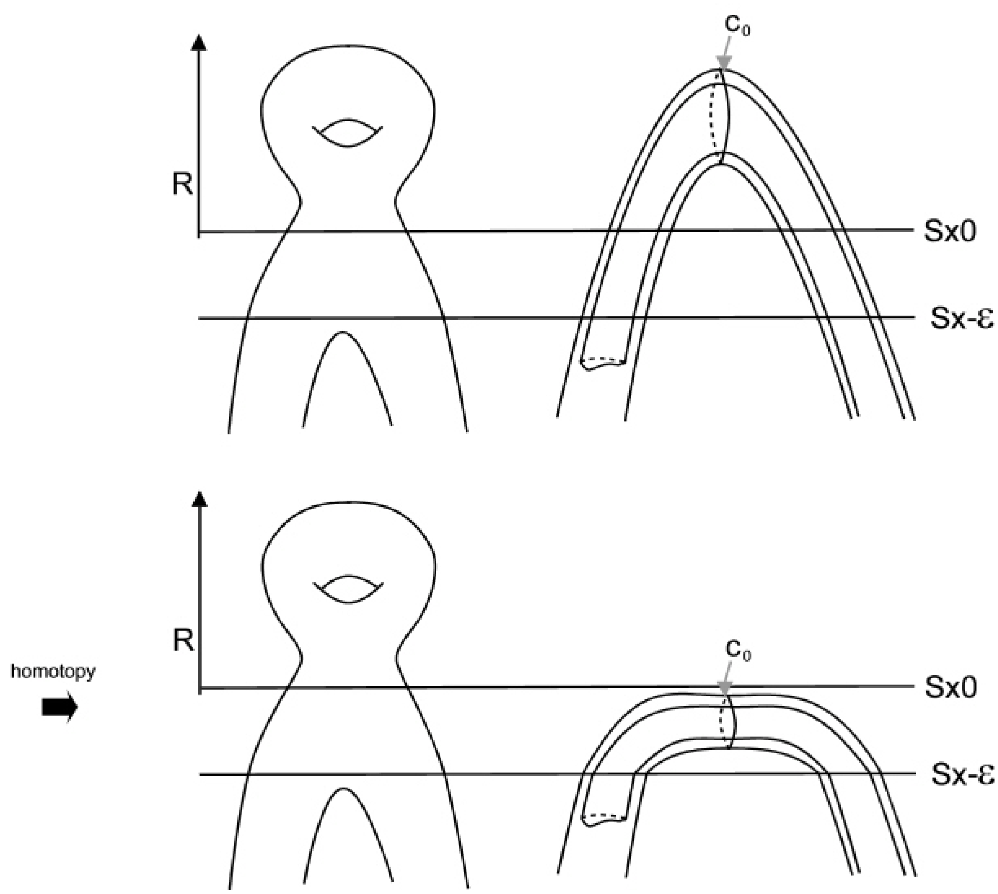

Figure 2.13: Homotopy that rescales the R coordinate of some components of $H_{-\epsilon}$.

Morse theory can be used to give handle decompositions of surfaces, which is what will be done here. In this proof, intersection properties of the projections of boundary curves and arcs to S are important. This makes it necessary to distinguish between two distinct methods of attaching handles, depending on the way the handle projects into S. Fix an orientation of S and a choice of projection π of M onto S. Suppose H^a contains two components, F_1 and F_2, that are subsets of a connected component F of $H^{a+\delta}$ for small δ. In other words, there is a component of $H^{a+\delta}$ obtained by attaching a handle to $F_1 \cup F_2$. A handle can be thought of as an oriented rectangle Q in M, whose boundary is a union of four arcs, each given an orientation as a subset of the boundary of Q. A pair of opposite sides of Q, q^1 and q^2, are glued along arcs on the boundary components of F_1 and F_2 respectively, in such a way that pairs of arcs with opposite orientation are glued together. In this way, an oriented surface F is obtained, such that F_1 and F_2 are oriented as subsets of F. Whenever F_1 and F_2 project onto subsets of S with opposite orientations, i.e. $\pi(F_1)$ is oriented as a subset of S, and $\pi(F_2)$ is oriented as a subset of $-S$, or vice versa, the handle Q has to be embedded in M with an odd number of half twists, otherwise the orientations of F_1 and F_2 can't be made to match up. The aim is to make a definition to distinguish between "ordinary" handles and those that contain twists, which will be called "bow tie" handles. There are two complications to doing this. A rectangle Q as in the previous example contains a subset without twists; the representative of the homotopy class of $H^{a+\delta}$ might be chosen such that this subset of Q could be viewed as an "ordinary" handle connecting two components of H^a. An "ordinary" handle, when given a half twist in one direction and a half twist in the other direction to cancel it out, could be viewed as two handles with twists. In order to avoid these problems, when determining whether or not a handle contains twists, a representative of the homotopy class of H^a is chosen that avoids all nonessential points of intersection of its boundary when projected into S. With this choice of representative of the homotopy class of H^a, a handle Q has twists if and only if there is no homotopy of Q in M relative to its boundary arcs q_1 and q_2 in $H(a)$ such that $\partial\pi(Q)$ is embedded in S.

Definition 41 (Bow tie handle)
Suppose that for arbitrarily small δ, $H^{a+\delta}$ is obtained from H^a by adding a handle Q as described above. Suppose also that the projection of ∂H^a to $S \times 0$ only contains essential intersections. Let $h(t)$ be a homotopy of $H^{a+\delta}$

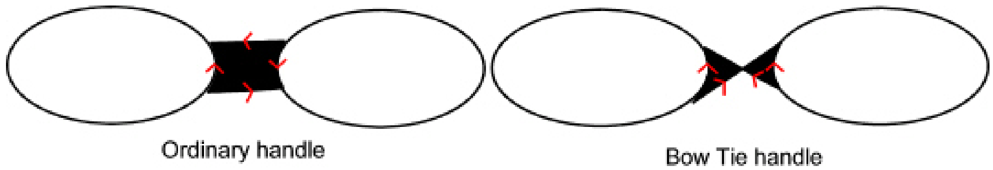

Figure 2.14: Examples of the two different types of handles.

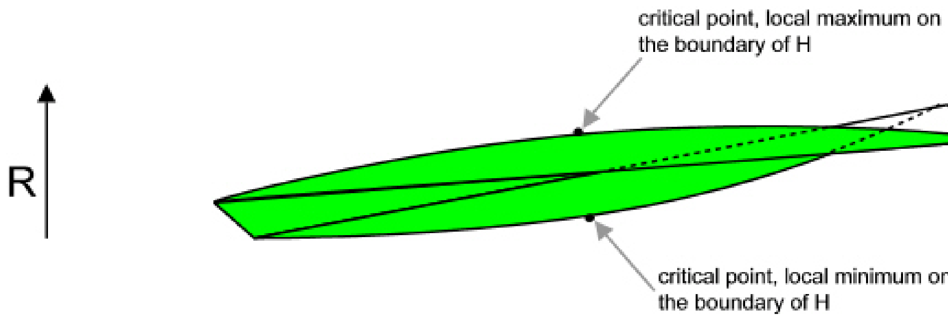

Figure 2.15: Cell decomposition of a bow tie handle. One side of the handle is shown in green.

in M that fixes H^a. If for arbitrarily small δ there does not exist a $h(t)$ such that the image of Q under the homotopy $h(1)(H^{a+\delta})$ has a boundary that projects one to one into S, Q will be called a bow tie handle.

The Morse function obtained by restricting the R coordinate to a suitable choice of representative of the homotopy class of a surface will not give a handle decomposition that contains a bow tie handle. A bow tie handle is a union of 2-handles of the handle composition obtained from R, as shown in figure 2.15.

There are arbitrarily many pant decompositions of H, and unless H is homotopic to a subset of S, it isn't the case that every pant decomposition of H gives rise to a surface producing sequence. It therefore doesn't make sense to prove results that are independent of the choice of Morse function, because most Morse functions on H give rise to handle decompositions that don't give any control over intersection numbers of curves projected into S. For the next results it is assumed that all Morse functions are obtained by restricting R to some embedded representative of the homotopy class of the surface H, as described.

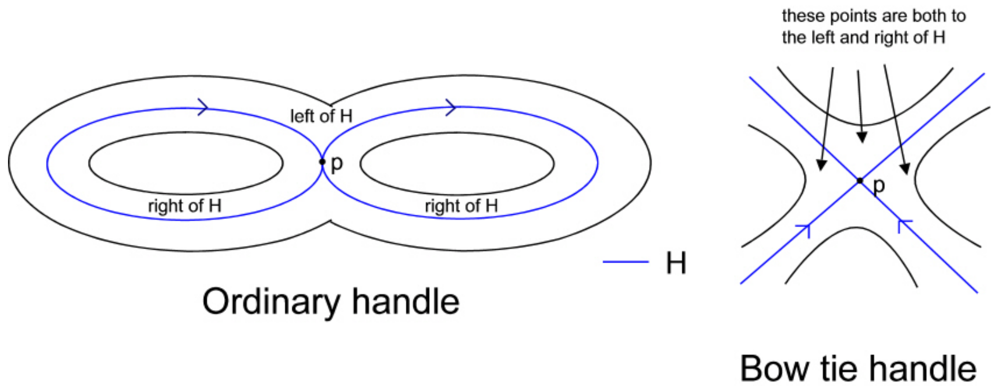

Figure 2.16: Intersection of the I bundle of H with $S \times a$

Since H is embedded, whenever a is not a critical value, $H(a) := H \cap (S \times a)$ is a union of curves and arcs that project one to one into S. If a is a critical value and $H_0^{a+\delta}$ is obtained from $H_0^{a-\delta}$ by adding a handle, $H(a)$ is a one dimensional cell complex. The representative of the homotopy class of H was chosen such that there can be at most one critical point for any value of R, so δ can be chosen small enough to ensure that $H^{a+\delta}$ is obtained from $H^{a-\delta}$ by attaching a single handle. Either this handle is a 2-disc attached along a boundary component of $H_{a-\delta}$ or there is a point p in $H^a_{a-\delta}$ such that $H^a_{a-\delta}$ has a component consisting of two 2-cells attached at a vertex p. If the handle is an ordinary handle, these two cells project onto two subsets of S, both of which are either oriented as a subset of S or as a subset of $-S$. If the handle is a bow tie handle with an odd number of twists, the two cells will have opposite orientations when projected onto subsets of S.

By theorems 6.1 and 6.3 of [16], it is possible to find a smooth injective map ϕ from the normal bundle of $H^{a+\delta}_{a-\delta}$ into M whose image, N, is embedded. Whenever the handle is an ordinary handle, $H^{a+\delta}_{a-\delta}$ is separating in N. In this case, the point p is a point at which two arcs in $H(a)$ touch but do not cross over. This is because if $H(a)$ were to consist of two arcs crossing over at p, one component of $\phi(H^{a+\delta}_{a-\delta} \times (I \setminus 0))$ would be connected up to the component on the other side, which is a contradiction to the assumption that the handle has a neighbourhood that could be projected one to one onto a subset of S. Therefore, the point p does not represent an essential point of intersection for the boundary component of $H^{a+\delta}_{a-\delta}$ containing $H(a)$. If the handle is a bow tie handle with an odd number of half twists, N is the image of a twisted I

bundle over $H_{a-\delta}^{a+\delta}$, and $H_{a-\delta}^{a+\delta}$ is not separating in N. Similarly, if the handle is a twisted handle, $H(a)$ has to have a point of intersection when projected into S. If the handle has more than one half twist, there will be more than one point of intersection, but not all of these points of intersection will be essential.

It can be assumed without loss of generality that 0 is not a critical value of the Morse function. $H \cap (S \times 0)$ therefore consists of a multicurve m_0 containing c_0 and a union of arcs, so the boundary curves of H_0 consist of m_0 and curves consisting of arcs of $H \cap (S \times 0)$ connected to arcs of $c_1 \cap (S \times [0, \infty))$. The curves $\partial H_0 \setminus m_0$ don't have any more points of intersection with c_0 than does c_1, but unlike c_1, there could arise essential points of (self)intersection when projected into S. If there are any arcs in the intersection of H_0 with $S \times a$, the boundary curves of H_0^a include curves consisting of arcs on the boundary of H_0 connected up to arcs of $H_0 \cap (S \times a)$.

The theorem will now be proven under the assumption that there are no bow tie handles. The assumption that H has smallest possible genus will then be used to rule out the necessity of bow tie handles for the given Morse function.

c_0 is a simple curve, so the surface $S \times 0$ could have been chosen to contain the boundary curve c_0 of H. This choice of the zero of the R coordinate results in a boundary curve c_0 consisting of degenerate critical points. First of all it is convenient to show that there exists an embedded representative of the homotopy class of H such that, by mapping R to $-R$ if necessary, the boundary curve c_0 has a collar in H, $c_0 \times [0, \epsilon)$, contained in $S \times [0, \infty)$. Once the existence of this collar has been established, it will be shown that there exists an embedded surface $H^{'}$ homotopic to H with the following properties:

1. $H^{'}$ is arbitrarily close to H in the Hausdorff topology
2. the restriction of R to $H^{'}$ is a Morse function, and
3. there exists a noncritical value r of R such that $S \times r$ intersects $H^{'}$ along the union of curves and arcs $H^{'(r)}$, where $H^{'(r)}$ contains a curve freely homotopic to c_0.

Suppose $S \times 0$ was chosen to contain the boundary curve c_0 of H. The existence of an embedded representative of the homotopy class of H whose boundary curve c_0 contains a collar in $S \times [0, \infty)$ follows from the assumption that there are no bow tie handles in H. Let $T(\epsilon)$ be a toroidal neighbourhood in M of the boundary curve c_0 of H, where $T(\epsilon)$ consists of all points within a distance 2ϵ of a point on the boundary corve c_0 of H, and let $T(\frac{\epsilon}{2})$ be the set of all points within distance ϵ of a point on the boundary curve c_0 of H. By theorem 6.3 of [16], ϵ can be chosen to be so small that $T(\epsilon) \cap H$ is connected. Let $c_0^{'}$ be the curve in $T(\frac{\epsilon)}{2}$ homotopic to c_0 with R coordinate ϵ such that $\pi(c_0^{'}) = c_0$, and let $c_0^{"}$ be the intersection of H with $\partial T(\epsilon)$. Since, by assumption, H does not have any bow tie handles, the intersection of H with $\partial T(\frac{\epsilon}{2})$ is homotopic on $\partial T(\frac{\epsilon}{2})$ to $c_0^{'}$. Also because of the assumption that H does not have bow tie handles, $c^{"}$ is homotopic in the closure of $T(\epsilon) \setminus T(\frac{\epsilon}{2})$ to $c^{'}$. It follows that $H \cap T(\epsilon)$ is homotopic in $T(\epsilon)$ to an embedded surface $c_0 \times [0, \epsilon] \cup A$, where A is an embedded annulus in $T(\epsilon) \setminus T(\frac{\epsilon}{2})$ with boundary curves $c_0^{'}$ and $c_0^{"}$. A surface chosen to coincide with H outside of $T(\epsilon)$ and with $c_0 \times [0, \epsilon] \cup A$ inside $T(\epsilon)$ can be smoothed off to give an embedded representative, H^c, of the homotopy class of H whose boundary curve c_0 has the desired collar.

It follows from the previous lemma that there is a surface $H^{'}$ homotopic to H arbitrarily close to H^c in the Hausdorff topology to which the restriction of the R coordinate is a Morse function. Whenever $H^{'}$ is sufficiently close to H^c in the Hausdorff topology, the boundary curve c_0 has a collar in $H^{'}$ such that the intersection of $S \times r$ with this collar contains a curve freely homotopic in $H^{'}$ to the boundary curve c_0, for some small, noncritical value r of R. H^c can be made arbitrarily close to H by choosing ϵ arbitrarily small, so $H^{'}$ can also be made arbitrarily close to H in the Hausdorff topology.

It can therefore be assumed without loss of generality that the zero of the R coordinate and the embedded representative of the homotopy class of H are chosen such that the restriction of R to H is a Morse function, $H(0)$ contains a curve freely homotopic to the boundary curve c_0 and no two critical points occur at the same value of R. With this choice of the zero of the R coordinate, the other boundary curve c_1 might intersect $S \times 0$ in a complicated way.

If a is so small that there are no critical points of R in the interval $[0, a]$, then H_0^a is a union of annuli whose core curves project onto a multicurve in S and perhaps some contractible components. Suppose now that a is large enough to ensure that there is only one critical value, b, in the interval $[0, a]$. If H_0^a contains a simply connected component that intersects some $S \times (a - \delta)$ along an arc or a contractible curve, and if this component wasn't in H_0^x for $x < b$, then the critical point has not changed the topology of the component

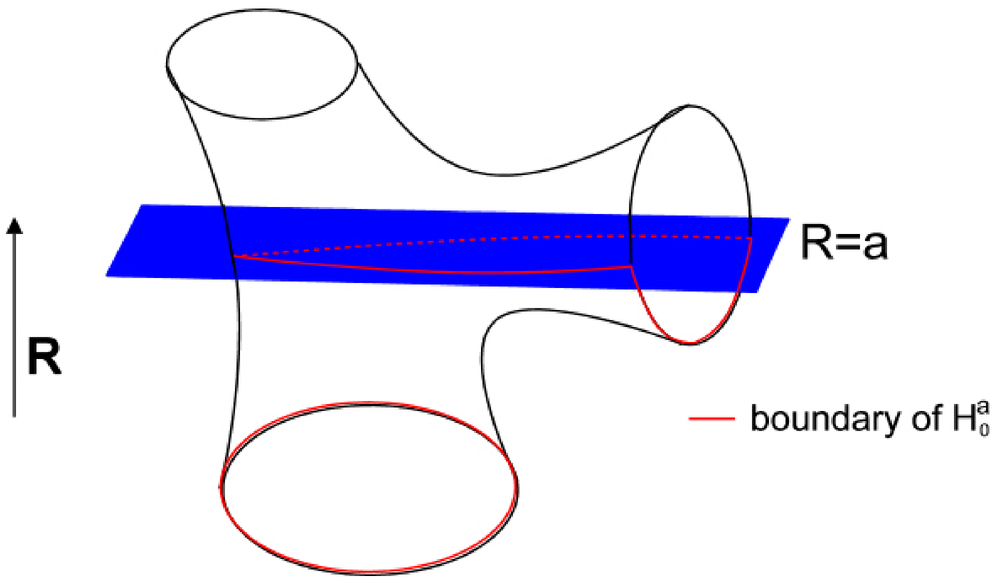

Figure 2.17: A local minimum on the boundary at $R = b$ for $b < a$.

of H_0^a with c_0 on its boundary. Similarly, if one of the boundary components of H_0^a has a local minimum at $R = b$, this can only change the representatives of the free homotopy classes of the curves on the boundary of H_0^x as x passes through b, as shown in figure 2.17.

If a is large enough for there to be a saddle point p in the interior of H_0^a, this saddle point could cancel out a local minimum as shown in figure 2.18. A critical point of this type also only changes the representatives of the homotopy classes of the curves on the boundary of H_0^a and/or the number of contractible components.

If a is now chosen such that in the interval $(0, a]$ there is either:

1. a local maximum (either in the interior or on the boundary) or

2. a saddle point that does not cancel out a local minimum,

then the topology of H_0^x changes as x moves through the critical value b. In particular, H_0^a is obtained from H_0^b (a disjoint union of contractible components and annuli whose core curves project onto a multicurve in S) by adding a handle. If this handle has one endpoint on a contractible component of H_0^b, again, the topology of the component to which the handle was added doesn't change when passing through the critical value. Otherwise, the endpoints of

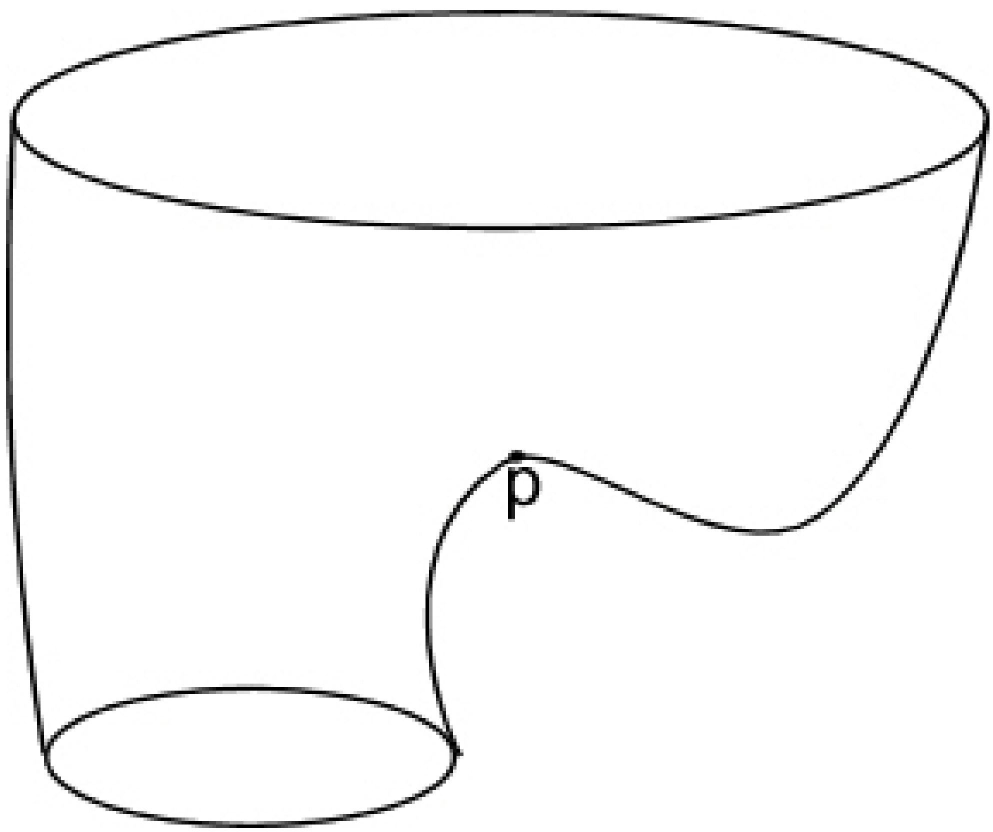

Figure 2.18: A saddle point cancelling out a local minimum.

the handle are either both on the same annulus or on two different annuli. Whenever both of the endpoints of the handle are on the boundary of the annulus with core curve c_0, H_0^a contains a pair of pants with boundary curves c_0 and the multicurve $\alpha \cup \beta$. (There is a second alternative here, namely that the handle has one endpoint on each boundary component of the annulus with core curve c_0. However, this doesn't happen here, because the handle is attached to the boundary component $H(b)$ which only contains one curve freely homotopic to c_0.) $\alpha \cup \beta$ is homotopic to a multicurve because it is a subset of the intersection of the embedded surface H with $S \times b$, where the assumption that the handle is not a bow tie handle is being used here.

$c_0 \cup \alpha \cup \beta$ is also a multicurve, because $\alpha \cup \beta$ is constructed by adding a single handle to c_0, where the handle is a subset of $S \times a$ without self intersections that meets the projection of c_0 onto $S \times a$ only at its endpoints (The assumption that the handle is not a bow tie handle is being used here also). Therefore the pair of pants projects onto a pair of pants in S and γ_1 can be taken to be $\alpha \cup \beta$, unless one of α or β is contractible. If one of α or β is contractible, β for example, then H contains an annulus with boundary curves c_0 in $S \times 0$ and α in $S \times a$ that intersects $S \times i$ for some values of i in a disconnected set.

