ISSN 1364-0380 (on line) 1465-3060 (printed)

Geometry & Topology
Volume 7 (2003) 329–397
Published: 19 June 2003

Period three actions on the three-sphere

JOSEPH MAHER
J HYAM RUBINSTEIN

Mathematics 253-37, California Institute of Technology
Pasadena, CA 91125, USA
and
Department of Mathematics and Statistics
University of Melbourne, Parkville VIC 3010, Australia

Email: maher@its.caltech.edu, rubin@ms.unimelb.edu.au

Abstract

We show that a free period three action on the three-sphere is standard, ie, the quotient is homeomorphic to a lens space. We use a minimax argument involving sweepouts.

AMS Classification numbers Primary: 57M60

Secondary: 57M50

Keywords: 3-manifold, 3-sphere, group action, spherical 3-manifold, lens space

Proposed: Cameron Gordon — Received: 6 April 2002
Seconded: Walter Neumann, Joan Birman — Revised: 1 June 2003

1 Introduction

1.1 Background

This paper presents a proof that a free period three action on the three-sphere is standard. In this section we give an outline of the proof, but we begin by giving a brief summary of some related results.

Thurston's Geometrization Conjecture implies that all 3-manifolds with finite fundamental groups are homeomorphic to quotients of S^3 by a finite group of isometries. This conjecture splits into the following two conjectures: the Poincaré Conjecture, which states that the universal cover of every closed 3-manifold with finite fundamental group is S^3, and the Spherical Spaceform Conjecture, which states that every 3-manifold whose universal cover is S^3 is homeomorphic to a quotient of S^3 by isometries.

The 3-manifolds which are quotients of S^3 by isometries are known as spherical or elliptic 3-manifolds. They have been classified by Hopf [2] and Seifert and Threlfall [12], by identifying the finite subgroups of $SO(4)$ which act freely on S^3. Wolf [15] has extended this to higher dimensions. There are more recent accounts of the 3-dimensional case in Scott [11] and Thurston [13].

Milnor [5] and Lee [3] investigated which finite groups might be able to act freely on S^3. They recovered the list of finite subgroups of $SO(4)$ which act freely, but also another family of groups. These extra groups are an infinite family of finite subgroups of $SO(4)$, which do not act freely on S^3 as subgroups of $SO(4)$, but may possibly have some non-standard free action on S^3. See table 1.

The group $Q(8n,k,l)$ is the semi-direct product of $\mathbb{Z}_{kl}$ with Q_{8n}, and has presentation $\{a,b,c \mid a^2 = (ab)^2 = b^{2n} = 1, c^{kl} = 1, aca^{-1} = c^r, bcb^{-1} = c^{-1}\}$, where $r \equiv -1 \bmod k$, and $r \equiv 1 \bmod l$. See Milnor [5] and Lee [3] for more details.

By work of Thomas [14], if the group is solvable, it suffices to show that the actions of its cyclic subgroups are standard. The only groups in table 1 which are not solvable are those containing the binary icosahedral group. In particular, showing that the cyclic actions are standard would eliminate the exceptional groups $Q(8n,k,l)$.

The first group action shown to be standard was $\mathbb{Z}_2$, by Livesay, [4]. His argument uses the standard sweepout by 2-spheres on S^3. An invariant unknotted curve is found by considering the intersections of the 2-spheres by their images

Subgroups of $SO(4)$ that act freely	Order
Cyclic, $\mathbb{Z}_n$	n
Quaternionic, $Q_{4n} \times \mathbb{Z}_m$, where n and m are coprime	$4nm$
Binary tetrahedral, $T \times \mathbb{Z}_m$, where m is coprime to 24	$24m$
Binary octahedral, $O \times \mathbb{Z}_m$, where m is coprime to 48	$48m$
Binary icosahedral, $I \times \mathbb{Z}_m$, where m is coprime to 120	$120m$
Index two subgroups of $Q_{4n} \times \mathbb{Z}_{2m}$	$4nm$
Index three subgroups of $T \times \mathbb{Z}_{3m}$	$24m$
Subgroups of $SO(4)$ which could only have non-standard actions	
$Q(8n, k, l)$	$8nkl$

Table 1: Finite groups that may act freely on S^3

under the involution. This was generalised to higher powers of two by Rice [7], Ritter [8] and Myers [6]. The other results due to Rubinstein [9, 10] use the fact that certain spherical manifolds contain embedded Klein bottles, and involve considering the intersections of the images of the Klein bottle under the group action.

Table 2 summarises which group actions are known to be standard.

Group	Reference
$\mathbb{Z}_2$	Livesay [4]
$\mathbb{Z}_4$	Rice [7]
$\mathbb{Z}_8$	Ritter [8]
Q_{2^k}	Evans, Maxwell [1]
$\mathbb{Z}_6, \mathbb{Z}_{12}, Q_{2^k}$	Rubinstein [9]
non-cyclic, non-quaternionic groups of order $2^a 3^b$	Rubinstein [10]
$\mathbb{Z}_{2^k}, \mathbb{Z}_{2^k.3}, Q_{2^k.3}, k \geqslant 2$	Myers [6]

Table 2: Group actions known to be standard

1.2 Acknowledgements

The first named author would like to thank his PhD advisor Daryl Cooper for his exceptional fortitude and patience during the writing of this paper. He would also like to thank the National Science Foundation for providing partial support during the writing of this paper. Both authors would like to thank the referee for many helpful comments.

1.3 Outline

An action of $\mathbb{Z}_3$ on S^3 is given by a diffeomorphism $g\colon S^3 \to S^3$, which is free, and which is period three. This means that g generates a group $G \cong \mathbb{Z}_3$, and as g is free, the quotient S^3/G is a manifold. If g is linear, namely an element of $SO(4)$, then the quotient is a lens space. We say the action of g is standard if g is conjugate by a diffeomorphism to an element of $SO(4)$. This is equivalent to the quotient being diffeomorphic to a lens space.