Similarly if one of the endpoints of the handle is on the boundary of the annulus with core curve c_0 and the other is on the boundary of another annulus whose core curve α is a subset of the multicurve m_0. β is then the curve obtained by connecting the annuli with core curves c_0 and α by a handle, and $c_0 \cup \alpha \cup \beta$ is a multicurve for the same reason as in the previous case. Again, γ_1 can be taken to be $\alpha \cup \beta$ unless one of α or β is contractible.

If the handle doesn't have an endpoint on the annulus with core curve c_0, then the intersection of H_0 with $S \times a$ will be a union of arcs plus a new multicurve, m_1, containing c_0. That m_1 is a multicurve follows from the assumption that there are no bow tie handles as before. As a result of lemma 40, it is possible to make an additional assumption on the representative of the homotopy class of H that simplifies the natation at this point. It will be assumed that there is no component of H_0 consisting of an annulus with core curve c_0 or a punctured sphere whose boundary curves are either contractible or homotopic to c_0. Therefore, whenever the component of H_0^a with c_0 on its boundary consists of an annulus with core curve c_0 or a bordered sphere whose boundary curves are either contractible or freely homotopic to c_0, if a is increased enough, there will either be a pair of pants in H with c_0 on its boundary or there will be another critical point of R on the component of H_0^a

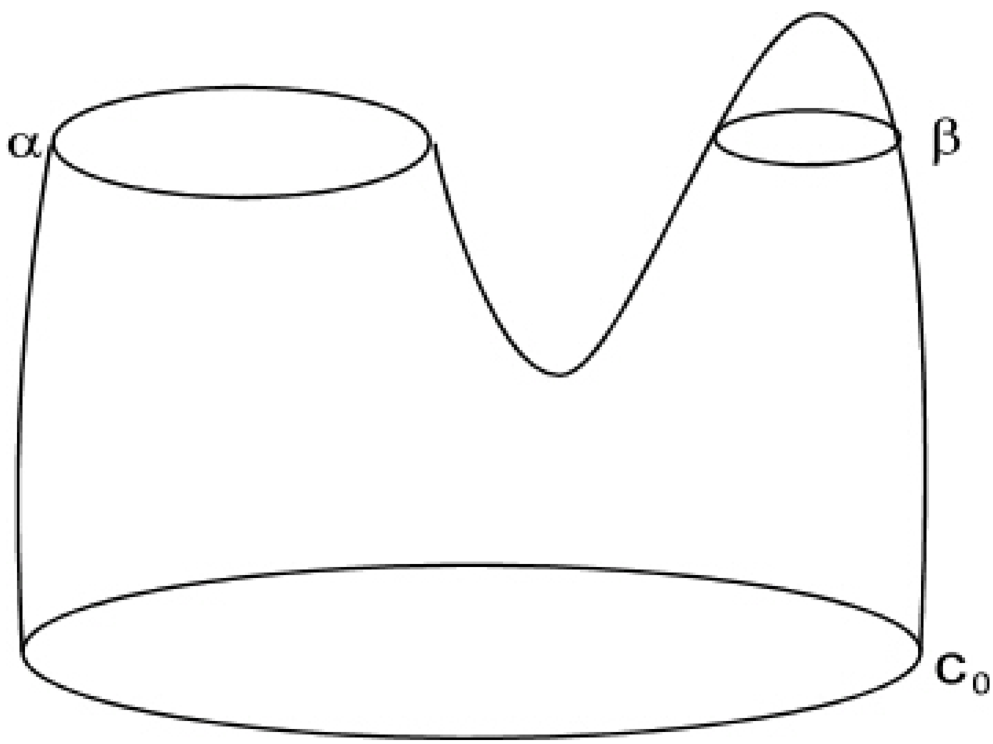

Figure 2.19: An annulus in H that intersects $S \times a$ in a disconnected set.

with c_0 on its boundary. Since there are only finitely many critical points, eventually the desired pair of pants is obtained, and γ_1 can be defined.

To construct γ_2, cut the pair of pants with boundary $c_0 \cup \gamma_1$ off H to obtain an embedded surface H_1 with boundary $c_1 - \gamma_1$. Since γ_1 is a multicurve, the previous argument can be applied with γ_1 in place of c_0 and H_1 in place of H. If γ_1 is not in the intersection of H with a level set of R, this involves using a modified Morse function, R_1, to obtain the second pair of pants. $\gamma_3, \gamma_4...$ are constructed similarly. For a surface H_γ constructed via a surface producing sequence, it is then clear that a representative of the homotopy class of H_γ in M can be found so that there is a choice of R coordinate that gives the same handle decomposition as the Morse function R when restricted to the subsurface of H bounded by $\gamma_1 - c_0$, R_1 when restricted to the subsurface of H bounded by $\gamma_2 - \gamma_1$, etc, so it is still valid to refer to "the" Morse function when different Morse functions were used on different subsets of H.

If H is not embedded, let D be the set of all double and triple points of H. By assumption, H is in general position, so, as discussed in chapter one of [14], D is a union of arcs with endpoints on the boundary, closed curves and triple points, such that D does not accumulate anywhere. It is possible to assume without loss of generality that there are no curves that are homotopic to c_1 or c_0, or arcs homotopic (relative to their endpoints) to a subarc of the boundary of H in D, because all such intersections can be removed by a homotopy.

Let d_1 be an arc or curve in D, and let a be a curve in the cell complex $H \cup D$ that isn't homotopic to a curve in H and whose intersection with H is connected. By theorem 3.3 of [25], there is a normal subgroup of $\pi_1(S)$ of finite index that contains curves homotopic to c_0 and c_1 but not a. The CW complex $H \cup D$ is compact, so $\pi_1(H \cup D)$ is finitely generated, and it is possible to choose a finite generating set consisting of curves in H and curves $a_1, a_2, ...a_n$ freely homotopic to curves in $H \cup D$ whose intersection with H is a connected arc. If $H \cup D$ is incompressible, for each of the a_i there is a normal subgroup of $\pi_1(S)$ of finite index containing c_0 and c_1 but not a_i. The intersection of all these subgroups is a normal subgroup N of finite index containing c_0 and c_1 but none of the a_i. Let $\tilde{S}$ be the covering space of S with $\pi_1(\tilde{S}) = N$. Then $\tilde{S} \times R$ is a covering space of M such that the lift $\tilde{H}$ of H to $\tilde{S} \times R$ is embedded. If it is not possible to choose all of the a_i such that they are not contractible in M, construct a covering space whose fundamental group contains c_0 and c_1 but none of the elements of a_i that are not contractible in M, and let D' be the set of all double and triple points of the lift of H, H', to this covering space. In this covering space, any curve in $H' \cup D'$ whose intersection with H' is a connected arc is contractible in the lift, M' of M. Since any 2-sphere in M' is also contractible, (Proposition D.3.17 of [2]), it follows that the points of intersection in D' are not essential.

It is being assumed that there are no bow tie handles in the handle decomposition of H given by the Morse function R. The definition of bow tie handle uses only local properties of the projection function that are preserved under the covering transformation, so it follows that there are no bow tie handles in the handle decomposition of $\tilde{H}$ in the handle decomposition given by the R coordinate of $\tilde{S} \times R$. The previous argument applies just as well to $\tilde{S}$ as to S, so $\tilde{H}$ is homotopic to a surface constructed from a surface producing sequence in the homology multicurve graph of $\tilde{S}$. In other words, $\tilde{H}$ is constructed by attaching surfaces with boundary that are homotopic to subsets of $\tilde{S}$. $\tilde{S}$ can be covered by a finite number of neighbourhoods that project one to one onto S, and so therefore can $\tilde{H}$. Since the boundary curves c_0 and c_1 of $\tilde{H}$ project one to one into S, these neighbourhoods can be used to construct a surface producing sequence γ in the homology multicurve graph of S, where H is homotopic to H_γ.

It remains to show that if H has smallest possible genus the surface can be constructed without bow tie handles. To do this, some properties of bow tie handles will be shown. Since the previous argument also applies to surfaces that are not embedded, it can be assumed (and will be assumed from now on) without loss of generality that all bow tie handles have a single half twist.

Definition 42 (Twisted skirt)
A twisted skirt is homotopic to an annulus with a bow tie handle attached. The bow tie handle is also required to have both endpoints on the same boundary component of the annulus. Since it is being assumed that all twisted handles only have a single half twist, a twisted skirt has only two boundary components, hence the name. Alternatively, a twisted skirt is a one holed Möbius band.

The handle decomposition of H by the Morse function R gives a pants decomposition of H, which is not allowed to contain any twisted skirts, because H is assumed to be oriented. As a consequence, a bow tie handle can't have both endpoints on the same boundary component of H_0^a. Suppose that the boundary of H_0^a consists of a union of (not necessarily simple) curves c_a. (Note that c_a is the boundary of H_0^a which is not the same as the intersection

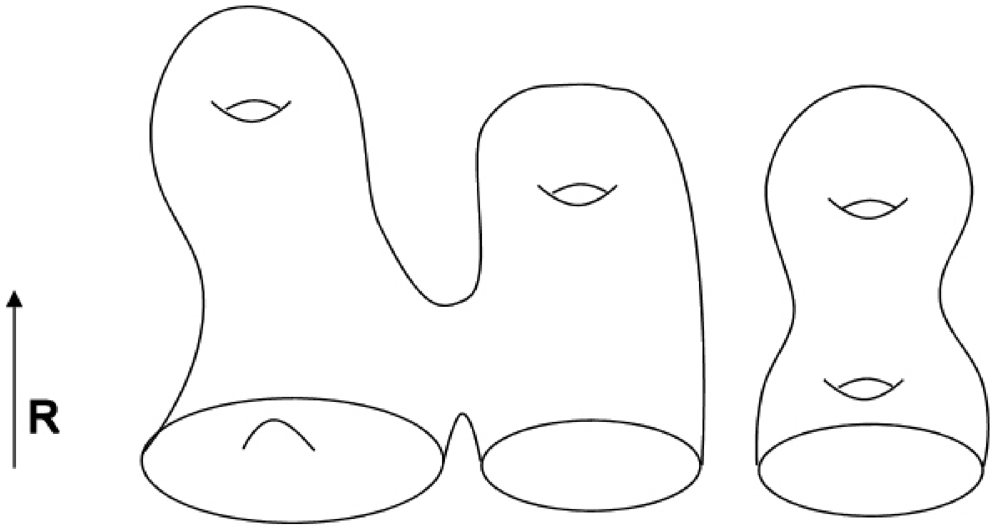

Figure 2.20: The order in which the handles are added does not depend only on the R coordinate of the corresponding critical point.

of H with $S \times a \cup S \times 0$). If $H_0^{a+\delta}$ is constructed from H_0^a by adding a bow tie handle with a half twist, then the boundary of $H_0^{a+\delta}$, $c_{a+\delta}$, could only have smaller intersection number with c_1 than c_a if the bow tie handle has both endpoints on the same boundary component of H_0^a. This is because, if the bow tie handle has endpoints on two different boundary components of H_0^a, $c_{a+\delta}$, when projected onto S, is a one dimensional cell complex with a subcomplex freely homotopic in M to c_a. It follows that the handle decomposition of H can't consist of bow tie handles only, otherwise c_1 would have to have nonzero intersection number with itself.

A handle decomposition of H defines a set of handles and an order in which they are attached to the component with boundary curve c_0. This ordering of the handles is not quite the same thing as the R coordinate of the corresponding critical points, since H_0 could have many components.

By changing the order in which handles are added, a union of (not necessarily simple) curves β_j on H can be constructed, where $H \setminus \beta_j$ has two components, and β_j is contructed from c_0 by attaching ordinary handles only. β_j is also chosen such that the component of $H \setminus \beta_j$ with boundary curves $c_1 - \beta_j$ has a handle decomposition consisting only of bow tie handles with a half twist. To see that a β_j with these properties exists, suppose i is as large as possible such that that the first i handles are all ordinary handles. The R

coordinate is chosen such that there is a multicurve freely homotopic to γ_i in the intersection of H with $S \times 0$, where γ_i is understood to be c_0 whenever $i = 0$. Let $H_0^a(\gamma_i)$ be the component of H_0^a with γ_i on its boundary, and let a_1 be the smallest value of R such that, for arbitrarily small δ, $H_0^{a_1+\delta}(\gamma_i)$ is obtained from $H_0^{a_1}(\gamma_i)$ by adding an ordinary handle. It might also be the case that there is no value of a_1 for which $H_0^{a_1+\delta}(\gamma_i)$ is obtained from the $H_0^{a_1}(\gamma_i)$ by adding an ordinary handle, because the corresponding critical values of R are less than zero. In this case it is necessary to work with the component of H_{a_1} with γ_i on its boundary, $H_{a_1}(\gamma_i)$, where a_1 is the largest negative number such that $H_{a_1-\delta}(\gamma_i)$ is obtained from $H_{a_1}(\gamma_i)$ by adding an ordinary handle. The argument is however exactly the same in this second case.

The multicurve γ_i can be homotoped inside $H_0^{a_1}(\gamma_i)$ to $\gamma_i^{'}$, where the subset of $H_0^{a_1}(\gamma_i)$ with the multicurve $\gamma_i^{'}$ on its boundary instead of γ_i has zero area i.e. the union of annuli with boundary curves $\gamma_i^{'} - \gamma_i$ "fill up" the surface $H_0^{a_1}(\gamma_i)$. The ordinary handle at $R = a_1$ can be attached to the union of annuli with boundary curves $\gamma_i^{'} - \gamma_i$ to obtain a surface with boundary $\beta_{i+1} - \gamma_i$, where it does not follow from the previous argument that β_{i+1} is a multicurve, because the boundary curves and the handles added are not all embedded in M at the same R coordinate. β_{i+2}, β_{i+3}... are constructed similarly, until β_j is obtained, after which there are no ordinary handles left. Since the handle decomposition of the subset of H with boundary $c_1 - \beta_j$ consists of bow tie handles only, β_j can't intersect c_1. c_1 is a simple curve, so β_j can't have self intersections, because the only way to reduce the number of self intersections of the boundary of a surface by adding a bow tie handle is to attach a handle with both endpoints on the same boundary component. $c_1 - \beta_j$ is therefore a multicurve. β_j can't contain more than one curve that is not contractible because it has to be possible to obtain the connected curve c_1 by adding bow tie handles only. If a bow tie handle with endpoints on two distinct, noncontractible simple curves is attached, the number of curves on the boundary is reduced by one, but the new boundary curve has a point of self intersection that can't be removed by attaching bow tie handles only. Therefore, β_j consists of a curve freely homotopic to c_1 and perhaps some contractible curves. The order in which the handles are attached can therefore be chosen such that the bow tie handles all have to have at least one endpoint on a contractible curve. Since $c_1 - \beta_j$ is a multicurve, the subset of H bounded by $c_1 - \beta_j$ has to be homotopic to a union of subsets of S, otherwise it would be possible to construct a surface with boundary $c_1 - c_0$

with smaller genus than H. A subset of H homotopic to a union of subsets of S can be embedded in M in such a way that R gives a handle decomposition of this set without bow tie handles. The theorem then follows from the proof of the result under the assumption that there are no bow tie handles.

□

Chapter 3

Freely Homotopic Curves

In the previous chapter, sequences of oriented multicurves were used to construct surfaces. However, not very much was shown about the multicurves themselves. In practice, it is often very helpful to have a bound on the number of curves in each multicurve. The number of free homotopy classes of curves in a given multicurve is bounded from above by $3g-3$, where g is the genus of S, and the multicurves are also constructed in such a way that they do not contain freely homotopic curves with opposite orientations. One advantage of the algorithm outlined in chapter one is that the number of curves in each of the multicurves is automatically kept bounded, as the next theorem shows. An example will be given to show that this is not true in general, even when a multicurve does not contain null homologous submulticurves.

Theorem 43
Let the curves c_o, c_1 be defined as usual, and $\{c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1\}$ a surface producing sequence constructed with a sequence of functions $f, f_1, f_2...f_j$ as described in the previous chapter. Then none of the multicurves γ_k contain curves that are freely homotopic.

Proof. To shorten the terminology, if a multicurve contains freely homotopic curves with the same orientation, will say that the multicurve has "doubled curves". Recall that, by definition, no multicurve in a surface producing sequence contains two freely homotopic curves with opposite orientation. Suppose that i is the largest number such that, for $l \leq i$, γ_l does not contain doubled curves. i is at least one, because if γ_1 were to contain doubled curves, $c_0 - \gamma_1$ would not bound a subset of S, which would contradict the assumption that $\{c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1\}$ is surface producing. For the same reason, the number of curves in any free homotopy class can only increase or decrease by one when going from γ_n to γ_{n+1} for any n. Suppose there is a curve c in γ_i such that γ_{i+1} contains two curves freely homotopic to c and with the same orientation. Then γ_{i+2} has to contain at least one copy of c, so c has to be disjoint from γ_{i+3}. Therefore c is disjoint from $\gamma_{i-1}, \gamma_i, \gamma_{i+1}, \gamma_{i+2}$ and γ_{i+3}.

Call h_n the function obtained from $\gamma_{n+2} - \gamma_n$. $0 \leq h_n \leq 2$, whenever γ_{n+2} and γ_n are chosen to be representatives of their free homotopy classes that only have essential points of intersection. A specific choice of representatives of the free homotopy classes is necessary for the rest of this proof. Suppose that γ_{i-1} and c_1 are in general position and only have essential points of intersection. $\gamma_i, \gamma_{i+1}, ...\gamma_j$ are representatives of their free homotopy classes that are subcomplexes of the CW complex $\gamma_{i-1} \cup c_1$ such that the subset of S bounded by $\gamma_{k+1} - \gamma_k$ coincides with f_{kmax} for $i - 1 \leq k$, where γ_{j+1} is understood to be c_1. With this choice of representatives of the free homotopy classes, h_{i-1max} is the same subset of S as f_{i-1max} and the subset of S on which h_{i-1} is greater than or equal to one is the same as the subset of S on which f_{i-1} is equal to a least one less than its maximum. Warning: these representatives of the free homotopy classes will not be in general position.

Also as discussed in the previous chapter, the subcomplexes γ_i and γ_{i+1} chosen in this way could have points of transversal intersection with c_1 that could be removed by choosing another subcomplex in the same free homotopy class on S. This is because there could be arcs of $c_1 \cap (S \setminus \gamma_{i-1})$ that are not on the boundary of f_{i-1max} but are freely homotopic to an arc of $c_1 \cap (S \setminus \gamma_{i-1})$ that is.

Definition 44 (Point of intersection for curves not in general position)
The left and right side of an oriented curve in S was defined in chapter two. Suppose a is an oriented curve and b is a second curve such that a and b are not in general position. If a and b coincide along some subarc, these points will only be counted as a point of intersection if b crosses from one side of a to the other.

γ_{i-1} and γ_{i+1} intersect in horizontal arcs only since the existence of a vertical arc would force the maximum of h_{i-1} to be at least 3, as can be easily verified in the diagram below.

It follows that adjacent arcs in a homotopy class of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ have opposite orientation. Since the maximum of h_{i-1} is two, any arc of

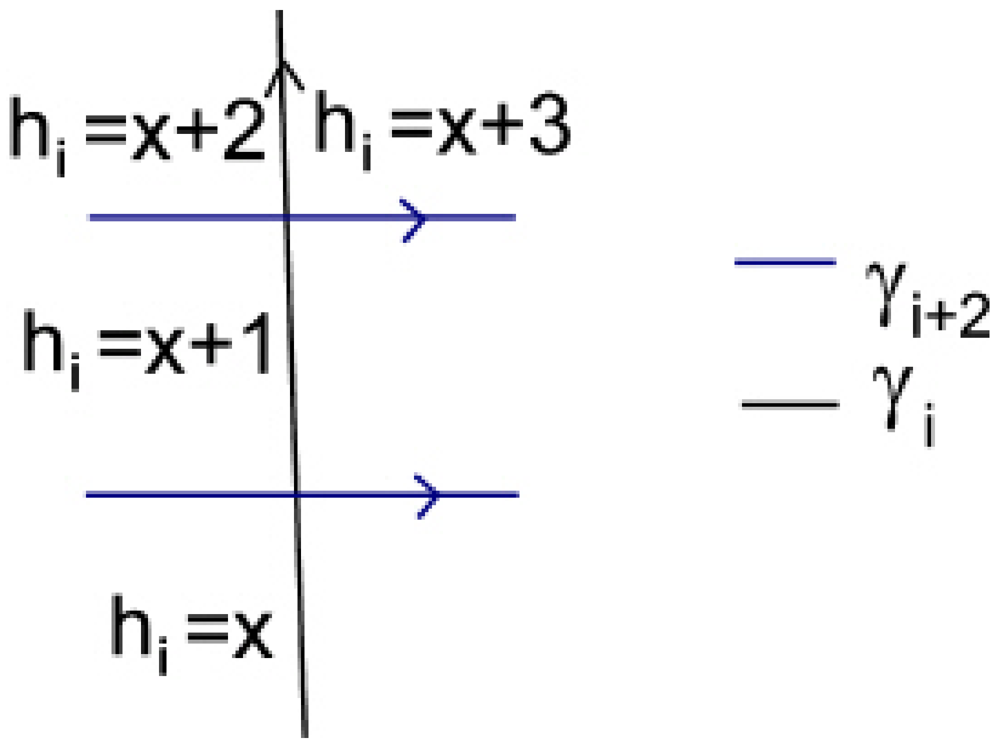

Figure 3.1: The existence of a vertical arc forces the maximum of f to be at least three.

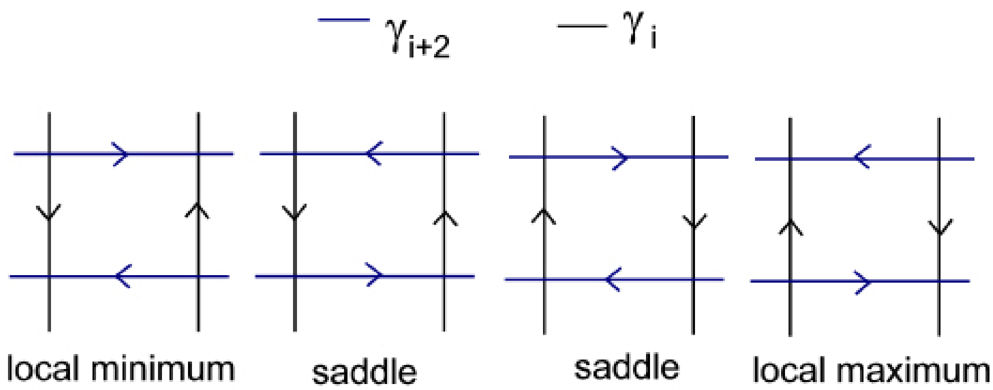

Figure 3.2: Components of $S \setminus (\gamma_{i-1} \cup \gamma_{i+1})$.

$\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ is either on the boundary of h_{i-1max} or h_{i-1min}. An arc of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ on the boundary of h_{i-1max} is to the right of γ_{i-1} whereas an arc of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ on the boundary of h_{i-1min} is to the left of γ_{i-1}, so by lemma 21 of chapter two, if a_k is an arc of γ_{i+1} on ∂h_{i-1max}, it can only be homotopic to other arcs on ∂h_{i-1max}. Therefore, this choice of representatives of the homotopy classes γ_k can be made without any two of the multicurves γ_{i-1}, γ_i or γ_{i+1} having points of intersection that can be removed by a free homotopy. In particular, this choice of the representatives of the free homotopy classes γ_{i+1} and γ_{i-1} won't have points of intersection, essential or otherwise, with the curve c in γ_i.

The curves on ∂f_{i-1max} or ∂h_{i-1max} are mostly constructed by alternately connecting arcs of $c_1 \cap (S \setminus \gamma_{i-1})$ to arcs of $\gamma_{i-1} \cap (S \setminus c_1)$, however it might also happen that a component of ∂f_{i-1max} is an entire curve contained in $\gamma_{i-1} \cup c_1$. This curve has to be a curve in γ_{i-1}, since c_1 is connected and therefore can't be on the boundary of f_{i-1max} unless $i-1=j$. Since ∂f_{i-1max} does not intersect c_1, this curve can't either.

Since γ_{i-1} and γ_{i+1} can intersect in horizontal arcs only, it follows that $S \setminus (\gamma_{i-1} \cup \gamma_{i+1})$ consists only of extrema or saddle points of h_{i-1}. The boundary of a component of $S \setminus (\gamma_{i-1} \cup \gamma_{i+1})$ therefore looks like one of the examples in the diagram below, in the sense that the arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ on its boundary are all oriented in such a way that the component of $S \setminus (\gamma_{i-1} \cup \gamma_{i+1})$ is either to their left or to their right, similarly for the arcs of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$.

If γ_{i-1}^{int} is the subset of curves in γ_{i-1} that intersects γ_{i+1} and γ_{i+1}^{int} is the subset of curves in γ_{i+1} that intersects γ_{i-1}, then $\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int}$ cuts S into pieces

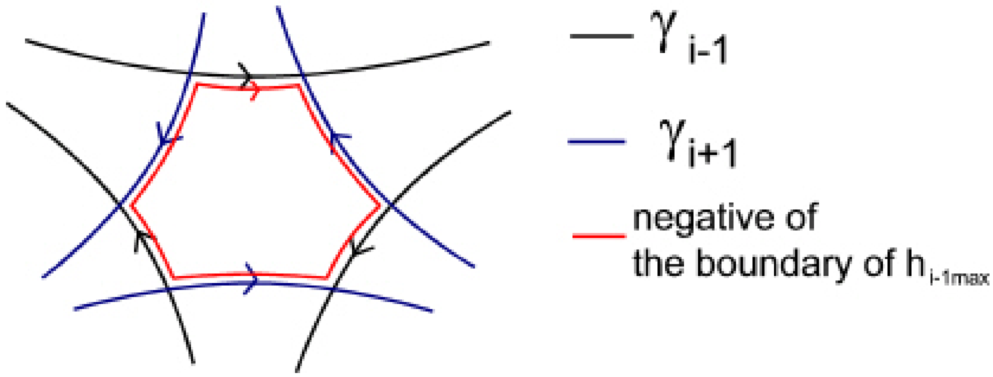

Figure 3.3: The null homologous curve $-\partial h_{i-1max}$.

that are extrema or saddle points of h_{i-1}. A curve such as c that is not in $\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int}$ has to be contained in one of these pieces. $\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int}$ can't contain any curves freely homotopic to c, so there can be at most one component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ in which the curves freely homotopic to c can be found.

Let $a_1, a_2, ...a_m$ be the arcs of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ on the boundary of h_{i-1max}. By assumption there are two curves in γ_{i+1} freely homotopic to c. Suppose γ_{i-1} does not have a curve freely homotopic to c. Then a curve freely homotopic to c has to be created by the surgery in which γ_i is obtained from γ_{i-1}. The only points of intersection of γ_{i+1} with γ_{i-1} are on the boundary of h_{i-1max}. This is because the maximum of h_{i-1} would otherwise have to be larger than two, since, as already discussed, $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ consists of horizontal arcs only, and each arc of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ to the right of γ_i is on the boundary of h_{i-1max}. A curve freely homotopic to c is therefore constructed by alternately connecting arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ to some subset of $\{a_k\}$ or by alternately connecting arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ to some subset of $\{-a_k\}$. In the second case, c is part of the null homologous multicurve $-\partial h_{i-1max}$, in which case it will be cancelled out, i.e. it will not be seen in γ_i. This has to do with the fact that the arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ on the boundary of h_{i-1max} are oriented in the opposite way to arcs of $\gamma_{i+1} \cap (S \setminus \gamma_{i-1})$ on the boundary of h_{i-1max}. As a result, the null homologous curve $-\partial h_{i-1max}$ can only be constructed by alternately connecting arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ on the boundary of h_{i-1max} to the arcs $\{-a_k\}$.

c is therefore constructed by alternately connecting arcs of $\gamma_{i-1} \cap (S \setminus \gamma_{i+1})$ that are not homotopic to arcs on the boundary of h_{i-1max} to some subset

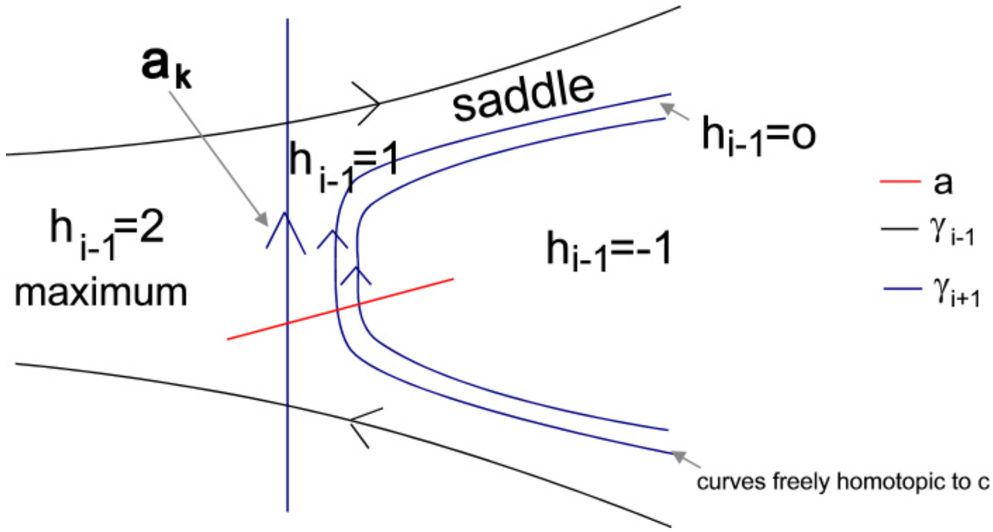

Figure 3.4: If there is no curve in γ_{i-1} freely homotopic to c, h_{i-1} has to to have a maximum greater than two or minimum less than zero.

of $\{a_k\}$, so c has to sit inside a component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ that is a saddle point. This contradicts the assumption that there is no curve in γ_{i-1} freely homotopic to c because otherwise, two curves in γ_{i+1} with the same orientation freely homotopic to c inside a saddle point of $S\setminus(\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ would force h_{i-1} to have a maximum greater than two, as shown in the diagram. It follows that γ_{i-1} has to contain a curve freely homotopic to c.

There remain three possibilities to consider; the curves in $\gamma_{i-1} \cup \gamma_{i+1}$ freely homotopic to c could be inside a component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ that is a maximum, minimum, or saddle. To start off with, consider the case in which the curves freely homotopic to c are contained in a minimum.

Inside this component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$, take a point p_1 at which h_{i-1} is equal to zero. Let p_2 be another point at which h_{i-1} is defined inside the same component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ as p_1 such that an arc connecting p_1 and p_2 crosses over one (and only one) of the curves in γ_{i+1} freely homotopic to c. It is also being assumed that all arcs are contained inside the one component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$. $h(p_2)$ is necessarily equal to one. Let p_3 be a point at which h_{i-1} is defined inside the same component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ as p_1 such that an arc connecting p_1 and p_3 crosses over one (and only one) of the curves in γ_{i+1} freely homotopic to c and the curve in γ_{i-1} freely homotopic to c. $h(p_3)$ is equal to zero. Let p_4 be a point at which h_{i-1} is defined inside

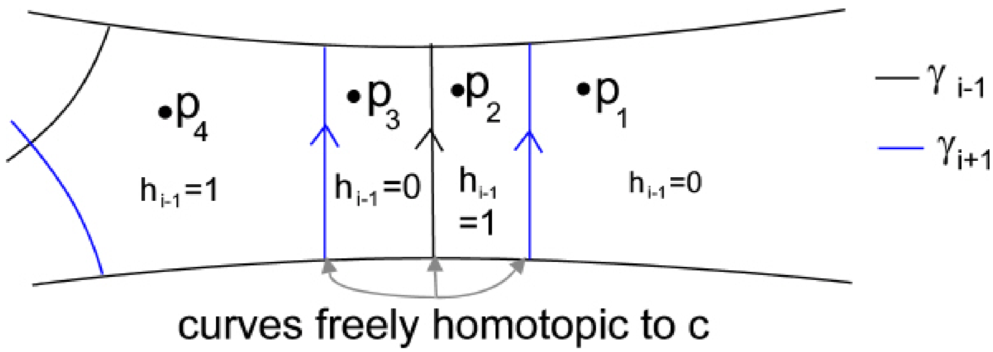

Figure 3.5: p_1, p_2, p_3 and p_4.

the same component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ as p_1 such that an arc connecting p_1 and p_4 crosses over every curve in $\gamma_{i+1} \cup \gamma_{i-1}$ freely homotopic to c. $h(p_4)$ is equal to one.

The subset of S on which h_{i-1} is at least equal to one therefore contains an annulus whose boundary curves are curves in γ_{i+1} and γ_{i-1} freely homotopic to c. The representatives of the free homotopy classes were chosen in such a way that the subset of S on which h_{i-1} is greater than or equal to one is the same as the subset of S on which f_{i-1} is equal to a least one less than its maximum. The subset on which f_{i-1} is at least one less than its maximum is the same as f_{imax}. It follows that f_{imax} contains a component consisting of an annulus with core curve c, as well as a component with an entire curve freely homotopic to c on its boundary. This is not possible, because γ_i is a multicurve and only contains one curve freely homotopic to c. It could happen that one of the curves freely homotopic to c on ∂f_{imax} is made up of a union of arcs of $c_1 \cap (S \setminus \gamma_i)$ and $\gamma_i \cap (S \setminus c_1)$, but then the third curve on f_{imax} freely homotopic to c could not be of this form unless γ_i had self intersections. In the diagram, the subset of the CW complex freely homotopic to c to the left can only have some arcs on the boundary of f_{imax}; the entire curve can't be on the boundary of the maximum.

If the curves in $\gamma_{i-1} \cup \gamma_{i+1}$ freely homotopic to c are inside a component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ that is a maximum, it follows that f_{i-1max} contains a component consisting of an annulus with core curve c, as well as a component with an entire curve freely homotopic to c on its boundary. This is impossible for the same reason as in the previous case.

The only other alternative is that the curves in $\gamma_{i-1} \cup \gamma_{i+1}$ freely homotopic

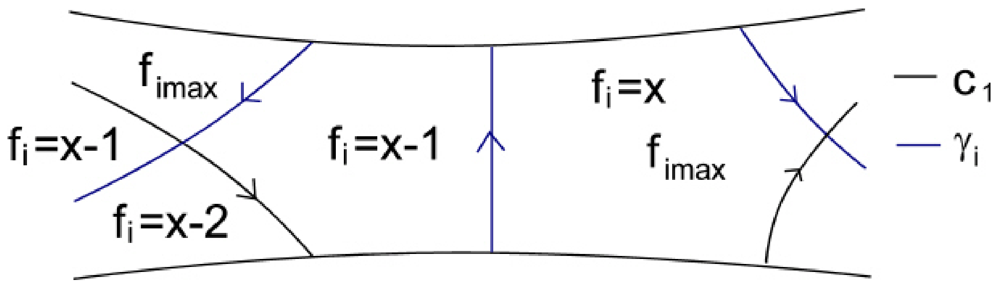

Figure 3.6: f_{imax} can't contain an annulus with core curve c in addition to a component with a curve freely homotopic to c on its boundary.

to c are inside a component of $S \setminus (\gamma_{i+1}^{int} \cup \gamma_{i-1}^{int})$ that is a saddle. Choosing the points p_1, p_2, p_3 and p_4 as before, h_{i-1} might increase or decrease when passing from p_1 to p_2, depending on the orientation of c. If h_{i+1} increases when passing from p_1 to p_2, f_{i-1max} contains a component consisting of an annulus with core curve c, as well as a component with an entire curve freely homotopic to c on its boundary. Otherwise, h_{i+1} decreases when passing from p_1 to p_2, in which case f_{imax} contains a component consisting of an annulus with core curve c, as well as a component with an entire curve freely homotopic to c on its boundary. As already discussed, neither of these two outcomes are possible. It follows that if γ_i doesn't have doubled curves, neither can γ_{i+1}.

□

If m is a multicurve homologous to and disjoint from c_1, lemma 15 requires that m contains a submulticurve homologous to c_1 that does not contain freely homotopic curves. However, if m intersects c_1, this is no longer true, as the next example shows. The problem seems to be that there could be a free homotopy class α with lots of elements in it that are all cancelled out by Dehn twists inside annuli with core curves whose union is homologous to $-\alpha$. If a sequence of homologous multicurves is constructed in some way other than that described in chapter one, it is not always possible to bound the number of curves in each of the multicurves by discarding null homologous submulticurves.

Example 45 (A multicurve m homologous to c_1 containing freely homotopic curves and no null homologous submulticurves)
In the diagram, c_1 is the curve drawn in blue, and m is the red curve. m

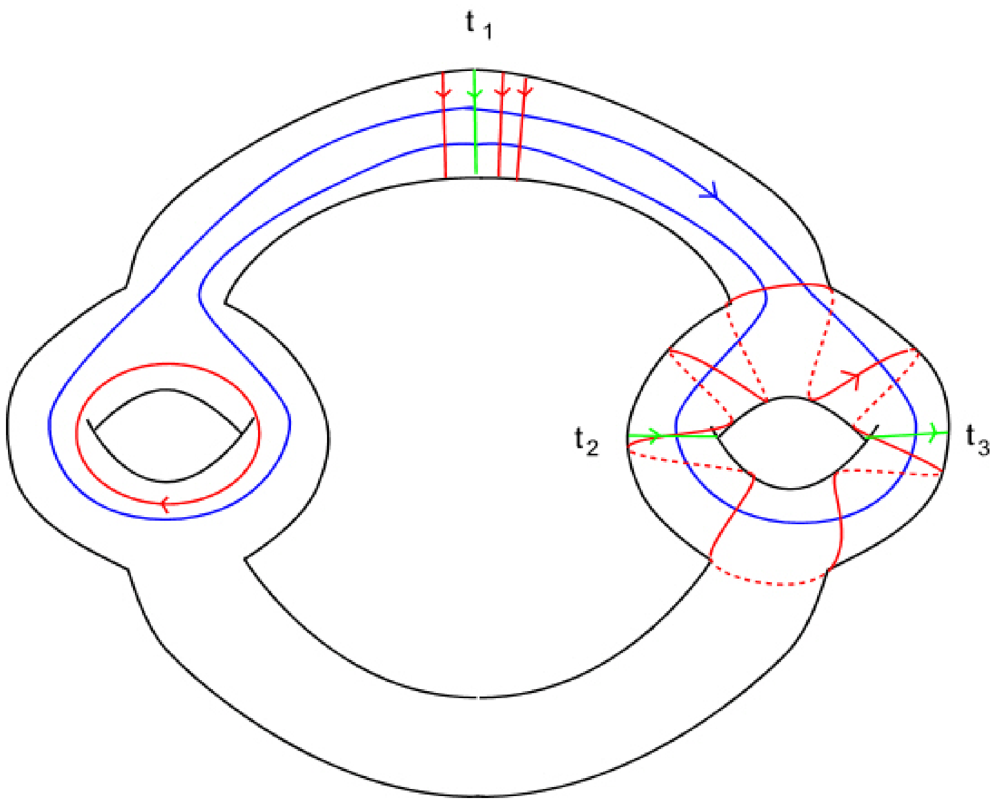

Figure 3.7: A multicurve m homologous to c_1 containing freely homotopic curves and no null homologous submulticurves.

is homologous to c_1 because the many curves in m freely homotopic to the curve t_1(shown in green) cancel out the Dehn twists inside the annuli with core curves t_2 and t_3(also shown in green).

Chapter 4

Twisting

4.1 Definition

The distance between two curves in the curve complex is bounded from above depending on the logarithm of the intersection number, as shown in [12], for example. The next example shows that the best possible bound on the distance in the homology multicurve graph depends linearly on the intersection number. In chapter two, an algorithm was devised to construct a sequence $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$, where $i(\gamma_{i+1}, c_1) + 2 \leq i(\gamma_i, c_1)$, so it follows that the distance between c_0 and c_1 in the homology multicurve graph is no more than $\frac{i(c_0,c_1)}{2} + 1$.

Example 46
Let c_0 and c_1 be the curves shown in figure 4.1. A simple calculation shows that the maximum of the function obtained from $c_1 - c_0$ is equal to $\frac{i(c_0,c_1)}{2} + 1$. In this case, it is also clear that the maximum of the function corresponds to the distance between c_1 and c_0 in the homology curve graph, because inside any annulus, it is only possible to perform one Dehn twist at a time when passing from γ_i to γ_{i+1}.

Example 47
c_1 and c_0 are the curves shown in figure 4.2. In this example, the distance in the curve graph is less than the maximum of the function obtained from $c_1 - c_0$. It is possible to construct a sequence $c_0, \alpha_1, \alpha_2, ...c_1$, where α_{i+1} is obtained from α_i by unwinding one twist in each of the four annuli shown, thereby decreasing the maximum of the function by two at each step. The

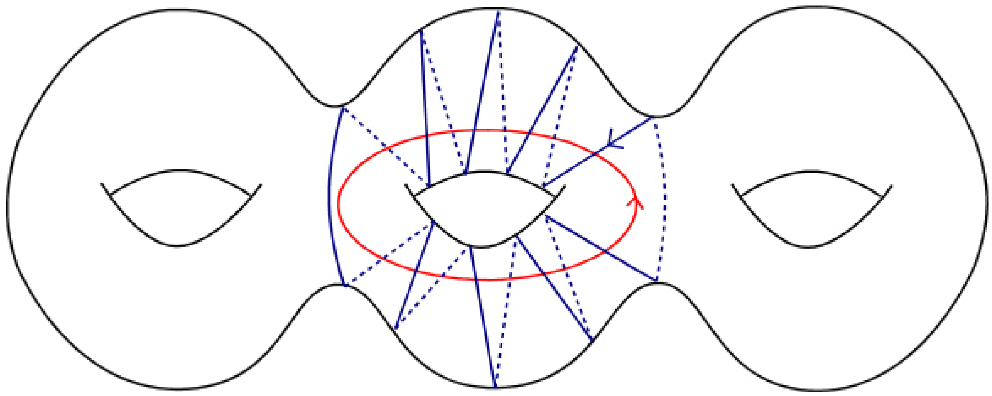

Figure 4.1: Example demonstrating that the best possible upper bound on the distance between c_0 and c_1 in the homology curve graph is given by $\frac{i(c_0,c_1)}{2}+1$.

sequence $c_0, \alpha_1, \alpha_2, ...c_1$ is not surface producing, however, because the null homologous multicurve consisting of the union of t_1, t_2, t_3 and t_4 does not bound a subset of S.

As shown in chapter two, the maximum of the function determines the number of multicurves in a surface producing sequence with endpoints c_0 and c_1 in the homology multicurve graph. Since not every sequence is surface producing, the distance between c_0 and c_1 in the homology multicurve graph can be smaller than this, as demonstrated in the previous example. However, the two concepts are not unrelated, as theorem 63 shows. When c_0 and c_1 are replaced by multicurves with freely homotopic curves, a further difference between the distance in the homology multicurve graph and the maximum of the function becomes apparent. Let nm_1 and nm_2 be the multicurves consisting of n copies of the homologous multicurves m_1 and m_2 respectively. The distance between nm_1 and nm_2 in the curve graph is the same as for m_1 and m_2, whereas the maximum of the function obtained from nm_1 and nm_2 is n times as large as the maximum of the function obtained from m_1 and m_2. This is related to the observation that the smallest genus surface with boundary curves $nm_1 - nm_2$ consists of n copies of a surface with boundary curves $m_1 - m_2$. A phenomenon called "twisting" will be defined. If m_1 is twisted with respect to m_2, then loosely speaking, in order to get to m_2 from m_1, many copies of a null homologous multicurve had to be added to

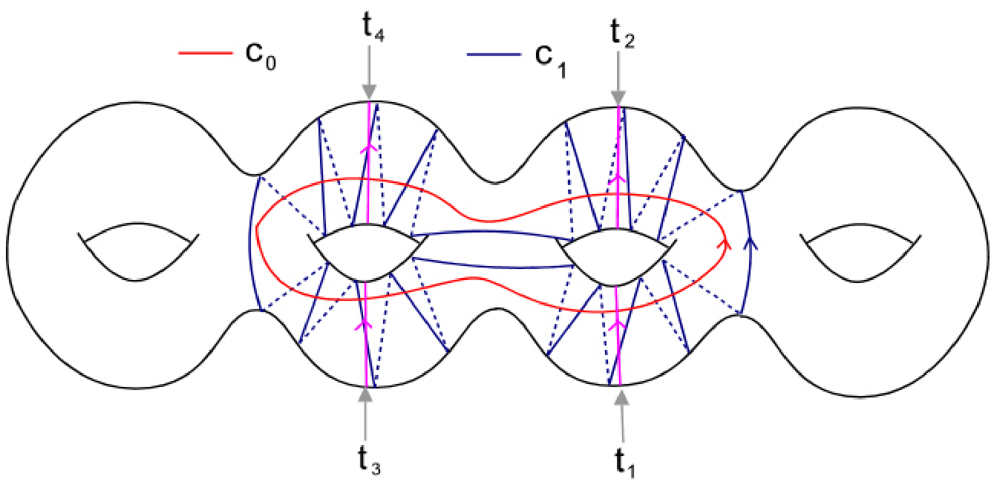

Figure 4.2: Example to show that the shortest arc in the homology curve graph with endpoints c_0 and c_1 is not always surface producing.

m_1. The distance between m_1 and m_2 in the curve complex does not depend on how many copies of the null homologous multicurve were added, but the distance in the homology multicurve graph can. If m_1 is twisted with respect to m_2 inside an annulus A with core curve α, when constructing a sequence $m_1, \gamma_1, ...\gamma_j, m_2$ in the curve graph it is possible to choose γ_1 such that γ_1 does not enter A, however this is not possible in the homology multicurve graph whenever m_1 has nonzero homology intersection number with α. In the next sections it will be shown that in the absence of twisting, a stronger bound on the distance in the homology multicurve graph in terms of the intersection number can be proven.

Remark 48
Unlike the curve graph, which is known to be hyperbolic ([18] and [4]), Example 47 can also be used to provide an example to show that the homology multicurve graph is not hyperbolic. Let v_1 be the curve c_0 from example 47, v_2 be c_0 Dehn twisted around t_1 and t_2 n times, and let v_3 be the curve c_0 Dehn twisted around t_3 and t_4 n times. v_1, v_2 and v_3 are the verticies of a triangle in the homology multicurve graph. For any fixed δ, n can be chosen large enough so that this triangle is not δ thin.

Suppose m_1 and m_2 are homologous multicurves without freely homotopic curves. In order to define twisting, it is necessary to make rigorous what is meant by the observation that a subarc of m_1 is obtained from a subarc of m_2 by Dehn twisting. In their paper [19] Masur and Minsky define a restriction $\mathcal{C}(Y)$ of the complex of curves, $\mathcal{C}(S)$, to a subset Y of S. A special case of this is when Y is an annulus A in S with incompressible boundary and core curve α. Let $\tilde{A}$ be the annular cover of S to which A lifts homeomorphically. There is a compactification of $\tilde{A}$ to a closed annulus $\hat{A}$ obtained in the same way as the usual compactification $\hat{S}$ of the universal cover $\tilde{S}$ of S. The vertices $\mathcal{C}_0(A)$ of $\mathcal{C}(A)$ are defined to be paths connecting the two boundary components of $\hat{A}$, modulo homotopies that fix the endpoints. Put an edge between any two elements of $\mathcal{C}_0(A)$ that have representatives with disjoint interiors. Distances d_A in $\mathcal{C}(A)$ are defined in the usual way by letting each edge have length one.

Let $\tilde{c}$ be the lift of a multicurve c on S to the covering space $\tilde{A}$, and let $\hat{c}$ be its lift to the compactification $\hat{A}$ of $\tilde{A}$. The reason for introducing $\hat{A}$ is that the lifts of curves have endpoints in $\hat{A}$, the existence of which follows from the fact that m_1 and m_2 are compact subsets of a hyperbolic surface.

Definition 49 (Nontrivial arcs)
The components of $\hat{m}_1$ and $\hat{m}_2$ that pass from one boundary component of $\hat{A}$ to the other, and the corresponding components of $\tilde{m}_1$ and $\tilde{m}_2$, will be called nontrivial arcs.

The Masur and Minsky definition of distance in a subsurface projection depends on a choice of free homotopy class of m_1 and m_2. In practice m_1 and m_2 are geodesics with respect to some hyperbolic metric on S. In this work, only the free homotopy class is important, so a small modification of the previous definition will be used that does not depend on the choice of representative of a free homotopy class.

Definition 50 (Distance d_A in the subsurface projection to an annulus A)
Suppose that $\tilde{m}_1$ and $\tilde{m}_2$ both have nontrivial arcs in $\tilde{A}$. $\hat{A}$ has two oriented boundary components, call them A_1 and A_2. Choose a point p_1 on A_1 and a point p_2 on A_2. Let $\hat{m}_1'$ be the image of $\hat{m}_1$ under a homotopy that slides the endpoints of $\hat{m}_1$ as far as possible along $\partial A \setminus (p_1 \cup p_2)$ in the direction given by the orientation of $\partial \hat{A}$, i.e. each component of $\hat{m}_1'$ has endpoints on p_1 and/or p_2. $\hat{m}_2'$ is defined analogously.

If there are nontrivial arcs of $\hat{m}_1'$ and $\hat{m}_2'$, the number of points of in-

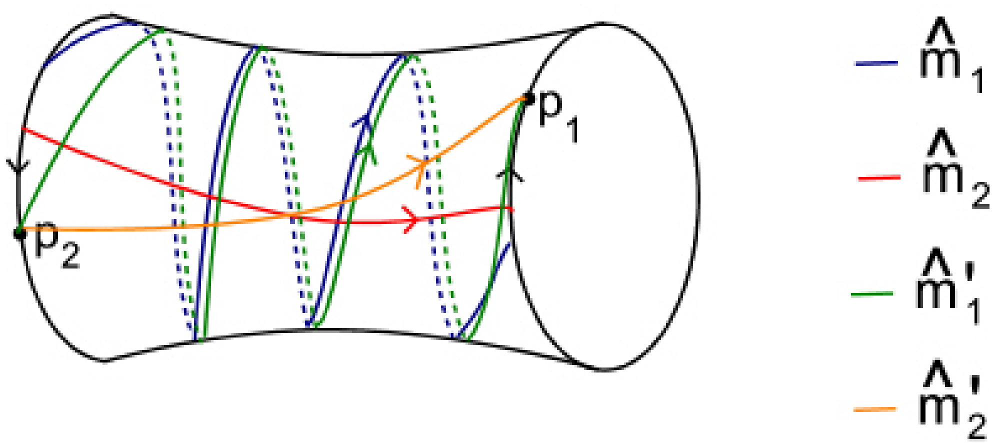

Figure 4.3: Construction of $\hat{m}_1'$ and $\hat{m}_2'$.

tersection between the nontrivial arcs (p_1 and p_2 are not counted as points of intersection), call it $i_A(m_1, m_2)$, depends on the choice of p_1 and p_2. As can be seen in the diagram, different choices of p_1 and p_2 could lead to a difference of at most two in the calculation of distance.

Choose p_1 and p_2 in such a way that $i_A(m_1, m_2)$ is minimised. With this choice of p_1 and p_2, let $d_A(m_1, m_2)$ be defined as $d_{\hat{A}}(\hat{m}_1', \hat{m}_2')$, where $d_{\hat{A}}(\hat{m}_1', \hat{m}_2')$ is the minimum distance in the subsurface projection, as defined by Masur and Minsky, between a component of $\hat{m}_1' \setminus (p_1 \cup p_2)$ and a component of $\hat{m}_2' \setminus (p_1 \cup p_2)$.

Definition 51 (κ-twisted)
m_1 is κ-twisted with respect to m_2 if there exists an annulus A in S with core curve α such that $d_A(m_1, m_2) \geq \kappa$. Alternatively, m_1 is said to be κ-twisted with respect to m_2 in A. Note that m_1 can't be twisted with respect to m_2 inside A if one or both of m_1 or m_2 does not intersect α, because then $d_A(m_1, m_2)$ is not defined.

It would be nice to be able to generalise the homology intersection number on S to an intersection number on $\tilde{m}_1$ and $\tilde{m}_2$. However, if the nontrivial components of $\tilde{m}_1$ and $\tilde{m}_2$ are considered separately from the other components, this intersection number is not homology invariant. If D_S is a fundamental domain of S in $\tilde{A}$, then (assuming each point of intersection on the boundary of D_S is counted once only) the number of points of intersec-

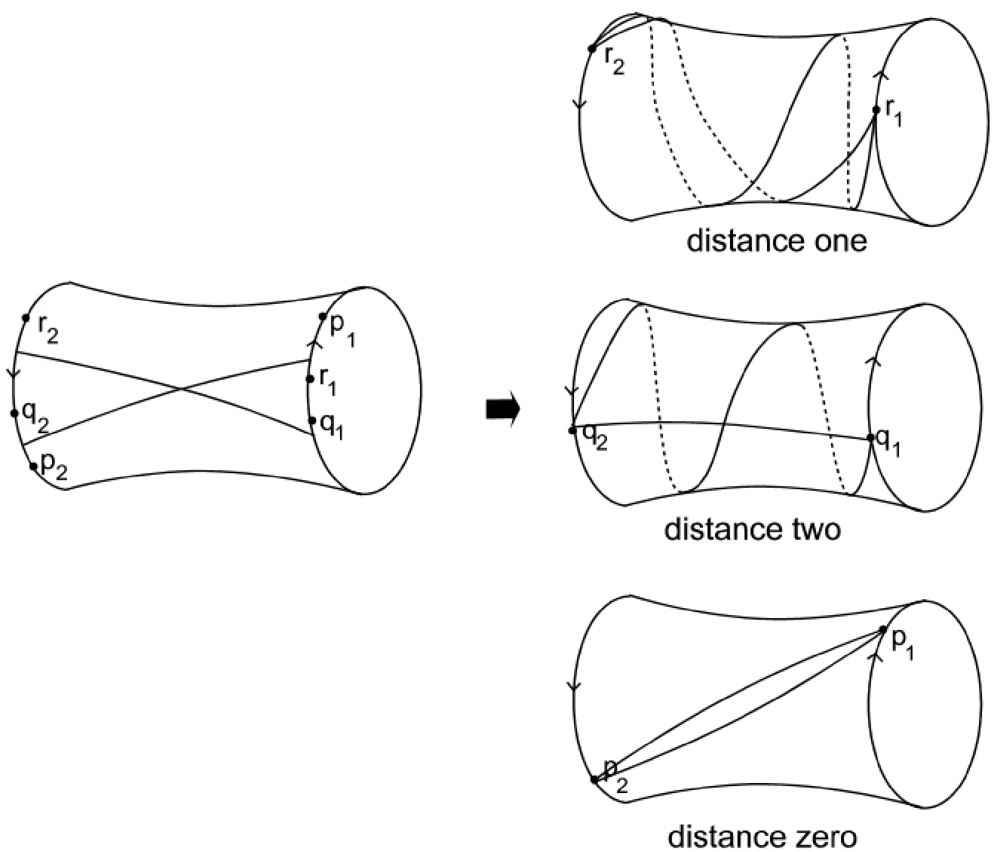

Figure 4.4: Different choices of p_1 and p_2

tion counted with orientation of $\tilde{m}_1 \cap D_S$ with $\tilde{m}_2 \cap D_S$ will be zero, because m_1 is homologous to m_2. The homology intersection number on S can be calculated by counting points of intersection with orientation, and given a metric on S or a choice of p_1 and p_2, this can be generalised to the nontrivial components of $\tilde{m}_1$ and $\tilde{m}_2$ in $\tilde{A}$.

Definition 52 ($i_{h,A}(m_1, m_2)$)
$i_{h,A}(m_1, m_2)$ is the intersection number of the nontrivial components of $\tilde{m}_1^{'}$ and $\tilde{m}_2^{'}$ counted with orientation. Only the absolute value of $i_{h,A}(m_1, m_2)$ will be needed here.

Given m_1 and m_2, in chapter two it was shown that it is possible to construct a multicurve homologus to m_1 that intersects m_2 less than m_1 by adding handles to m_1 corresponding to horizontal arcs of $m_2 \cap (S \setminus m_1)$. Since any multicurve homologous to m_2 will always have zero homology intersection number with m_2, this can be repeated until a multicurve homologous to and disjoint from m_2 is obtained.

Definition 53 (Twisting that cancels out in an annulus A)
Suppose that m_1 is twisted with respect to m_2 inside an annulus A. If the nontrivial arcs of $\tilde{m}_2$ are not all oriented in the same way, there exist horizontal arcs of $\tilde{m}_1 \cap (\tilde{A} \setminus \tilde{m}_2)$ with both endpoints on nontrivial arcs of $\tilde{m}_2$. Similarly, if the nontrivial arcs of $\tilde{m}_1$ are not all oriented in the same way, there exist horizontal arcs of $\tilde{m}_2 \cap (\tilde{A} \setminus \tilde{m}_1)$ with both endpoints on nontrivial arcs of $\tilde{m}_1$. By adding handles to m_1 and m_2 corresponding to these horizontal arcs it is possible to construct multicurves m_{1-} and m_{2-} homologous to m_1 and m_2 whose nontrivial arcs have fewer points of intersection. This can be repeated until eventually the multicurves m_{1--} and m_{2--} are obtained, where m_{1--} and m_{2--} have the property that the nontrivial arcs of $\tilde{m}_{1--}$ and $\tilde{m}_{2--}$ are all oriented in the same way. If $d_A(m_{1--}, m_{2--}) \leq 3$, the twisting inside A is of the type that cancels out in an annulus or it will be said to cancel out in A.

Definition 54 (κ-twisting that does not cancel out in A)
m_1 will be said to be κ-twisted with respect to m_2 of the type that does not cancel out in an annulus if m_{1--} is κ-twisted with respect to m_{2--} in A.

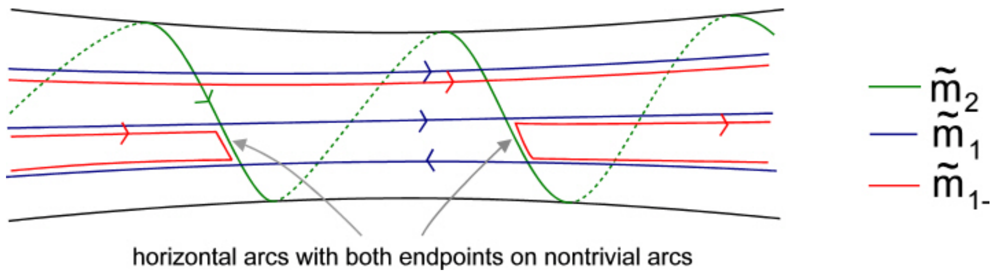

Figure 4.5: Construction of m_{1--} and m_{2--}.

Remark 55 (Convention regarding the orientation of α)
If α is the core curve of an annulus A in which m_1 is twisted with respect to m_2, whenever the twisting doesn't cancel out in A, α has nonzero homology intersection number with m_1 and m_2. This defines a "preferred orientation" on a nontrivial arc. Given a nontrivial arc a of $\tilde{m}_1$ oriented in the same way as the majority of the nontrivial arcs of $\tilde{m}_1$, the orientation of α determines the direction in which a has to be Dehn twisted in order to reduce the intersection number with m_2. For example, if m_1 is twisted with respect to m_2 inside an annulus with (oriented) core curve α, then m_2 is twisted with respect to m_1 in an annulus with core curve $-\alpha$.

To construct an example of twisting that cancels out inside an annulus with core curve α, let c be a curve that intersects α once, and $\alpha^{'}$ a curve homologous to α that intersects α. Let m_1 be the curve c Dehn twisted around α κ times, and let m_2 be the curve c Dehn twisted around $\alpha^{'}$ κ times. Since α is homologous to $\alpha^{'}$, $\alpha^{'}$ has zero homology intersection number with α, and in particular, $i_{h,A}(m_1, m_2) = 0$, (where A is, as usual, the annulus with core curve α). If $\alpha^{'}$ had been chosen so that it does not intersect α, then m_1 would still be twisted with respect to m_2, but this twisting would not be of the type that cancels out inside an annulus.

In this example, m_1 is twisted with respect to m_2 in A, where A is the annulus with core curve α, shown in green in the diagram. For simplicty, m_1 is only 2-twisted with respect to m_2 in A, but the number of twists can be clearly chosen to be arbitrarily large. m_1 and m_2 both have homology intersection number ± 1 with α. Inside $\tilde{A}$ there are many horizontal arcs of $\tilde{m}_1 \cap (S \setminus \tilde{m}_2)$, one of which is marked in gray in figure 4.6. Adding handles to m_2 corresponding to these arcs reduces the intersection number of m_2 with α, and in the process the twists inside the annuli with core curves t_1 and t_2

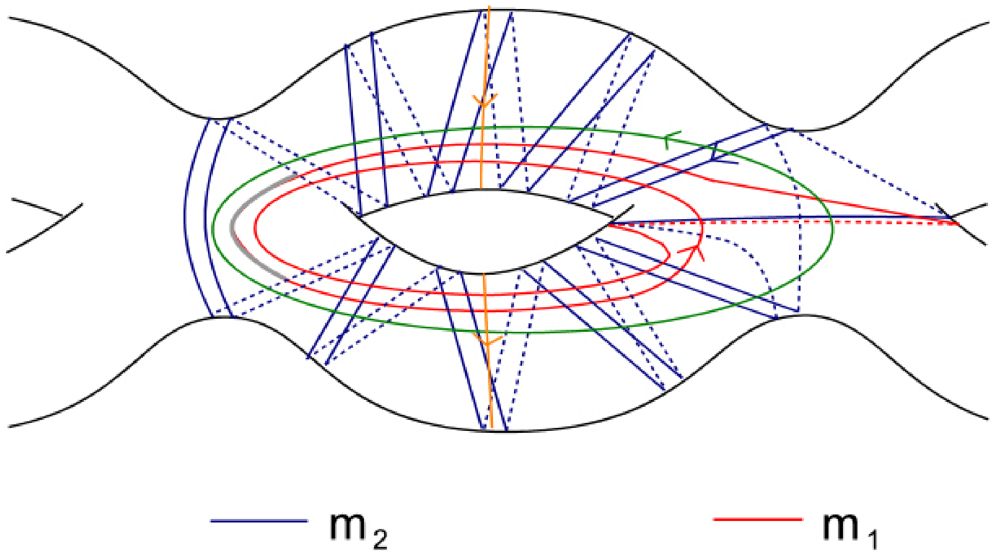

Figure 4.6: Example of twisting that cancels out in an annulus.

(the curves drawn in orange) are undone. Once the intersection number with α has been reduced in this way as much as possible, the resulting multicurve contains a curve freely homotopic to m_1 plus a union of null homologous curves.

When constructing a surface with boundary $m_1 - m_2$, it will be shown that the type of twisting that does not cancel out in an annulus is the main problem, so from now on, "twisting" will be taken to mean "twisting of the type that does not cancel out inside an annulus" unless otherwise stated. Twisting defines a marking on S, as the next lemma shows.

Lemma 56
If m_1 is twisted with respect to m_2 inside an annulus A with core curve α, whenever m_1 is also twisted with respect to m_2 inside an annulus B with core curve β and $i(\beta, \alpha) \neq 0$, it is not possible that m_1 is κ-twisted with respect to m_2 of the type that does not cancel out in an annulus in both A and B.

Proof. $m_1 - m_2$ is null homologous, but in general it is not a multicurve because m_1 and m_2 intersect. It will be assumed that m_1 and m_2 are in general position, and are representatives of the free homotopy classes $[m_1]$ and $[m_2]$ that only have essential points of intersection, so a null homologous

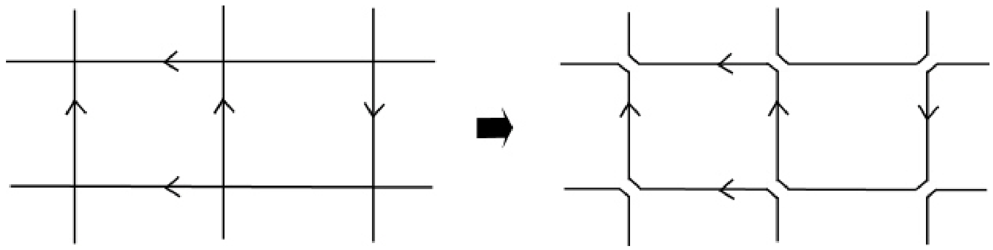

Figure 4.7: Resolving points of intersection.

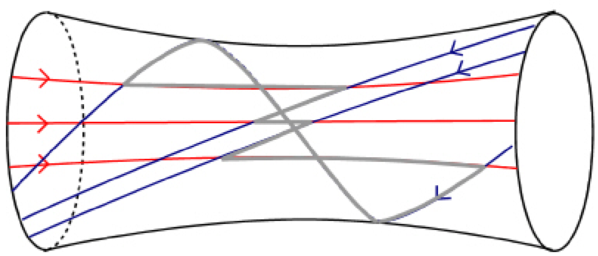

Figure 4.8: A curve in m_{1-2} freely homotopic to χ

multicurve n can be formed by cutting out the points of intersection and reconnecting the resulting arcs in such a way that the orientations match up. (This resolving of points of intersection is an example of attaching "bow tie" handles as defined in chapter two) Remove all contractible curves from n and call the resulting multicurve m_{1-2}. The various intersection numbers of m_1, m_2 or subarcs of m_1 or m_2 with m_{1-2} are defined by treating m_1, m_2, m_{1-2} and all subarcs of these multicurves as subcomplexes of the one dimensional cell complex $m_1 - m_2$ and counting the points of intersection of the subcomplexes according to definition 44.

m_{1-2} depends on the representatives of the free homotopy classes $[m_1]$ and $[m_2]$. The only assumptions being made is that these representatives are in general position and have the smallest possible intersection number. This choice of representatives therefore might not coincide with the choice of representatives used to calculate distances in subsurface projections.

What curves could be contained in m_{1-2} for any allowed choice of representatives of the free homotopy classes $[m_1]$ and $[m_2]$? If C is an annulus in S with core curve χ that has nonzero homology intersection number with m_1, all components of $m_1 \cap C$ and $m_2 \cap C$ are oriented in the same way (i.e.

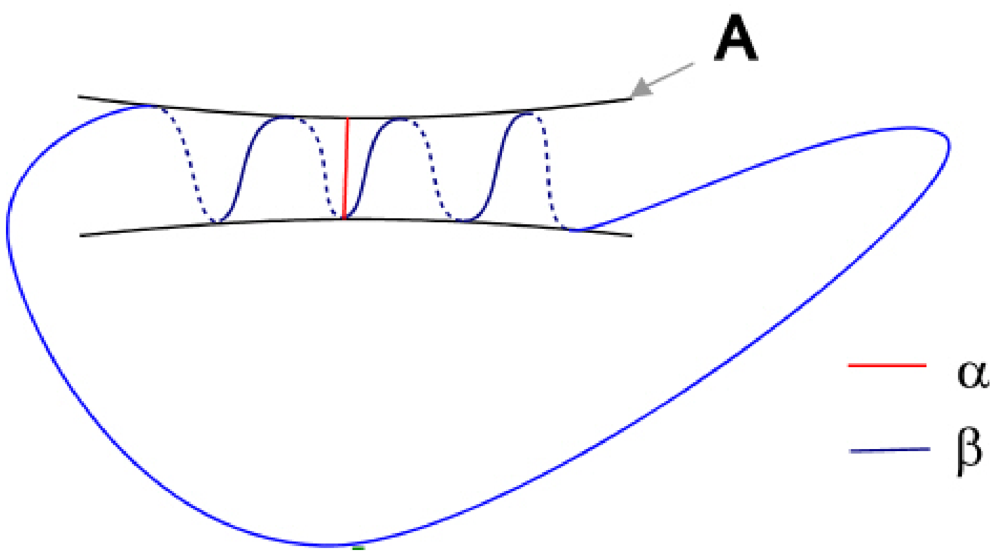

Figure 4.9: β is κ-twisted with respect to m_1 in A.

$i(m_1, \chi) = i_h(m_1, \chi)$ and $i(m_2, \chi) = i_h(m_2, \chi)$), there will be curves in m_{1-2} homologous to χ whenever $d_C(m_1, m_2)$ is large enough. In this case there will be a curve in m_{1-2} freely homotopic to χ as long as there is a nontrivial arc in $\tilde{m}_2$ that intersects a nontrivial arc of $\tilde{m}_1$ twice. If the intersection number of one or both of m_1 and m_2 with χ is much larger than its homology intersection number with χ, there might not be a curve freely homotopic to χ in m_{1-2}, even if the distance between m_1 and m_2 in the subsurface projection to C is large.

Suppose m_1 is κ-twisted with respect to m_2 inside A and B and that this twisting does not cancel out in either A or B. If $\alpha \cap (S \setminus m_1)$, $\alpha \cap (S \setminus m_2)$, $\beta \cap (S \setminus m_1)$ and $\beta \cap (S \setminus m_2)$ only consist of vertical arcs, it follows from the previous argument that this is a contradiction, because then the multicurve m_{1-2} would have to contain curves freely homotopic to both α and β, which it can't, since they intersect. Otherwise, to start off with, choose a metric that makes m_2 look "twisted" in A and B, while m_1 looks "straight". It is possible to assume w.l.o.g. that β is κ-twisted with respect to m_1 in A (the other alternative is that α could be κ-twisted in B with respect to m_1).

If the non-trivial components of $m_2 \cap B$ aren't all oriented in the same way construct m_{2--} as in the definition of "twisting that cancels out inside an annulus". m_{2--} will be at least κ-twisted with respect to m_1 in B by

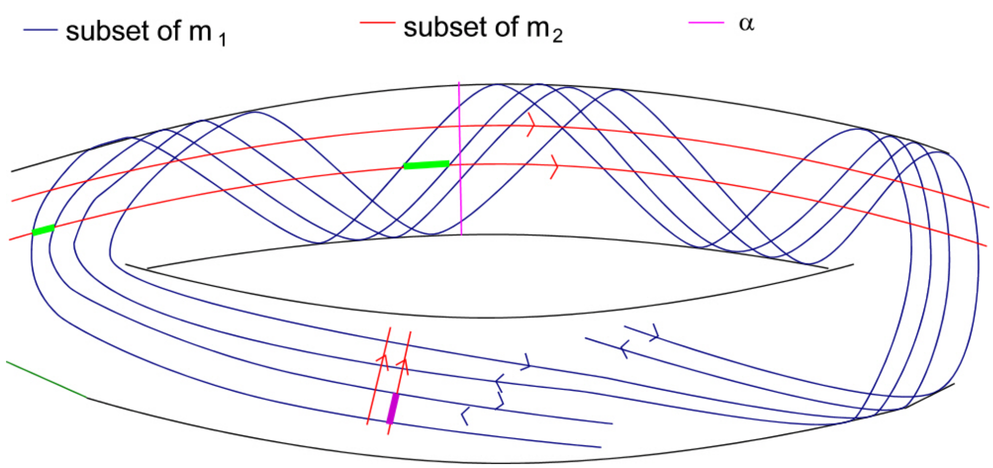

Figure 4.10: The arc drawn in purple is a horizontal arc of $m_2 \cap (S \setminus m_1)$ with both endpoints on nontrivial arcs of $\tilde{m}_1$. This arc is homotopic to the horizontal arcs of $m_2 \cap (S \setminus m_1)$ with both endpoints on nontrivial arcs of m_1 in A drawn in green.

assumption. Also, given two adjacent, nontrivial arcs of $\tilde{m}_2$ with opposite orientation (the tilde here refers to the lift to the annular covering space $\tilde{B}$ of S corresponding to B), a horizontal arc of $\tilde{m}_1 \cap (\tilde{B} \setminus \tilde{m}_2)$ with endpoints on these two arcs is homotopic to a horizontal arc of $m_1 \cap (S \setminus m_2)$ with endpoints of two nontrivial arcs of $m_2 \cap A$. Therefore m_{2--} will also be κ-twisted with respect to m_1 in A. It does not follow that $i_h(m_{2--}, \alpha) = i(m_{2--}, \alpha)$, because in the lift to $\tilde{A}$ there could be horizontal arcs of $m_1 \cap (S \setminus m_2)$ with endpoints on nontrivial arcs of m_2 in the lift to $\tilde{A}$ that are not homotopic to horizontal arcs of $\tilde{m}_1 \cap (\tilde{B} \setminus \tilde{m}_2)$.

Since twisting is a property that is symmetric in the two multicurves, the same argument with m_1 in place of m_2 and m_{2--} in place of m_1 shows that it is possible to construct m_{1--} such that all the nontrivial components of $\tilde{m}_{1--}$ are oriented in the same way and such that m_{1--} is κ-twisted with respect to m_{2--} in A and B. It follows that $i_h(m_{1--}, \alpha) = i(m_{1--}, \alpha)$, because if an arc of m_{1--} intersects α, it also intersects β.

Now choose a metric that makes m_{1--} look "twisted" in A and B, while m_{2--} looks "straight". Since all nontrivial components of $\tilde{m}_{2--}$ are oriented in the same way, it also follows that $i_h(m_{2--}, \alpha) = i(m_{2--}, \alpha)$. This is a contradiction, because by construction, it is also true that $i_h(m_{1--}, \beta) =$

$i(m_{1--}, \beta)$ and $i_h(m_{2--}, \beta) = i(m_{2--}, \beta)$, i.e. the null homologous multicurve constructed from $m_{1--} - m_{2--}$ would have to contain curves homologous to both α and β.

$\square$

4.2 Interval Exchange Maps

It has already been suggested that if the distance between m_1 and m_2 in the homology curve graph is large in comparison with $i(m_1, m_2)$, then m_1 is twisted with respect to m_2. Another way of looking at this is that if the maximum of f is large in comparison with $i(m_1, m_2)$, the reduction in the intersection number can't be large at each step. Given a surface producing sequence $m_1, \gamma_1, \gamma_2 ... \gamma_j, m_2$ constructed as in chapter two (only with the multicurves m_1 and m_2 in place of c_0 and c_1), the maximum possible reduction in intersection number with m_2 when passing from γ_i to γ_{i+1} is bounded above by twice the number of horizontal arcs of $m_2 \cap (S \setminus \gamma_i)$. Interval exchange maps will be used to investigate the connection between twisting and the proportion of horizontal arcs.

If each homotopy class of arcs of $m_1 \cap (S \setminus m_2)$ is represented by a letter, choosing a point in $m_1 \cap m_2$ as a starting point, any component of m_1 can be represented by a word. If m_1 is twisted with respect to m_2, a word representing a component of m_1 will have "syllables" that are repeated more than κ times. For example, m_1 might be represented by a word that looks like abcbcbcbcbcbcbcbcbcbefggg. However, repeated syllables is not a sufficient condition to ensure twisting, because two arcs that are homotopic might not be homotopic inside the annulus in question. The next definition will be used to make the concept "locally homotopic" precise.

Definition 57 (a-ladder)
From lemma 22 of chapter two it is clear that an arc in the oriented homotopy class a can't be homotopic to an arc in the oriented homotopy class $-a$. An "a-ladder" is a union of arcs of $m_1 \cap (S \setminus m_2)$ (or arcs of $m_2 \cap (S \setminus m_1)$) in the oriented homotopy class a that can be homotoped into each other with without crossing over an element of the homotopy class $-a$, and the subarcs of m_2 (or m_1) along which the endpoints of an arc have to be moved through by any such homotopy. A "long" ladder is a ladder with many steps.

An a-ladder looks like a ladder unless it is degenerate, see figure 4.11.

If m_1 and m_2 are assumed to be homologous, since the elements in a are all oriented in the same way, f is monotone on the ladder and increases by one for each step. If an a-ladder is degenerate, m_1 is automatically twisted with respect to m_2 if it is long enough. Whenever a is a homotopy class of arcs of m_1, the sides of an a-ladder consist of vertical arcs of $m_2 \cap (S \setminus m_1)$.

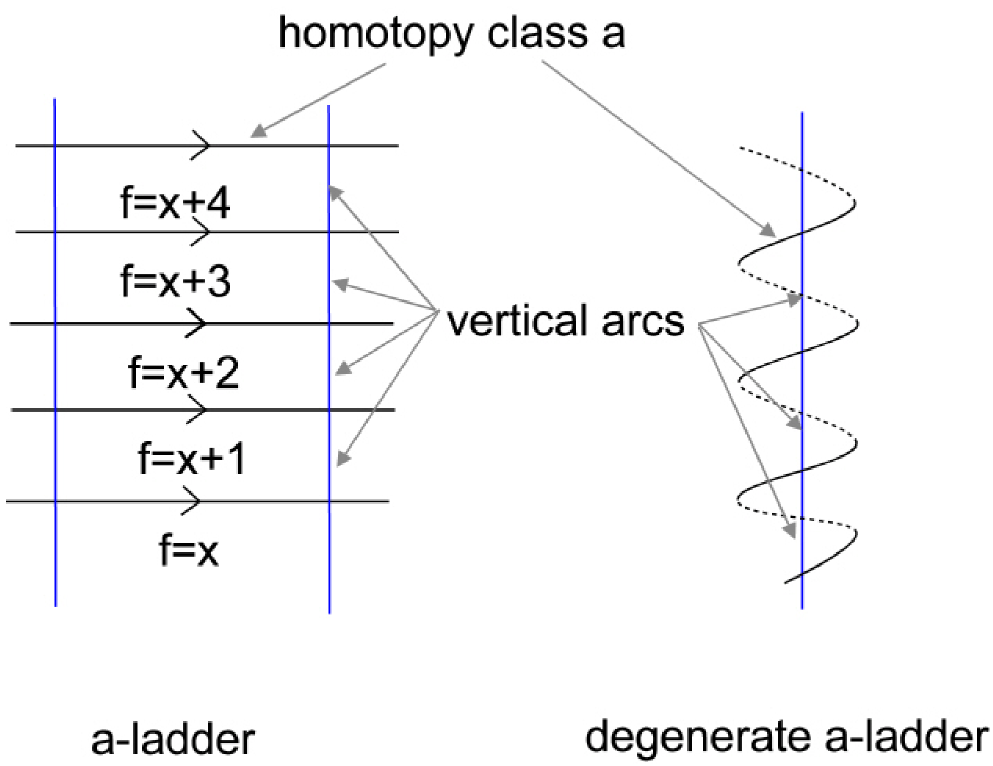

Figure 4.11: Examples of ladders.

If this a-ladder has n "steps", there has to be at least $2n-2$ vertical arcs of $m_2 \cap (S \setminus m_1)$ on its sides. If there don't exist long ladders, then either there aren't many representatives of any of the homotopy classes of arcs i.e. the intersection number is small, or there are many horizontal arcs ensuring that many arcs are adjacent to a homotopic arc with the opposite orientation. Since the converse is not true, the assumption that "all a-ladders are short" is a stronger condition than the assumption that the proportion of horizontal arcs is large.

Definition 58 (Interval Exchange Map, from [3])
Consider a compact interval I in R that is decomposed into a finite union $I = I_1 \cup I_2 \cup ... \cup I_n$ of intervals I_i with disjoint interiors. Choose another decomposition of $I = J_1 \cup J_2 \cup ... \cup J_n$ into intervals J_j with disjoint interiors such that, for every i, there is an isometry $\phi_i : I_i \to J_i$. The collection of the ϕ_i defines a 'map' $\phi : I \to I$. This map is in general 1-to-2 at the endpoints of the I_i, but is well-defined everywhere else. Such a ϕ is an interval exchange map.

m_1 and m_2 give rise to an interval exchange map, where I is a connected subarc of m_2 consisting of a union of vertical arcs of $m_2 \cap (S \setminus m_1)$. In other words, m_1 can only cross over I from left to right or right to left, but not both. I could be one side of an a-ladder, for example. For a fixed surface S, there is at most a bounded number of homotopy classes of arcs of $m_1 \cap (S \setminus I)$ relative to I. Along I, between any two oriented arcs in a particular homotopy class there can only be arcs from the same oriented homotopy class. An interval I_i is determined by the starting points of a homotopy class a_i of $m_1 \cap (S \setminus I)$. The homotopy class of m_1 can be chosen such that the length of an interval is proportional to the number of points of intersection of m_1 with the interval. Since the the function f obtained from $m_1 - m_2$ is single valued, the distance along I between the starting point and the endpoint of an arc a_i is proportional to its homology intersction number with m_2. ϕ maps the interval I_i to the interval J_i. The requirement that m_1 can't cross over I from both left to right and right to left is used here to ensure that all arcs of $m_1 \cap (S \setminus I)$ leave I from the one side and return to I on the other. If this were not so, the intervals I_i might not all have disjoint interiors. Similarly for the intervals J_i.

Before interval exchange maps can be used to prove a connection between long ladders and twisting, it is necessary to bound the number of homotopy

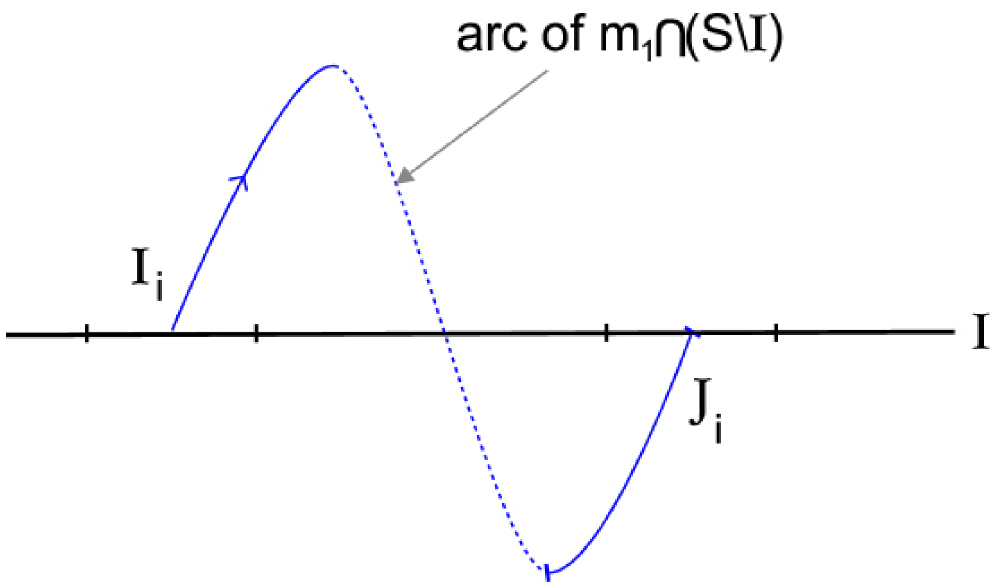

Figure 4.12: A homotopy class of arcs of m_1 determines how the subinterval I_i is mapped to J_i.

classes of arcs. This is done by showing that each homotopy class of arcs has a representative on the boundary of a component of $S \setminus (m_1 \cup m_2)$ that makes a nonzero contribution to $\chi(S)$.

Definition 59 (Nontrivial Component of $S \setminus (m_1 \cup m_2)$)
A nontrivial component of $S \setminus (m_1 - m_2)$ is any component of $S \setminus (m_1 - m_2)$ that is not a contractible rectangle, for example, a hexagon, annulus or pair of pants.

Lemma 60
The number of homotopy classes of arcs of $m_1 \cap (S \setminus m_2)$ is bounded above by $-3\chi(S)$, the number of homotopy classes of vertical arcs of $m_1 \cap (S \setminus m_2)$ is bounded above by $-2\chi(S)$.

Proof. To calculate the contribution of each component of $S \setminus (m_1 - m_2)$ to $\chi(S)$, $m_1 \cup m_2$ is treated as a subcomplex of a larger CW complex T consisting of a union of curves without triple points such that $S \setminus T$ consists only of contractible pieces. The contribution of a component of $S \setminus T$ is then

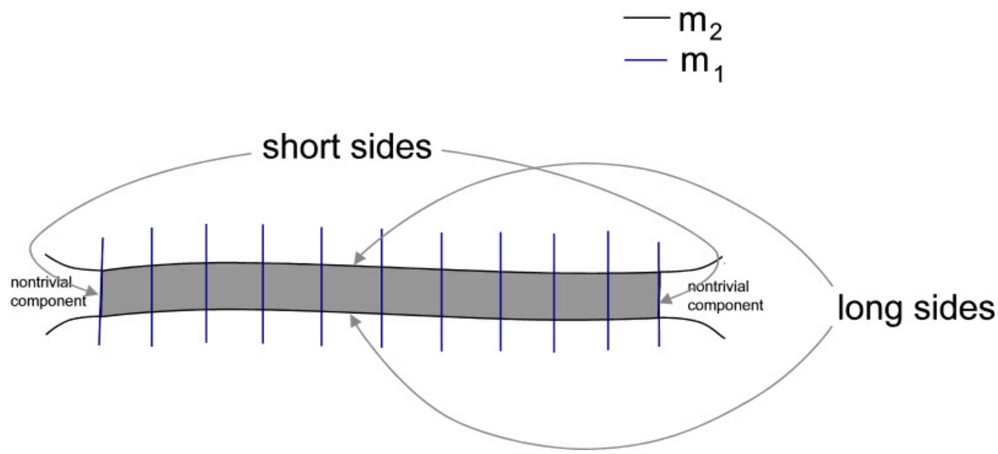

Figure 4.13: Ends of a homotopy class of arcs.

taken to be equal to

$$1 - \frac{E}{2} + \frac{V}{4} \tag{4.1}$$

where E is the number of edges and V the number of verticies. Each annulus has to contribute at least $-\frac{1}{2}$ to $\chi(S)$ because $m_1 \cup m_2$ does not contain freely homotopic curves, so no annulus can be a union of rectangles. It is also possible to check that any other nontrivial component of $S \setminus (m_1 \cup m_2)$ contributes at least $-\frac{1}{2}$ to $\chi(S)$, and that a trivial component does not contribute to $\chi(S)$.

A homotopy class of arcs of $m_1 \cap (S \setminus m_2)$ can be thought of as a rectangle on S; the two "short" sides are homotopic arcs of $m_1 \cap (S \setminus m_2)$, and the two "long" sides are arcs of m_2 along which the endpoints of an arc in the homotopy class has to be moved by homotopies that take it to all other representatives of the homotopy class. The two short sides of the rectangle will be called the "ends" of the homotopy class.

The ends of a homotopy class are arcs on the boundary of a nontrivial component of $S \setminus (m_1 \cup m_2)$, for example a hexagon. A contractible component of $S \setminus (m_1 \cup m_2)$ with n sides contributes $-\frac{n-4}{4}$ to $\chi(S)$; a non-contractible component of $S \setminus (m_1 \cup m_2)$ with n arcs of $m_1 \cap (S \setminus m_2)$ and $m_2 \cap (S \setminus m_1)$ on its boundary contributes even more than $-\frac{n-4}{4}$ to $\chi(S)$. The maximum number of homotopy classes of arcs of $m_1 \cap (S \setminus m_2)$ is achieved when the connected components of $S \setminus (m_1 \cup m_2)$ are all rectangles and hexagons,

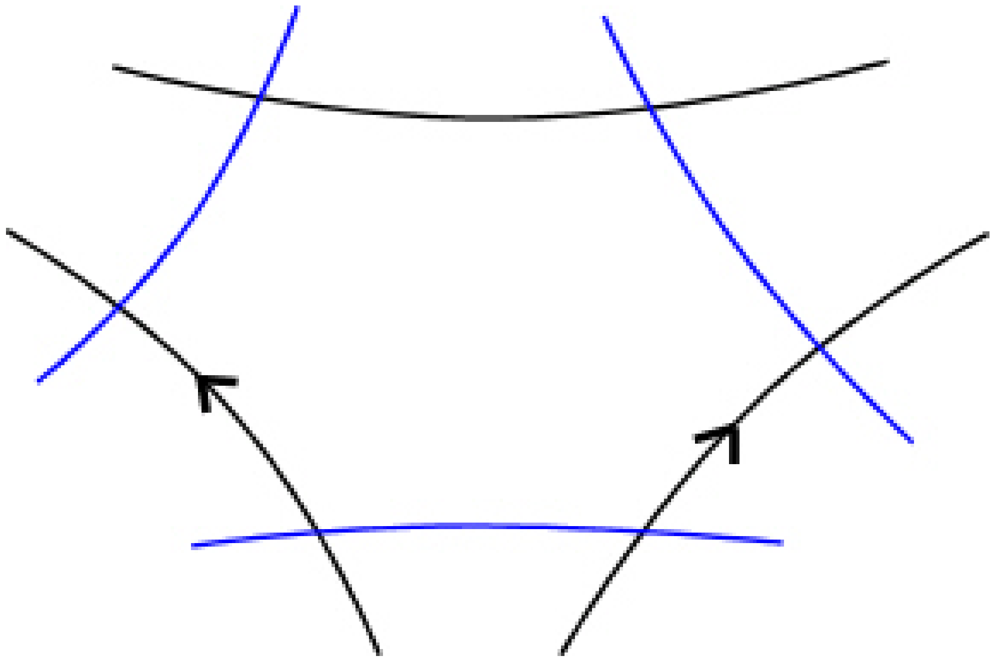

Figure 4.14: The two arcs shown in black are oriented in such a way that the blue arc between them is vertical. However the third black arc is oriented, one of the blue arcs has to be horizontal.

since there has to be some component that is not a rectangle, and apart from rectangles, hexagons have the largest number of edges for the smallest contribution to $\chi(S)$. In this case there are $-2\chi(S)$ hexagons amongst the components of $S \setminus (m_1 \cup m_2)$. Each hexagon has three arcs of $m_1 \cap (S \setminus m_2)$ on its boundary, and since each rectangle that represents a homotopy class of arcs of $m_1 \cap (S \setminus m_2)$ has two short sides, the given bound on the number of homotopy classes of arcs follows.

The boundary of a hexagon can't be made up of vertical arcs only; this has to do with the fact that the number of arcs of $m_1 \cap (S \setminus m_2)$ on its boundary is odd, as can easily be verified in figure 4.14. Therefore, if the connected components of $S \setminus (m_1 \cup m_2)$ are all either rectangles or hexagons, for every homotopy class of horizontal arcs there can be at most two homotopy classes of vertical arcs. If there is a connected component of $S \setminus (m_1 \cup m_2)$ that is not a rectangle or hexagon, then there might be more than two homotopy classes of vertical arcs for every homotopy class of horizontal arcs, but the number of homotopy classes of vertical arcs is still no more than $-2\chi(S)$ because the

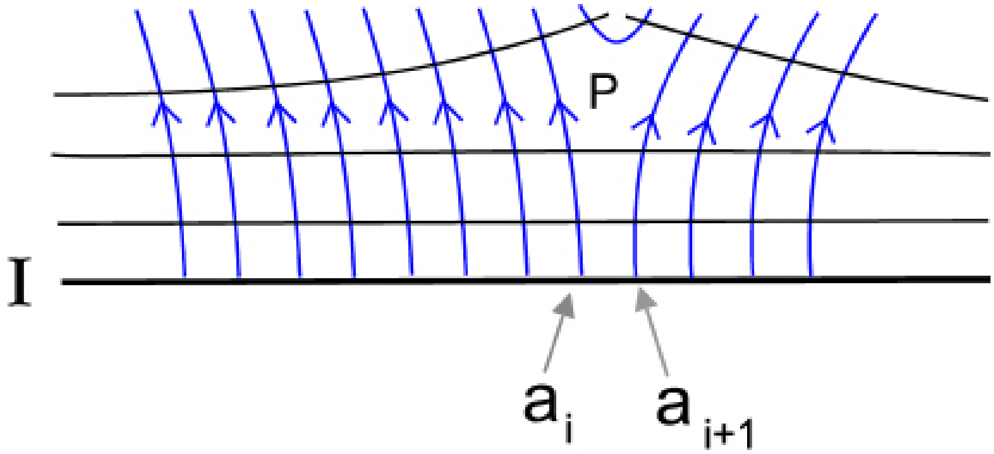

Figure 4.15: A nontrivial component P of $S \setminus (m_1 \cup m_2)$ sends two arcs of $m_1 \cap (S \setminus I)$ off in different directions.

total number of homotopy classes of arcs is correspondingly smaller. □

Lemma 61
Let v be equal to the number of homotopy classes of vertical arcs of $m_2 \cap (S \setminus m_1)$. Given an interval exchange map arising from a subinterval I of m_2, the number of homotopy classes of $m_1 \cap (S \setminus I)$, i.e. the number of intervals being exchanged by the interval exchange map, is less than or equal to $v + 1$.

Proof. Let a_1 and a_2 be two arcs of $m_1 \cap (S \setminus I)$ with adjacent starting points on I. If a_1 and a_2 are not homotopic, there has to be a non-rectangular component P of $S \setminus (m_1 \cup m_2)$ that sends the two arcs off in different directions.

The arcs a_1 and a_2 are oriented as subsets of m_1, and it follows from the definition of I that the sub-arcs of a_1 and a_2 that connect the starting points of a_1 and a_2 on I to the boundary of P are oriented in the same way. Using the terminology from lemma 60, there is a homotopy class of vertical arcs of $m_2 \cap (S \setminus m_1)$ that has an end on the boundary of P. Similarly, if a_1 and a_2 are arcs with adjacent endpoints on I that are not homotopic, a homotopy class of vertical arcs has an end on the boundary of the non-rectanglar component of $S \setminus (m_2 \cup m_1)$ at which the two arcs come together. Therefore, two ends of homotopy classes of vertical arcs of $m_2 \cap (S \setminus m_1)$ are needed to split a homotopy class of arcs of $m_1 \cap (S \setminus I)$ into two homotopy classes of arcs.

Since each homotopy class of vertical arcs of $m_2 \cap (S \setminus m_1)$ has two ends, the number of homotopy classes of $m_1 \cap (S \setminus I)$ is less than or equal to $v + 1$.

□

Definition 62 ($c \circ d$)
For oriented arcs c and d, with endpoint $d =$ starting point of c, $c \circ d$ is defined to be the arc formed by connecting the two arcs c and d at the endpoint of d.

Theorem 63
Let $c_0, \delta_1, \delta_2, ...\delta_k, c_1$ be an arc in the homology multicurve graph such that none of the δ_i contain freely homotopic curves. The shortest surface producing sequence, $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$ has length no more than $-3\chi(S)$ times the length of $c_0, \delta_1, \delta_2, ...\delta_k, c_1$.

Proof. Let f_i be the function obtained from $c_1 - \gamma_i$. Recall that γ_{i+1} is obtained from γ_i by attaching handles to γ_i corresponding to f_{imax}. In chapter two it was also shown that there are horizontal arcs $a_1, a_2...$ with endpoints on γ_i such that "attaching a handle corresponding to f_{imax}" results in the same multicurve as attaching handles to γ_i corresponding to the arcs $a_1, a_2...$ and discarding a null homologous multicurve freely homotopic to $-\partial f_{imax}$. In the case of δ_i, $\delta_{i+1} - \delta_i$ bounds a union of, possibly overlapping, subsets of S. δ_{i+1} can therefore also be constructed from δ_i by attaching handles corresponding to horizontal arcs and discarding null homologous submulticurves.

Let I be an oriented arc in S that intersects δ_i for some i. There are a certain number of homotopy classes of arcs of $\delta_i \cap (S \setminus I)$ relative to I. The orientations on I and δ_i makes it possible to define an ordering of the starting points of the arcs of $\delta_i \cap (S \setminus I)$ along I. Let h be a homotopy of δ_i that changes this ordering without moving any arcs over ∂I. Since δ_i does not contain freely homotopic curves, h has to introduce self intersections of δ_i. Similarly, if m also intersects I and is a representative of a multicurve chosen such that it only has essential points of intersection with δ_i, then any homotopy of δ_i and/or m that changes the ordering of the starting points of $\delta_i \cup m$ along I without moving any arcs over ∂I has to either create points of intersection or move one curve past another curve in the same free homotopy class.

Suppose δ'_{i+1} is obtained from δ_i by attaching handles corresponding to the horizontal arcs $a_1, a_2...a_k...a_n$ and at least one arc of the form $v_1 \circ a_k \circ v_2$

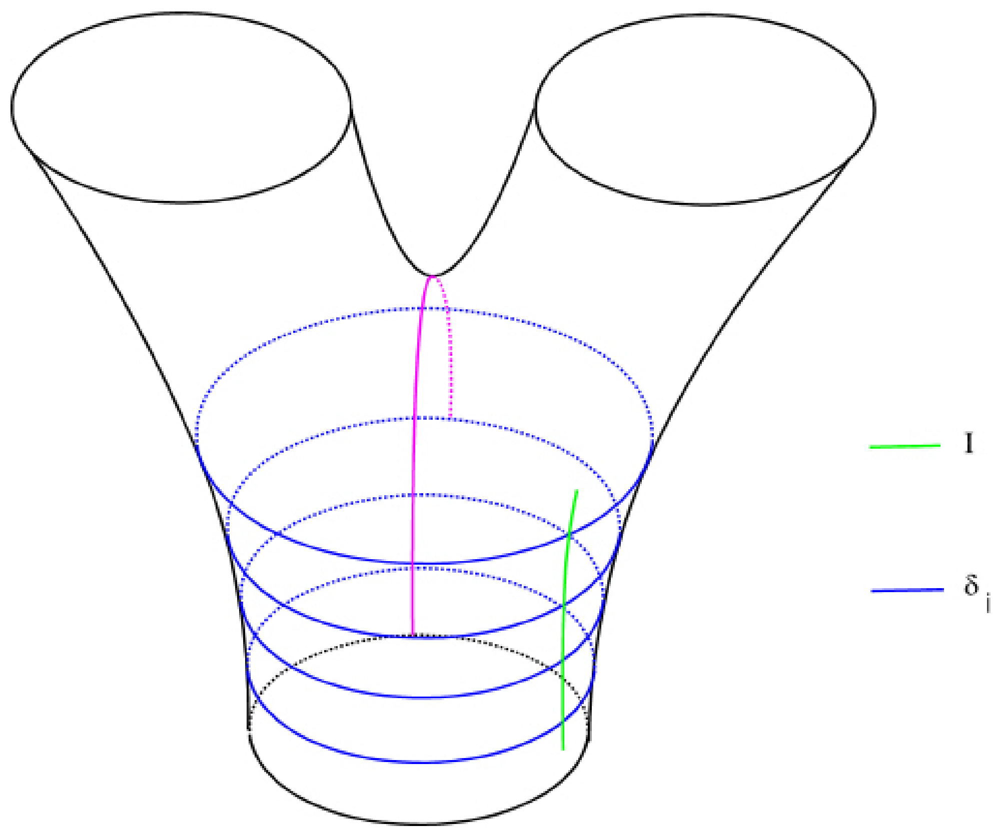

Figure 4.16: If δ_i could contain freely homotopic curves, the points of intersection of δ_i with the horizontal arc shown in pink can be removed by a homotopy that changes the ordering of the points of intersection of δ_i with the interval I, without creating points of self-intersection of δ_i.

for vertical arcs v_1 and v_2, where δ_{i+1} is obtained from δ'_{i+1} by discarding null homologous submulticurves. As usual, an arc with endpoints on δ_i is assumed not to be homotopic with fixed endpoints to a subarc of δ_i. Since the sequence $c_0, \delta_1, ...\delta_k, c_1$ is not required to be surface producing, it is possible to assume without loss of generality that none of the δ_i contain null homologous submulticurves, since otherwise it is possible to discard these submulticurves without increasing the number of multicurves in the sequence. In δ'_{i+1} there are one or two curves that were created by the surgery in which a handle corresponding to $v_1 \circ a_k \circ v_2$ was attached. δ_{i+1} has to contain at least one of these curves, otherwise there was no need to attach the handle corresponding to $v_1 \circ a_k \circ v_2$ at all, because the multicurve formed by attaching all the handles other than the handle corresponding to $v_1 \circ a_k \circ v_2$ would have to contain a null homologous submulticurve that could simply have been discarded without attaching a handle to it.

Call a curve in δ_{i+1} "new" if it was created by one of the surgeries in which δ_{i+1} is obtained from δ_i. Either

1. all new curves in δ_{i+1} are freely homotopic to other curves in δ_{i+1} i.e. δ_{i+1} contains doubled curves,
2. all new curves are homotopic to curves in δ_i, i.e. δ_{i+1} is a submulticurve of δ_i, or
3. neither 1 nor 2.

Let I be a compact arc in S chosen to pass through an arc in the homotopy class v_1 or v_2. In this third case, if the order of the arcs along I is altered to remove the points of intersection with δ_i of the attached handle corresponding to $v_1 \circ a_k \circ v_2$, it has to induce points of intersection elsewhere. In other words, δ_{i+1} has essential points of intersection with δ_i. Since δ_{i+1} is not a submulticurve of δ_i, and is not allowed to intersect δ_i or contain doubled curves, in order to obtain δ_{i+1} from δ_i it is only possible to attach handles corresponding to pairwise disjoint horizontal arcs with endpoints on δ_i contained in $S \setminus \delta_i$.

Attaching more than one handle to δ_i corresponding to arcs in the same homotopy class does not change the resulting multicurve up to homotopy. From lemma60, there are no more than $-3\chi(S)$ pairwise disjoint homotopy classes of horizontal arcs in $S \setminus \delta_i$ with endpoints on δ_i. Adding a handle to δ_i corresponding to each of these handles gives a multicurve with at most the number of curves in δ_i plus $-3\chi(S)$. Let $\tilde{\delta}_{i+1}$ be the multicurve obtained from δ_i by adding handles, such that δ_{i+1} is obtained from $\tilde{\delta}_{i+1}$ by discarding null homologous submulticurves. On average it therefore possible to discard no more than $-3\chi(S)$ curves when passing from $\tilde{\delta}_{i+1}$ to δ_{i+1}, i.e. on average, the maximum of the function obtained from $c_1 - \delta_i$ can't be more than $-3\chi(S)$ larger than the function obtained from $c_1 - \delta_{i+1}$. □

4.3 Existence of Twisting

In this section, interval exchange maps are used to show the existence of twisting whenever the proportion of horizontal arcs is sufficiently small. The small amount of ambiguity in the definition of twisting will be used to construct examples to show that a stronger result is not possible.

Theorem 64
If the proportion of horizontal arcs is less than $\frac{1}{\sqrt{-3\chi(S)(1-2\chi(S))\kappa i(m_1,m_2)}}$, m_2 has to be κ-twisted with respect to m_1. Also, the κ-twisting guaranteed by this theorem is of the type that does not cancel out in an annulus.

Proof. It follows from lemma 60 that there is at least one homotopy class of arcs of $m_1 \cap (S \setminus m_2)$ with at least

$$\frac{-i(m_1, m_2)}{3\chi(S)} \tag{4.2}$$

elements in it, and therefore there has to be a ladder L with length greater than $\sqrt{\frac{\kappa i(m_1,m_2)(1-2\chi(S))}{-3\chi(S)}}$ i.e.

$$\frac{-i(m_1, m_2)}{3\chi(S)} \tag{4.3}$$

divided by the number of horizontal arcs, which is by assumption less than

$$\sqrt{\frac{i(m_1, m_2)}{-3\chi(S)(1 - 2\chi(S))\kappa}} \tag{4.4}$$

Let I be one of the sides of the ladder L. By lemma 61, it follows that the number of intervals in the interval exchange map is less than or equal to $1 - 2\chi(S)$. Whenever the number of horizontal arcs is less than $\frac{1}{\kappa(1-2\chi(S))}$ multiplied by the number of steps in L, m_1 has to pass through each of the homotopy classes of $m_1 \cap (S \setminus I)$ on average more than κ times before passing through a horizontal arc. Therefore there is a subarc b of m_1 without horizontal arcs that passes through a homotopy class h_1 of $m_1 \cap (S \setminus I)$ more than κ times. Because b doesn't pass through any horizontal arcs, f is monotone on b, so once it has passed through an element of the homotopy class h_1 it can't pass through other homotopy classes and then come back to h_1 because the other homotopy classes are either above or below h_1 on the ladder. It follows that b has to pass through h_1 more than κ times in a row, i.e. m_1 is κ-twisted with respect to m_2.

Suppose that L has more than $\sqrt{\frac{\kappa i(m_1,m_2)(1-2\chi(S))}{-3\chi(S)}}$ steps. It remains to show that the twisting that has just been shown to exist can't cancel out in an annulus. Let α be freely homotopic to the curve obtained by connecting up the endpoints of h_1 by a subarc of I. None of the arcs $\alpha \cap (S \setminus m_1)$ can be horizontal, or, since I is a subarc of m_2, the ladder can not be very long. To be more specific, if one or more of the arcs of $\alpha \cap (S \setminus m_1)$ were horizontal, then L would have to have fewer than $\sqrt{i(m_1, m_2)}$ steps, because in this case the intersection number of each arc of $m_1 \cap (S \setminus I)$ is larger than the number of steps in the ladder.

None of the arcs of $\alpha \cap (S \setminus m_2)$ can be horizontal either, because then the homotopy class h_1 couldn't consist of vertical arcs only. Since all the nontrivial arcs of $m_1 \cap A$ are oriented in the same way and all the nontrivial arcs of $m_2 \cap A$ are oriented in the same way, the twisting can't cancel out in A.

□

From now on, "twisting" will be used to mean "twisting of the type that does not cancel out in an annulus".

In general, it is not possible to obtain a bound better than the square root in the previous theorem, as the next example shows.

Example 65
Let a_1 and a_2 be two homologous curves on S that intersect, and b_1 and b_2 be two homologous curves on S such that b_1 intersects a_1 once and b_2 intersects a_2 once. m_1 is the curve constructed as follows: Dehn twist b_1 inside the annulus with core curve a_1 n times, and call the resulting curve c_1. d_1 is the curve b_1 Dehn twisted around a_1 $n+1$ times. c_1 intersects d_1 once, so let e_1 be the curve d_1 Dehn twisted around c_1 n times, and let f_1 be the curve d_1 Dehn twisted around c_1 $n+1$ times. f_1 intersects e_1 once, so this can be repeated arbitrarily often. Let m_2 be the multicurve constructed in the same way as m_1 only with a_2 instead of a_1 and b_2 instead of b_1. m_1 is homologous to m_2 since a_1 is homologous to a_2 and b_1 is homologous to b_2. If n is chosen to be small, one for example, then m_1 is not twisted with respect to m_2.

In Example 65, the proportion of horizontal arcs and the length of ladders depends somehow on the square root of the intersection number. m_1 is homologous to $n_b b_1 + n_a a_1$, where n_b and n_a are large, relatively prime, and $n_b < n_a$. The number of horizontal arcs of $m_1 \cap (S \setminus m_2)$ increases linearly with n_a and n_b, whereas the total number of arcs depends on the square of n_a and n_b. To illustrate this point, consider the special case in which b_1 is freely homotopic to b_2. If there are h horizontal arcs and v vertical arcs amongst $a_2 \cap (S \setminus a_1)$, the number of horizontal arcs amongst $m_2 \cap (S \setminus m_1)$ is approximately hn_a, whereas the total number of arcs is approximately $(h+v)n_a^2$, as demonstrated in figure 4.17.

Another, more quantitative, way of describing what is happening in Example 2 is that there exists a ladder over an interval I with an associated interval exchange map ϕ and a sequence $I_2 \supset I_3 \supset I_4...$ of subintervals of I such that when ϕ is restricted to any of the intervals I_i the same (up to rescaling of the interval) interval exchange map is obtained. For example, suppose I is a subinterval of m_1 that makes up one side of a ladder, for which there are two homotopy classes of arcs of $m_2 \cap (S \setminus I)$, a_1 and b_1. Take N_{a_1} to be the number of arcs in the homotopy class a_1 and N_{b_1} to be the number of arcs in the homotopy class b_1, and let $n_{a_1} = \frac{N_{a_1}}{N_{a_1}+N_{b_1}}$ and $n_{b_1} = 1 - n_{a_1}$. Suppose for example that after m_2 has passed through an arc in the homotopy class b_1, it passes through at least two arcs in the homotopy class a_1 before returning to the homotopy class b_1.

If an arc in the homotopy class a_1 has its endpoint on I to the right of its starting point as in figure 4.18, then the first $N_{a_1} - 2N_{b_1}$ arcs of $m_2 \cap (S \setminus I)$, counting from left to right, pass through three arcs in the homotopy class a_1 before passing through an arc in the homotopy class b_1, all others pass through two arcs in the homotopy class a_1. Let I_1 be the subarc of I whose intersection number with m_2 is N_{b_1} and such that $m_2 \cap (S \setminus I_1)$ is an arc that either passes through two arcs in the homotopy class a_1 before passing through an arc in the homotopy class b_1 (call this homotopy class c_1) or it passes through three arcs in the homotopy class a_1 before passing through an arc in the homotopy class b_1 (call this homotopy class d_1). The condition that the interval exchange map associated with I_1, c_1 and d_1 is the same as the interval exchange map associated with I, a_1 and b_1 is therefore

$$\frac{N_{a_1} - 2N_{b_1}}{N_{b_1}} = \frac{N_{a_1}}{N_{a_1} + N_{b_1}} \tag{4.5}$$

and since $n_{a_1} + n_{b_1} = 1$, it follows that $n_{a_1} = -1 + \sqrt{3}$. I_2 and the homotopy classes e_1 and f_1 of arcs of $m_2 \cap (S \setminus I_2)$ are constructed from

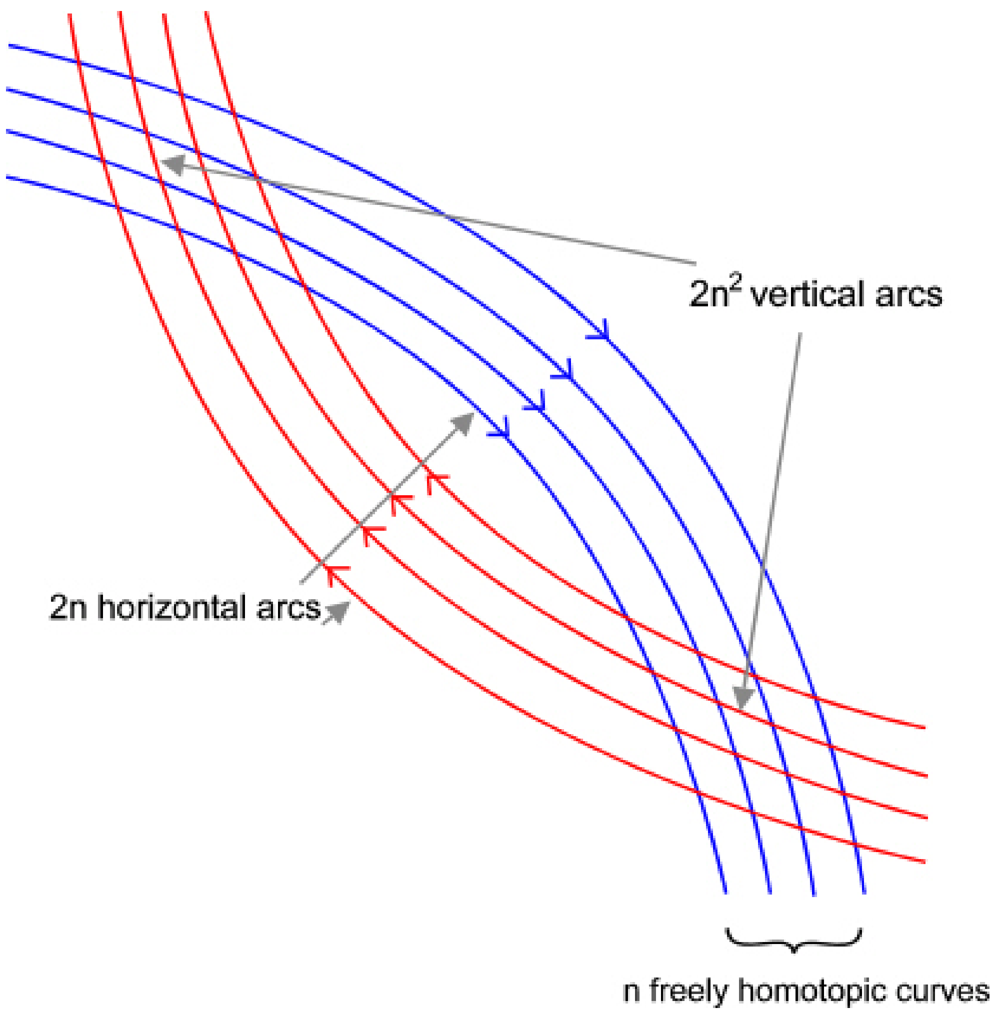

Figure 4.17: In this diagram, it can be seen that as the number of curves in a given free homotopy class is increased, the number of horizontal arcs increases linearly while the number of vertical arcs increases quadratically

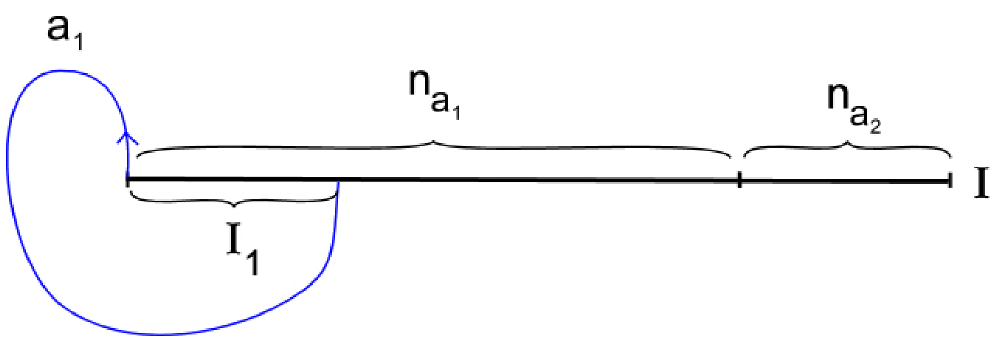

Figure 4.18: Interval exchange map with two intervals.

I_1 in the same way as I_1 from I, and the associated interval exchange map has to be the same as the interval exchange map associated with I_1, c_1 and d_2 for the same reason. Similarly for I_3, I_4 etc. The curves a_1, b_1... in Example 65 are related to the arcs a_1, b_1... by using subarcs of I to connect up the endpoints of the arcs a_1, b_1... It is easy to check that the curve b_1 constructed by connecting up the endpoints on I of the arc b_1 intersects the curve a_1 constructed by connecting up the endpoints of the arc a_1 once, as claimed. Similarly for c_1 and d_1, e_1 and f_1, etc.

As discussed at the beginning of this chapter, there is a small amount of ambiguity in defining distance in the subsurface projection to an annulus. This ambiguity was used in the previous example to construct a curve m_2 that is "almost twisted" with respect to m_1 and has a small proportion of horizontal arcs.

Chapter 5

Counting Horizontal Arcs

Suppose that $c_0, \gamma_1, ...\gamma_j, c_1$ is a geodesic in the homology curve graph constructed as in chapter two. Call h_i the function obtained from $\gamma_i - c_0$ and f_i the function obtained from $c_1 - \gamma_i$. It follows from theorems 43 and 64 that in the absence of twisting there exists a lower bound on the proportion of horizontal arcs of $c_1 \cap (S \setminus \gamma_i)$. However, $i(c_1, \gamma_i) = i(c_1, \gamma_{i+1}) + 2m$ where m is equal to the number of horizontal arcs of $c_1 \cap (S \setminus \gamma_i)$ homotopic to an arc on the boundary of the maximum of f_i. In general, not all horizontal arcs are homotopic to a horizontal arc on the boundary of the maximum or minimum of the function. The next theorem shows that it is possible to divide $c_0, \gamma_1, ...\gamma_j, c_1$ up into a uniformly bounded number of subarcs to which the theorems 43 and 64 can be applied to each subarc to directly obtain a bound on the reduction in the intersection number at each step.

Theorem 66
$c_0, \gamma_1, ...\gamma_j, c_1$ can be broken up into n subarcs such that for each subarc, $\gamma_l, \gamma_{l+1}..., \gamma_{m-1}, \gamma_m$, a horizontal arc of $\gamma_k \cap (S \setminus \gamma_l)$ to the right of γ_l is either homotopic to an arc on the boundary of the maximum of the function f_{lk} obtained from $\gamma_k - \gamma_l$ for all $l < k \leq m$, or its two endpoints are points of intersection that can be removed by a homotopy of γ_l. Also, n is bounded from above by $\frac{-11\chi(S)}{2} - 1$.

Proof. Once the representatives of the free homotopy classes of c_0 and c_1 in S have been fixed, recall that it was shown in chapter two that the algorithm for constructing the sequence $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$ uniquely determines a choice of representative of the free homotopy class of the multicurve γ_i for each i.

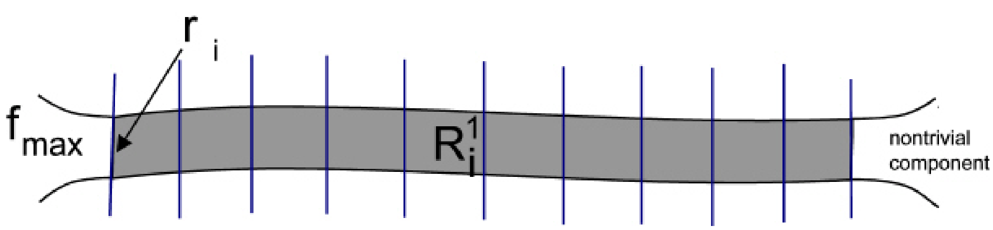

Figure 5.1: The rectangle representing a homotopy class of arcs.

This choice is made such that for each i, f_{imax} is the subset of S bounded by $\gamma_{i+1}-\gamma_i$. It was shown that this choice of representatives of the free homotopy classes are embedded, oriented subcomplexes of the one dimensional CW complex $c_0 \cup c_1$. The disadvantage of this choice is that as shown in diagram 2.10, the multicurves γ_i might have nonessential points of intersection with c_1 and that as subcomplexes, they aren't all pairwise in general position.

It is automatically true that every horizontal arc of $\gamma_2 \cap (S \setminus c_0)$ is homotopic to an arc on the boundary of either h_{2max} or h_{2min}. This is because $0 \leq h_2 \leq 2$, so a horizontal arc is either part of the boundary of a component of $S \setminus (c_0 \cup \gamma_2)$ on which h_2 equals one and part of the boundary of a component of $S \setminus (c_0 \cup \gamma_2)$ on which h_2 equals zero (h_{min}) or it is part of the boundary of a component of $S \setminus (c_0 \cup \gamma_2)$ on which h_2 equals two (h_{max}) and part of the boundary of a component of $S \setminus (c_0 \cup \gamma_2)$ on which h_2 equals one.

This proof involves defining "patches" consisting of unions of components of $S \setminus (c_0 \cup c_1)$. If every horizontal arc of $\gamma_i \cap (S \setminus c_0)$ to the right of c_0 is homotopic to an arc on the boundary of h_{imax} but not every horizontal arc of $\gamma_{i+1} \cap (S \setminus c_0)$ to the right of c_0 is homotopic to an arc on the boundary of h_{i+1max}, then it will be shown that the patch p_i has to be altered in some nontrivial way to obtain the patch p_{i+1}, and that this can't happen arbitrarily often.

In section 4.2, a nontrivial component of $S \setminus (c_1 - c_0)$ was defined to be any component of $S \setminus (c_1 - c_0)$ that is not a contractible rectangle, for example, a hexagon, annulus or pair of pants. A homotopy class of arcs of $c_1 \cap (S \setminus c_0)$ was treated as a rectangle with one pair of sides (the "short" sides) consisting of arcs of $c_1 \cap (S \setminus c_0)$ on the boundary of a nontrivial component of $S \setminus (c_0 \cup c_1)$, and the other pair of sides, (the "long" sides) subarcs of c_0 along which the endpoints of one short side of the rectangle have to be moved by a homotopy that takes it to the other side of the rectangle.

Let R^1_1, $R^1_2 ... R^1_a$ be the rectangles representing the homotopy classes of horizontal arcs $r_1, r_2 ... r_a$ of $c_1 \cap (S \setminus c_0)$ with at least one representative on ∂f_{max}. The first patch, p_1, is defined to be $f_{max} \cup R^1_1 \cup R^1_2 \cup ... R^1_a$. c_0 is altered inside p_1 to obtain the multicurve γ_1. By construction, γ_1 is the same subcomplex of $c_0 \cup c_1$ as c_0 outside of p_1, in other words $\gamma_1 \cap (S \setminus p_1) \subseteq c_0 \cap (S \setminus p_1) \subseteq \gamma_1 \cap (S \setminus p_1)$. γ_1 is taken to be the representative of its homotopy class chosen such that f_{max} is the subset of S bounded by the multicurve $\gamma_1 - c_0$. The points of intersection of γ_1 with c_1 in or on the boundary of p_1 can all be removed by a homotopy. With this choice of the representative of the free homotopy class of γ_i, let f_i be the function obtained from $c_1 - \gamma_i$. γ_2 is constructed from γ_1 in the same way as γ_1 from c_0, and is also chosen to be the representative of its free homotopy class such that f_{1max} is the subset of S bounded by $\gamma_2 - \gamma_1$. Similarly for γ_3, γ_4 etc.

$p_2 := f_{1max} \cup R^2_1 \cup R^2_2 \cup ... \cup p_1$, where each of the R^2_i are rectangles representing homotopy classes of horizontal arcs of $c_1 \cap (S \setminus \gamma_1)$ on the boundary of f_{1max}. p_2 is a union of components of $S \setminus (c_0 \cup c_1)$ chosen such that all the points of intersection of γ_2 with c_1 inside or on the boundary of p_2 can be removed by a homotopy.

$p_3 := f_{2max} \cup R^3_1 \cup R^3_2 \cup ... \cup p_2$, where each of the R^3_i are rectangles representing homotopy classes of horizontal arcs of $c_1 \cap (S \setminus \gamma_2)$ on the boundary of f_{2max}. Similarly for p_4, p_5, etc. Each multicurve γ_i is a subcomplex of $c_0 \cup c_1$, and so each patch is a union of components of $S \setminus (c_0 \cup c_1)$. Since the boundaries of the patches are fixed within their free homotopy classes, it makes sense to claim for example that a given patch contains another, even when the boundaries are freely homotopic.

An example is helpful at this point to clarify the purpose of the next part of the argument.

Example 67
Suppose c_0 and c_1 intersect on a subset of S as shown in the diagram.

Recall that, due to the choice of representatives of the free homotopy classes γ_1, h_{max} is the subset of S bounded by $\gamma_1 - c_0$, and so it is the same subset of S as f_{max}. h_{imax} is also the subset of S bounded by $\gamma_1 - c_0$. Similarly, the subset of S bounded by $\gamma_2 - c_0$ is the subset of S on which h_i is no less than its maximum value minus one, and coincides with the subset of S on which f is no less than its maximum value minus one, etc. The sequence $c_0, \gamma_1, \gamma_2 ... \gamma_j, c_1$ depends on the curves c_0 and c_1, however, it follows from the previous observation that if γ_i had been used instead of c_1 to construct the

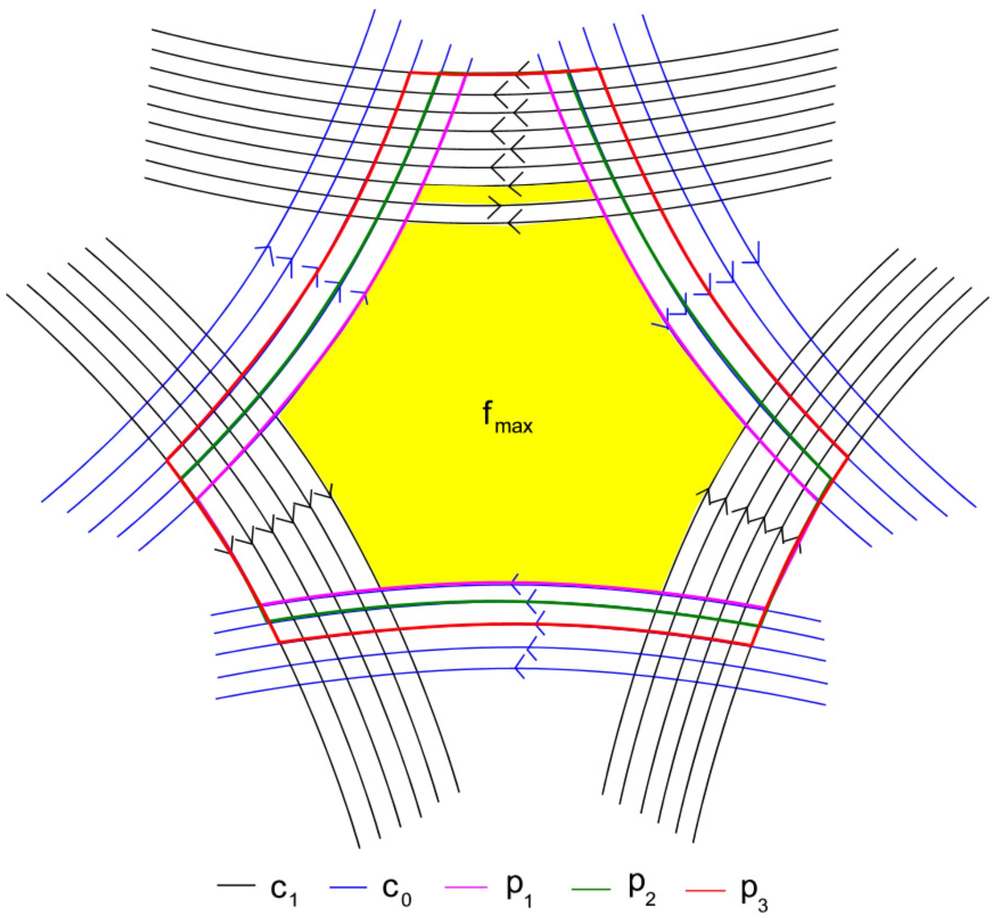

Figure 5.2: Various patches.

sequence, the same multicurves $\gamma_1, \gamma_2 ... \gamma_{i-1}$ would have been obtained. Let f_{lm} be the function obtained from $\gamma_m - \gamma_l$. f_{lmmax} is the same subset of S as f_{lmax}, so if the multicurves γ_l and γ_m are used in place of c_0 and c_1, the same multicurves γ_{l+1}, $\gamma_{l+2} ... \gamma_{m-1}$ are obtained.

h_{imax} also has the property that, if a component of $S \setminus (c_0 \cup c_1)$ has a boundary arc in common with h_{imax}, this component will be contained in h_{i+1max}. Also, if h_{i+1max} contains a component of $S \setminus (c_0 \cup c_1)$ that is not in or adjacent to h_{imax}, then this component has to be a local maximum of f. Considering all arcs shown in the diagram, every horizontal arc of $\gamma_4 \cap (S \setminus c_0)$ to the right of c_0 is homotopic to an arc on the boundary of h_{4max} (or f_{max}). The same is true for γ_3. This is necessarily the case because $p_4 \setminus f_{max}$ does not contain any nontrivial components of $S \setminus (c_0 - c_1)$, all local extrema of f inside p_4 other than f_{max} are rectangles with boundary arcs of $c_1 \cap (S \setminus c_0)$ that are homotopic to arcs on the boundary of f_{max}, and the boundary of p_4 is freely homotopic (both in S and in p_4) to the boundary of f_{max}.

It will be said that "the patch p_i has to be altered in a nontrivial way to obtain the patch p_{i+1}" if at least one of the following four possibilities occurs:

1) f_{i+1} has a maximum on a subset of S disjoint from p_i.

One advantage of using these patches instead of the sets f_{imax} for all possible values of i is that f_{imax} could have arbitrarily many components, but a patch can't. If f_{imax} has many components, then all but a uniformly bounded number of them will be contractible rectangles. It is only necessary to add a finite number of handles to γ_i to obtain γ_{i+1}. The number of components of p_i is equal to the number of handles that have to be added to γ_i to obtain a multicurve freely homotopic to γ_{i+1}, because it is only necessary to add at most one handle for each homotopy class of arcs with a representative on ∂f_{imax}.

2) p_{i+1} contains a nontrivial component of $S \setminus (c_0 \cup c_1)$ that p_i doesn't, as shown in figure 5.3, for example.

3) p_{i+1} "loops back on itself". p_{i+1} will be said to "loop back on itself" whenever p_{i+1} is obtained from p_i by attaching a handle to a component of p_i, as long as this handle does not come about from a nontrivial component of $S \setminus (c_0 \cup c_1)$ contained in p_{i+1} but not in p_i. In other words, when passing

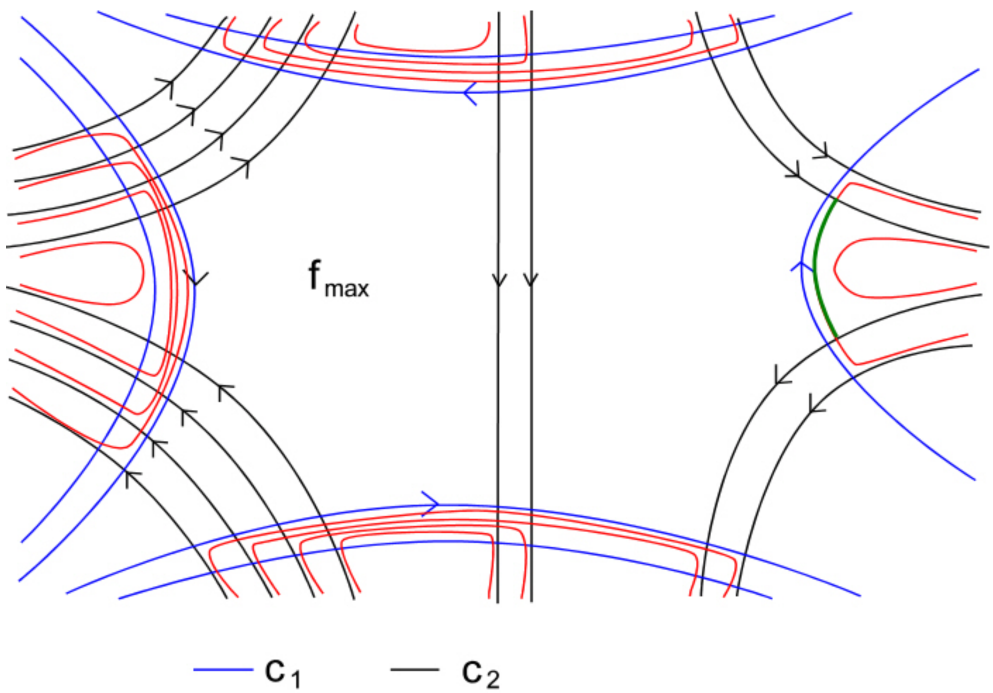

Figure 5.3: The arc marked in green is a horizontal arc of $\gamma_4 \cap (S \setminus c_0)$ to the right of c_0 that is not homotopic to an arc on the boundary of h_{4max}.

from p_i to p_{i+1}, the boundary of a component of p_i develops a new point of contact with itself.

If p_{i+1} loops back on itself there is a component of ∂p_{i+1} that is not homotopic in p_{i+1} to a component of ∂p_i. Suppose p_{i+1} loops back on itself and that b is a boundary component that is not freely homotopic in p_{i+1} to any boundary component of p_i. If b is contractible in S, it might happen that patches can loop back on themselves for arbitrarily many values of i, so it is necessary to investigate this possibility. Suppose b is contractible in S but not in p_{i+1}, and let d be the disk in $S \setminus p_{i+1}$ with boundary $-b$. Let $\gamma_{i+1}^{'}$ be a multicurve freely homotopic to γ_{i+1} that intersects c_1 as little as possible and does not contain any contractible curves. Then $\gamma_{i+1}^{'}$ can be chosen such that it does not enter the set p_{i+1}. Since it was chosen not to contain any contractible curves, it does not enter $p_{i+1} \cup d$ either. The same is true for $\gamma_{i+2}^{'}, \gamma_{i+3}^{'}...c_1$. If c_0 or γ_k for $k < i$ intersects γ_{i+1} inside or on the boundary of d, then these points of intersection can be removed by a homotopy.

If any of the multicurves $\gamma_{i+2}^{'}, \gamma_{i+3}^{'}...$ had been used in place of $\gamma_{i+2}, \gamma_{i+3}...$ to construct p_{i+1}, p_{i+1} would contain the disk d. The claims made in the statement of the theorem don't break down on the geodesic segment $c_0, \gamma_1...\gamma_{i+1}$ for this reason. If this is all that happens, any horizontal arc of $\gamma_{i+1} \cap (S \setminus c_0)$ to the right of c_0 that is not homotopic to an arc of the boundary of h_{max} can therefore be removed by a homotopy.

If b is not contractible in S, there is a pants decomposition of S containing the boundary curves of p_{i+1}. Whenever a patch loops back on itself and the new boundary curves are not homotopic in S to the boundary curves of the previous patch, the number of pants in the pants decomposition of the patch increases by one, so it is only possible for this to happen for at most $\frac{-3\chi(S)}{2}$ different values of i.

4)Two components of p_i come together when passing from p_i to p_{i+1}.

If none of these four possibilities occur for $l \leq m$ then by construction, the subarc has the properties claimed in the theorem. To calculate the bound, n, on the number of subarcs, it remains to count the number of times each of the four different possibilities could happen.

If a new component of f_{i+1max} appears it is either a nontrivial component of $S \setminus (c_0 \cup c_1)$ or it is a rectangle with a pair of sides consisting of horizontal arcs of $c_1 \cap (S \setminus c_0)$ to the right of c_0 that are not homotopic to arcs in an other component of p_i, i.e. there is at least one new nontrivial component of

$S \setminus (c_0 \cup c_1)$ on the boundary of the patch. There can be at most $-2\chi(S)$ nontrivial components of $S \setminus (c_0 \cup c_1)$. If neither 3) nor 4) happens, 1) and 2) together can occur at most $-2\chi(S) - 1$ times. 4) can only happen if a patch has more than one component, so the worst case scenario is that p_1 has one component, 2) happens as often as possible (namely $-2\chi(S) - 1$ times) and the last patch has only one component. As already discussed, 3) effectively only happens at most $\frac{-3\chi(S)}{2}$ times. The number n of subarcs is at most one more than the number of times one of the four possibilities could happen, so adding everything up gives the bound $\frac{-11\chi(S)}{2} - 1$ for n.

□

An example 47 was given to show that the shortest possible surface producing sequence with endpoints c_0 and c_1 is not always surface producing. The arguments given in the previous proof and in Theorem 63 make it seem plausible that the arc $c_0, \gamma_1, \gamma_2, ...\gamma_j, c_1$ might be piecewise geodesic in the homology curve graph, with each of the n subarcs $\gamma_l, \gamma_{l+1}, ...\gamma_m$ being a geodesic arc. The reason that this is not entirely clear is that γ_i might not contain a curve that passes through an arc on the boundary of f_{imax} and on the boundary of f_{imin}. In this case, it is possible to construct a shorter sequence by attaching handles to γ_i corresponding to f_{imax} and f_{imin}. This shorter sequence will not be surface producing, as discussed in chapter two.

Chapter 6

Calculating Bounds on Genus

In this chapter, the results of the previous sections will be combined to prove an upper bound on the genus of the surfaces constructed in chapter two. Recall that all twisting will be assumed to be of the type that does not cancel out in an annulus. From example 46, it is clear that if c_0 is twisted with respect to c_1, the existing bound of $\frac{i(c_0,c_1)}{2}$ on the number of multicurves in a surface producing sequence connecting c_0 and c_1 can't be improved upon. In chapter two it was shown that the smallest genus surface with boundary curves $c_1 - c_0$ is constructed via a surface producing surface. Without any restrictions on c_0 and c_1, the best possible bound on the genus of the surface with boundary $c_1 - c_0$ is therefore $\frac{i(c_0,c_1)g}{4} + \frac{g}{2}$, where g is the genus of S. The main result of this section is the following:

Theorem 68
Let c_0, c_1 and M be as defined. If c_0 is not κ-twisted with respect to c_1, there is a surface in M with boundary $c_1 - c_0$ with genus less than or equal to

$$\frac{gc}{2k}\left(\sqrt{\frac{i(c_0,c_1)}{c}} - 1\right) + \frac{g(c+1)}{2} \tag{6.1}$$

where $c = \frac{-11\chi(S)}{2} - 1$ and $k = \frac{1}{\sqrt{-3\chi(S)(1-2\chi(S))(\kappa+3)}}$

In order to prove theorem 68, it is necessary to show that the twisting restriction on c_0 and c_1 excludes the possibilty that γ_i could be twisted with respect to c_0 or c_1. This is the purpose of the next two lemmas.

Lemma 69
Let m_1, m_2 and m_3 be homologous multicurves. If m_1 is not κ-twisted with respect to m_2 and m_3 is λ-twisted with respect to m_2 in A, then m_3 is at least $(\lambda - \kappa)$-twisted with respect to m_1 in A.

Proof. If m_2 is λ-twisted with respect to m_3 inside an annulus A with oriented core curve α, since it is being assumed that the twisting is of the type that does not cancel out in A, it follows from the definition that m_2 has to have nonzero homology intersection number with α. The homology intersection number of m_1 with α is therefore also nonzero. The worst case scenario is when m_2 is $(\kappa - 1)$-twisted with respect to m_1 in the annulus with core curve α, i.e. taking the non trivial arcs of $\tilde{m}_2$ as a reference, m_1 is twisted with respect to m_2 in the same direction as m_3 is twisted with respect to m_2. Recall that the points p_1 and p_2 used in the definition of twisting are chosen in such a way as to give the smallest possible distance in the subsurface projection. The only difficulty in this lemma is that the choice of the points p_1 and p_2 used to define distance could be different when defining the distance between m_1 and m_2 and m_2 and m_3. Suppose p_1 and p_2 are the points chosen to define the distance between m_1 and m_2 in the subsurface projection to A. Homotope the endpoints of the nontrivial components of $\tilde{m}_1$, $\tilde{m}_2$ and $\tilde{m}_3$ onto p_1 and p_2 as described in the definition of distance in the subsurface projection, and call these arcs with fixed endpoints $\tilde{m}_1^p$, $\tilde{m}_2^p$ and $\tilde{m}_3^p$. Then $d_A(m_1, m_2)$ is equal to the smallest possible number of Dehn twists needed to be performed on a nontrivial arc of $\tilde{m}_1^p$ to obtain a nontrivial arc of $\tilde{m}_2^p$. It was also seen that a different convention for the choice of points p_1 and p_2 could affect the distance calculation by at most two. However, in this worst case scenario, the points p_1 and p_2 are already chosen such that $\tilde{m}_1^p$ is as close to $\tilde{m}_2^p$, and therefore $\tilde{m}_3^p$, as possible. Where p_1 and p_2 lie on the boundary of $\hat{A}$ in relation to the endpoints of $\hat{m}_3$ could only affect the distance between $\tilde{m}_1^p$ and $\tilde{m}_3^p$ by one. Therefore at least $\lambda - \kappa$ Dehn twists are needed to get from a component of $\tilde{m}_3^p$ to $\tilde{m}_1^p$, from which the lemma follows. □

Lemma 70
If c_0 is not κ-twisted with respect to c_1, the γ_i constructed as in chapter two can't be $\kappa + 3$-twisted with respect to c_0 or c_1.

Proof. The difficulty here is that even when c_0 is not twisted with respect to c_1, there can exist subcomplexes of $c_0 \cup c_1$ that are. It can happen that

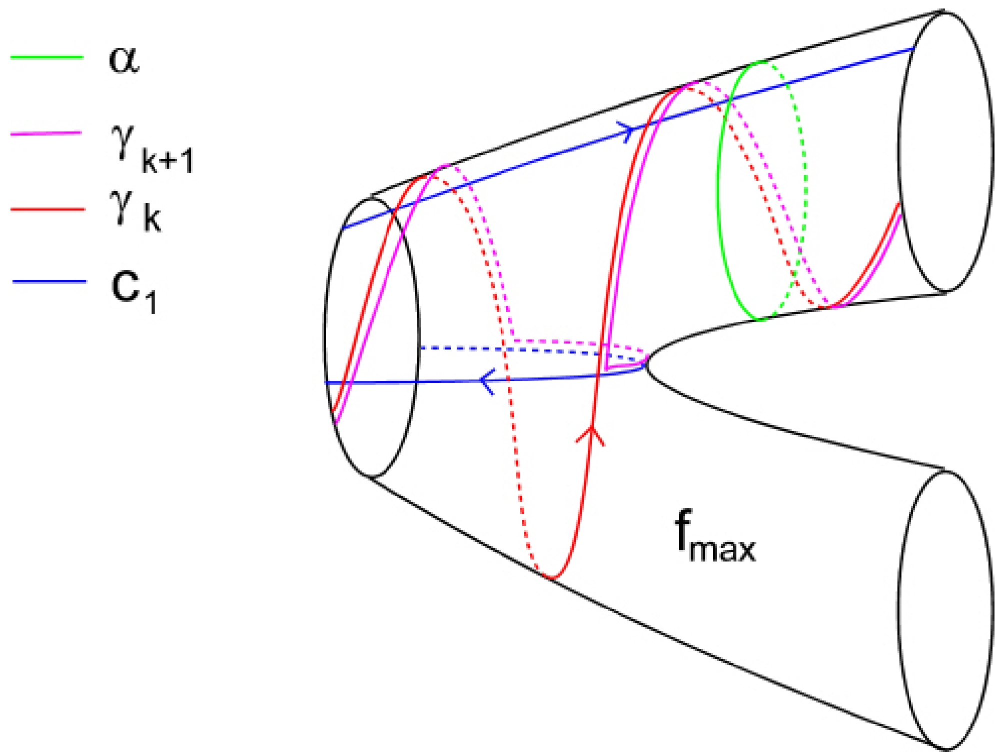

Figure 6.1: Distances in subsurface projections to annuli do not always decrease monotonically with intersection number.

attaching a handle to γ_k gives a multicurve whose intersection number with c_1 is less and whose distance from c_1 in the subsurface projection to an annulus with core curve α is larger than that of γ_k, as shown in the diagram.

In this proof, the representatives of the free homotopy classes will be chosen as in the previous chapter. Let f_k be the function obtained from $c_1 - \gamma_k$.

Suppose γ_i is $(\kappa+3)$-twisted with respect to c_0 in an annulus A with core curve α. Suppose also that $i(c_0, \alpha) = i(c_1, \alpha) = |i_h(c_0, \alpha)|$, in other words, in the lift to the covering space $\tilde{A}$, all the nontrivial arcs of $\tilde{c}_0$ and $\tilde{c}_1$ are oriented in the same way. In this case, the nontrivial arcs of each of the $\tilde{\gamma}_k$ are also all oriented in the same way. This is because otherwise f_k would have to have some local extremum along α that f does not have. This is not possible, because f_k is equal to f outside of f_{kmax} and is constant inside f_{kmax}.

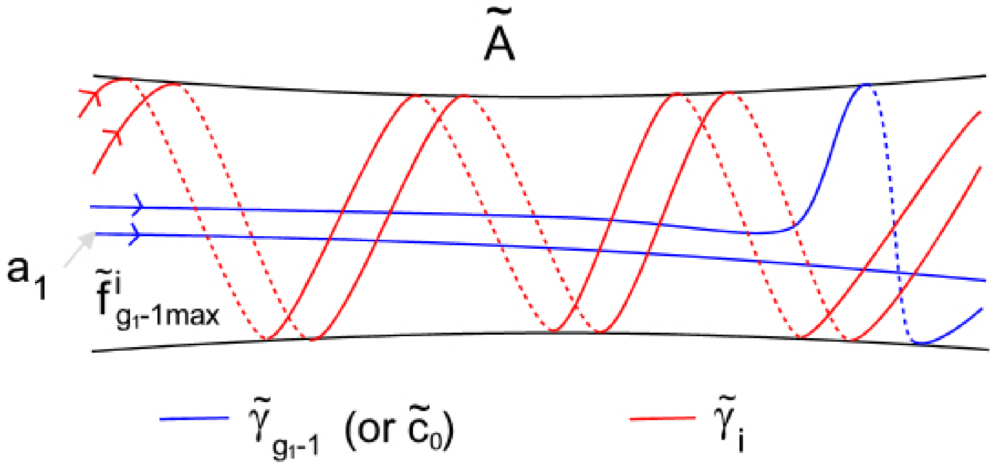

Figure 6.2: The arc a_1.

For this proof it is important that for any k, γ_{k+1} restricted to the closure of f_{kmax} consists of a union of subarcs of c_1, and coincides with c_0 elsewhere. This follows from the choice of the representatives of the free homotopy classes.

As discussed elsewhere, for $g_1 < i$, if γ_i is used in place of c_1 to construct γ_{g_1} from γ_{g_1-1}, the same multicurve would have been obtained. Let f_k^i be the function obtained from $\gamma_i - \gamma_k$. For some $g_1 < i$, the distance between γ_{g_1} and γ_i in the subsurface projection to A has to become one less than the distance between c_0 and γ_i in the subsurface projection to A. For this to happen, f_{g_1-1max} (or, equivalently $f^i_{g_1-1max}$) has to "enter" the annulus A. To be more precise, the lift to $\tilde{A}$ of f_{g_1-1max} has to have an arc a_1 of $\tilde{\gamma}_{g_1-1} \cap (S \setminus \tilde{c}_1)$ as part of its boundary, where a_1 is an arc contained in a nontrivial component of $\tilde{\gamma}_{g_1-1}$.

Recall that γ_{g_1-1} is a subcomplex of $c_0 \cup c_1$, and since all arcs of $\gamma_{g_1-1} \cap (S \setminus c_1)$ on the boundary of f_{g_1-1max} are subarcs of c_0, a_1 is also a subarc of $\tilde{c}_0$. If the distance between γ_{g_1} and γ_i in the subsurface projection to A is to be decreased further, it is necessary to Dehn twist each of the nontrivial components of $\tilde{\gamma}_{g_1}$ once before coming back to "the same" nontrivial component and twisting it a second time. Let $g_1 < g_2 < i$ be as small as possible such that γ_{g_2} is one unit closer than γ_{g_1} to γ_i in the subsurface projection to A. Then the lift of f_{g_2-1max} has to have an arc a_2 of $\tilde{\gamma}_{g_2-1}$ on its boundary, where a_2 is also a subarc of the same nontrivial

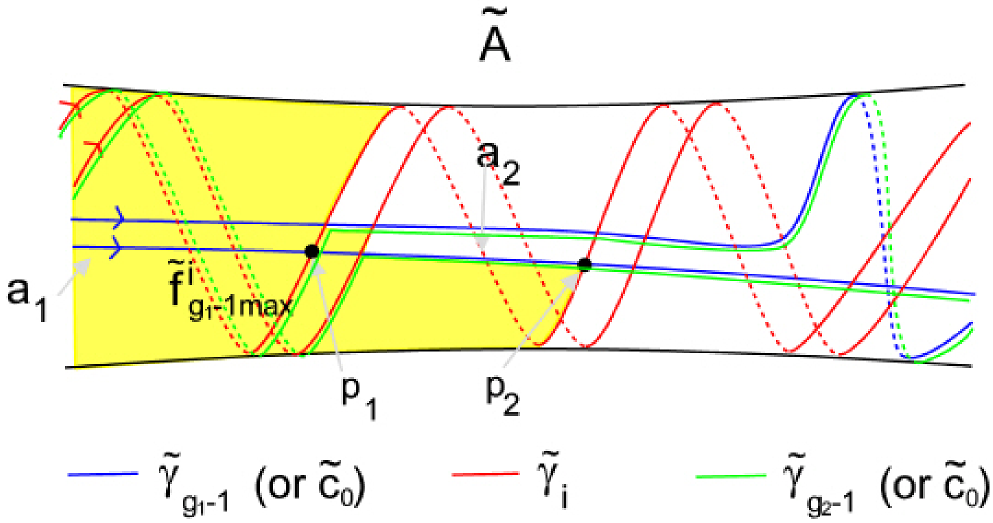

Figure 6.3: The arc a_2.

component of $\tilde{\gamma}_{g_1-1}$ as a_1. (This claim makes sense, because the arcs of $\tilde{\gamma}_{g_1-1}$ on the boundary of $\tilde{f}_{g_2-1max}$ coincide with the arcs of $\tilde{\gamma}_{g_2-1}$ on the boundary of $\tilde{f}_{g_2-1max}$, which also coincide with arcs of $\tilde{c}_0$ on the boundary of $\tilde{f}_{g_2-1max}$.)

The closure of the lift to $\tilde{A}$ of f_{g_2max} therefore contains a_1 and a_2. Let p_1 be one endpoint of a_1 and p_2 be an endpoint of a_2 connected to p_1 by a component of $\tilde{\gamma}_i$. p_1 and p_2 lie on the same component of $\tilde{c}_0$, since both a_1 and a_2 are both subarcs of the same nontrivial component of $\tilde{c}_0$. The assumption that all the nontrivial arcs of $\tilde{c}_0$ and $\tilde{c}_1$ are oriented in the same way is used here to ensure that f is monotone along any subarc of $\tilde{\gamma}_k$ connecting p_1 and p_2. Since the closure of f_{g_2max} contains p_1 and p_2, it will also have to contain any arcs of $\tilde{c}_0, \tilde{\gamma}_n$ and $\tilde{c}_1$ connecting p_1 and p_2. By construction, f_{g_2-1max} has to contain an arc of γ_{g_2} that is obtained from the intersection of an arc of c_0 with f_{g_2-1max} by Dehn twisting twice around α. This argument can be continued until a $g_\kappa \leq i$ is obtained. All components of the intersection of $\tilde{c}_0$ with the lift of $f_{g_\kappa max}$ are κ-twisted with respect to all components of the intersection of $\tilde{\gamma}_{g_{\kappa+3}}$ with the lift of $f_{g_{\kappa+3}max}$. Since the intersection of γ_{g_κ} with the interior of $f_{g_\kappa max}$ coincides with the intersection of c_1 with the interior of $f_{g_\kappa max}$, this contradicts the assumption that c_1 is not κ-twisted with respect to c_0.

If the nontrivial arcs of $\tilde{c}_0$ and $\tilde{c}_1$ are not all oriented in the same way, the

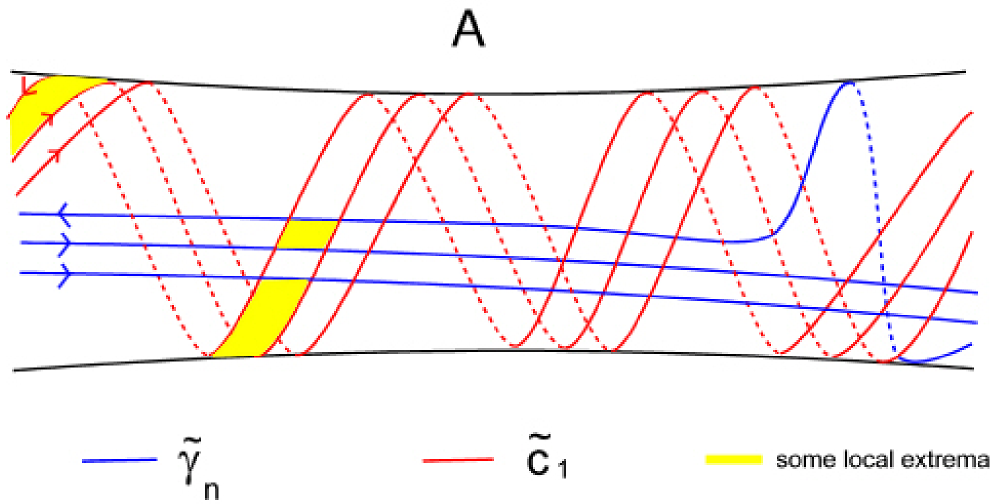

Figure 6.4: Local extrema of f inside A.

existence of a subarc $\gamma_n, \gamma_{n+1}...\gamma_i, ...\gamma_{m-1}, \gamma_m$ of the sequence $c_0, \gamma_1, ...\gamma_j, c_1$ is shown, where γ_n and γ_m are chosen such that all nontrivial arcs of $\tilde{\gamma}_n$ and $\tilde{\gamma}_m$ are oriented in the same way and $d_A(c_0, \gamma_n) < 2$ and $d_A(c_1, \gamma_m) < 2$. That this is possible follows from the observation that if the distance in the subsurface projection to A between c_0 and some γ_{n-1} is no more than one, and the distance between c_0 and γ_n in the subsurface projection to A is two, then the nontrivial arcs of $\tilde{\gamma}_{n-1}$ all have to be oriented in the same way. This remains true for $\gamma_{n+1}, \gamma_{n+2}...\gamma_m$, whenever the distance between γ_{m-1} and c_1 in the subsurface projection to A is at least two. A proof of this makes use of the fact that if the nontrivial arcs of $\tilde{\gamma}_{n-1}$ are not all oriented in the same way, there are local extrema of f "inside A", as shown in figure 6.4.

To put this more precisely, whenever the nontrivial arcs are not all oriented in the same way, the lift, $\tilde{f}$, of f to $\tilde{A}$ is not monotone along the nontrivial arcs of $\tilde{c}_0$ and $\tilde{c}_1$. It therefore has to have local extrema whose boundary contains horizontal arcs of $\tilde{c}_1 \cap (\tilde{A} \setminus \tilde{c}_0)$ with both endpoints on nontrivial arcs of $\tilde{c}_0$. Adding a handle corresponding to one of these local extrema to γ_k for some k transforms two nontrivial arcs of $\tilde{\gamma}_k$ into two trivial arcs of $\tilde{\gamma}_{k+1}$, thereby reducing the intersection number with α. Recall the definition of the patch p_i from the previous chapter. p_{n-1} has to contain all the local maxima in $\tilde{A}$. This is because any arc in the CW complex $\tilde{c}_0 \cup \tilde{c}_1$ is homotopic to an arc in $\tilde{f}_{n-1max}$, by the choice of n. Therefore the intersection number of γ_n with α has been reduced as much as possible, i.e. every nontrivial arc is oriented in the same way. By lemma 31, the same argument applies to γ_m and c_1.

To see where the "+3" in the statement of the lemma comes from, recall that the distance in the subsurface projection was defined to be the minimum distance between two nontrivial arcs. It might be the case that the nontrivial arcs of $\tilde{c}_0$ that are "least twisted" with respect to the nontrivial arcs of $\tilde{c}_1$ are all oriented in the opposite direction to the majority of the nontrivial arcs, and therefore do not appear in γ_n. Similarly for γ_m. If c_1 is not κ-twisted with respect to c_0, it might be the case that γ_m is $\kappa+1$-twisted with respect to γ_n, but not $\kappa+2$-twisted. From lemma 69, it follows that if γ_i is $\kappa+3$-twisted with repect to c_0, it has to be at least $\kappa+2$-twisted with respect to γ_n. The lemma then follows in this case by replacing c_0 and c_1 by γ_n and γ_m in the original argument.

If γ_i is $\kappa+3$-twisted with respect to c_1 instead of c_0, by lemma 31 the argument is the same, only with c_1 and c_0 interchanged.

□

Proof of theorem 68. Suppose $c_0, \gamma_1, ...\gamma_j, c_1$ is a surface producing sequence constructed as in chapter two. Since c_0 is not κ-twisted with respect to c_1, by lemma 70, γ_i can't be $\kappa+3$-twisted with respect to c_1 for any i. By theorem 43, γ_i can't have freely homotopic curves with the same orientation for any i, so theorem 64 gives a bound of

$$\frac{1}{\sqrt{-3\chi(S)(1-2\chi(S))(\kappa+3)i(m_1,m_2)}} \tag{6.2}$$

on the proportion of horizontal arcs of $c_1 \cap (S \setminus \gamma_i)$. If for each i every horizontal arc of $c_1 \cap (S \setminus \gamma_i)$ is homotopic to an arc on the boundary of f_{imax}, then equation 6.2 provides a lower bound on the proportional decrease in the intersection number with c_1 at each step. Let $t_0 = i(c_0, c_1)$, and let t_n be the recurrence relation

$$t_{n+1} := t_n - \frac{2}{k}\sqrt{t_n}\ , \text{ where } k = \frac{1}{\sqrt{-3\chi(S)(1-2\chi(S))(\kappa+3)}}. \tag{6.3}$$

From the previous arguments it is clear that t_n is an upper bound for $i(\gamma_n, c_1)$. $j-1$ is bounded from above by the smallest value of n such that $t_n \leq 2$. This recurrence relation is not easy to solve exactly, however, since the decrease in intersection number at each step is an even number and is *at least* $t_{n+1} - t_n$, the bound on $j-1$ is simpler than the solution to this recurrence relation. Assuming that the decrease in intersection number is as small as possible at each step, this decrease has to remain constant until

$$\sqrt{\frac{i(\gamma_m, c_1)}{k^2}}$$

decreases by one. Therefore, the square root of the intersection number decreases by one after at most $\frac{1}{k}$ steps, from which it follows that

$$j \leq \frac{1}{k}\left(\sqrt{i(c_0, c_1)} - 2\right) + 1 \qquad (6.4)$$

If there are horizontal arcs of $c_1 \cap (S \setminus \gamma_i)$ that are not homotopic to an arc on the boundary of f_{imax}, by theorem 66, it is possible to break the sequence $c_0, \gamma_1, ...\gamma_j, c_1$ up into at most $\frac{-11\chi(S)}{2} - 1$ subsequences of the form $\gamma_l, \gamma_l, ...\gamma_m$ such that for any $l \leq i \leq m$, a horizontal arc of $\gamma_m \cap (S \setminus \gamma_i)$ is either homotopic to an arc on the boundary of f_{imax} or the points of intersection on its boundary can be removed by a homotopy. From the proof of theorem 66, it is clear that if the sequence is broken up into the subsequences $c_0, ...\gamma_{s_1}$, $\gamma_{s_1}, ...\gamma_{s_2}$, $\gamma_{s_2}, ...\gamma_{s_3}$, etc. then the sum of the intersection numbers $i(c_0, \gamma_{s_1})$, $i(\gamma_{s_1}, \gamma_{s_2})$, $i(\gamma_{s_2}, \gamma_{s_3})$, etc. is no more than $i(c_0, c_1)$, so the worst case scenario is when the sequence has to be broken into $\frac{-11\chi(S)}{2} - 1$ subarcs, where

$$i(\gamma_{s_i}, \gamma_{s_{i+1}}) = i(c_0, \gamma_{s_1}) = i(\gamma_{s_n}, c_1) = \frac{i(c_0, c_1)}{\frac{-11\chi(S)}{2} - 1}$$

In this case, the bound on j is obtained by replacing $i(c_0, c_1)$ by

$$\frac{i(c_0, c_1)}{\frac{-11\chi(S)}{2} - 1}$$

in equation 6.4, and multiplying everything by $\frac{-11\chi(S)}{2} - 1$. The theorem follows from the bound on j by adding one to obtain the number of subsets of S that are attached in the construction of the surface, and multiplying the result by $\frac{g}{2}$ to obtain a bound on the genus of the surface. □

Bibliography

[1] Baryshnikov, Y. and Ghrist, R. (2009) 'Euler integration over definable functions' arXiv:0909.4054v1

[2] Benedetti, R. and Petronio, C. (1992) Lectures on Hyperbolic Geometry Springer.

[3] Bonahon, F. (1997), 'Geodesic Laminations on Surfaces' *Contemporary Mathematics* **00**.

[4] Bowditch, B. (2006), 'Intersection numbers and hyperbolicity of the curve complex' *J. reine angew. Math* **598**, 105-129

[5] Bowditch, B. (2007), 'Length bounds on curves arising from tight geodesics'*Geom. Funct. Anal.***17** 1001-1042.

[6] Bowditch, B. (2008), 'Tight geodesics in the curve complex' *Invent. math.* **171**, 281-300.

[7] Bredon, G. E. (1993) *Topology and Geometry*, Springer Verlag New York.

[8] Broaddus, N., Farb, B., and Putman, A. (2009), 'Irreducible Sp-representations and subgroup distortion in the mapping class group' arXiv:0707.2262v3

[9] Farb, B. and Margalit, D. (2009) *A Primer on Mapping Class Groups version 3.1* available at www.math.uchicago.edu/ margalit/mcg/mcgv31.pdf

[10] Fulton, W. (1995) *Algebraic Topology A First Course*, Springer Verlag New York.

[11] Jaco, w. (1980) *Lectures on Three-Manifold Topology*, American Mathematical Society

[12] Hamenstädt, U., Lecuire, C. and Otal, J. (2010) *Applications of Teichmller Theory to 3-Manifolds: 39 Oberwolfach Seminars* Birkhäuser.

[13] Harvey, W. (1981), 'Boundary structure of the modular group' Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference (I. Kra and B. Maskit, eds), Ann. of Math. Stud. 97, Princeton

[14] Hempel, J. (1976) *3-Manifolds* Princeton University Press.

[15] Hempel, J. (2001), '3-manifolds as viewed from the curve complex' Topology **40**(3), pp. 631-657.

[16] Hirsch, M. (1997) *Differential Topology* Springer.

[17] Leasure, J. (2002), 'Geodesics in the complex of curves of a surface' PhD Thesis. Available at http://hdl.handle.net/2152/1700

[18] Masur, H., Minsky, Y.(1999), 'Geometry of the complex of curves I: Hyperbolicity' *Invent. Math.***138**, 103-149.

[19] Masur, H., Minsky, Y. (2000), 'Geometry of the complex of curves II: Hierarchical Structure' *Geometric and Functional Analysis***10(4)**

[20] Matsuzaki, K. and Taniguchi, M. (1998)*Hyperbolic Manifolds and Kleinian Groups*Clarendon Press, Oxford.

[21] Milnor, J. (1965), *Lectures on the h-cobordism Theorem*, Princeton University Press.

[22] Minsky, Y. (2006), 'Curve complexes, surfaces and 3-manifolds' Proceedings of the International Congress of Mathematicians, Madrid, Spain.

[23] Penner, R. (1988) 'A Construction of Pseudo-Anosov Homeomorphisms', *Transactions of the American Mathematical Society*, Volume 310 Number 1

[24] Rafi, K. (2005) 'A characterization of short curves of a Teichmller geodesic', *Geometry and Topology*, Volume 9, 179-202

[25] Scott, P. (1978), 'Subgroups of surface groups are almost geometric'*J. London Math. Soc.***17**, 555-565.

[26] Shackelton, J. 'Tightness and computing distances in the curve complex' arXiv:math/0412078v3 14 Apr 2005

[27] Thurston, W. (1986), 'A norm for the homology of 3-Manifolds', *Mem. Amer. Math.Soc.***339**.