We can show that the action is standard by finding an invariant unknotted circle in S^3. An unknotted invariant circle has an invariant neighbourhood which is a solid torus. The complement of this neighbourhood is also an invariant solid torus. The quotient of a solid torus by a finite group acting freely is again a solid torus, so the quotient manifold is the union of two solid tori, a lens space.

We will find an invariant unknotted curve by studying sweepouts of S^3. A sweepout of S^3 is a family of surfaces which "fill up" the manifold. A simple example is the foliation of S^3 by 2-spheres with two singular leaves which are points. Think of the leaves as parameterised by time, starting with one singular leaf at $t=0$ and ending with the other one at $t=1$. At a non-singular time t, the sweepout consists of a single 2-sphere S_t. We can think of the union of the leaves as a 3-manifold, in this case S^3, with a height function for which each level set is a 2-sphere. The map from the leaf space to S^3 is degree one, and is an embedding on each level set of the height function.

For our purposes, we require a more general definition, in which the sweepout surfaces at non-singular times may be finitely many spheres. We say a generalised sweepout is a 3-manifold M, with a height function h on it, so that the level sets at regular values are unions of 2-spheres, and a degree one map $g\colon M \to S^3$, which is an embedding on each level set of the height function. We shall think of the height function on M as time. The 3-manifold M is in fact either S^3, or a connect sum of $S^2 \times S^1$'s.

We can can look at the three images of the sweepout spheres under the group $\mathbb{Z}_3$. Generically, they will intersect in double curves and triple points. In order to distinguish the three images of the spheres under $\mathbb{Z}_3$, we will colour them red, blue and green.

By general position, we can arrange that the sweepout spheres intersect transversely for all but finitely many times, and that the non-transverse intersections all come from a finite list of possibilities, corresponding to critical points of the height function on either M itself, or on the double or triple point sets of M. We call these non-transverse intersections *moves*, and we can describe a sweep-out by drawing the configurations in S^3 in between the critical times, with each neighbouring pair of pictures differing by a non-transverse intersection. Critical points of the height function on M change the number of 2-spheres by one. So a 2-sphere can either appear or disappear, or two spheres can either split apart or join together. We will call the appearance of a sphere an *appear* move, and the disappearance of a sphere a *vanish* move. A critical point which splits two spheres apart will be called a *cut* move, and a critical point which joins two spheres together will be called a *paste* move. Critical points of the height function on the double set change the number of double curves, by either creating or destroying a double curve, which we will call *birth* or *death* moves respectively, or by saddling curves together, which we will call saddle moves. Critical points of the height function on the triple set change the number of triple points, which we will call *triple point* moves. In fact the number of triple points is always a multiple of six, as every triple point has three images under G, and there must also be an even number of triple points.

We look for a sweepout that is "simple", by defining a complexity for sweepouts, and showing how to change the sweepout to reduce complexity. We say that the *complexity* of the sweepout at a generic time t is the ordered pair (n, d), where n is the number of triple points, and d is the number of simple closed double curves without triple points. We order the pairs (n, d) lexicographically. We say that the complexity of a sweepout is the maximum complexity that occurs over all generic times. We say that a sweepout is a minimax sweepout if it has minimal complexity. The complexity only depends on the graph of double curves, so moves which do not change the graph of double curves do not change complexity.

At a generic time, the double curves form an equivariant graph in S^3, which may contain an invariant unknotted curve. The basic strategy is to show that a minimax sweepout must have an unknotted invariant curve in its graph of double curves at some time during the sweepout. This follows if we can show that an arbitrary local maximum can either be reduced in height by changing the sweepout, or else contains an invariant unknotted curve in its graph of double curves. A minimax sweepout contains local maxima that cannot be removed, so it contains an unknotted invariant curve.

We now give a more detailed outline of the argument.

1.4 Modifications

In order to apply this strategy we need some way to simplify a sweepout with many triple points. One of the basic operations we will use is to change the sweepout by cutting out a subset of the sweepout, and replacing it with a different subset. We now give an informal description of this procedure, which we will call a *modification*. A precise description of this is given in Section 2.6.

Imagine choosing an equivariant 3-dimensional subset N of S^3, and watching the images of the sweepout surfaces inside it for some time interval I. Assume that the boundary of $N \times I$ is disjoint from any of the moves of the sweepout. We can think of the image of the sweepout in $N \times I$ as a sequence of pictures of spheres and planar surfaces in N, with each picture in the sequence differing from its neighbours by a move. If we draw a sequence of pictures in N which is different, but which agrees with the original one on the boundary of $N \times I$, then we can try to construct a sweepout by replacing the original map into $N \times I$ by a new one determined by the new set of pictures. We need to check that the new map really is a sweepout, by checking that the level sets are still spheres, and that the projection to S^3 is still degree one, but if it is a sweepout, then we say that we have produced a new sweepout from the original one by a modification.

The two main modifications we will use will involve removing double curves, and removing bigons. A *bigon* is a subdisc of a sweepout surface whose boundary consists of a pair of double arcs which share common endpoints, and whose interiors are disjoint. We say a subsurface of a sweepout sphere is *double curve free* if its interior is disjoint from all of the double curves.

Removing double curves Suppose we have a double curve that bounds a double curve free disc Δ in a sweepout sphere, for some time interval I. Think of the sweepout surface containing the disc as "horizontal", and the other sweepout surface intersecting the double curve as "vertical". Then we can change the sweepout to remove this double curve for a subinterval of I, by "pinching off" the vertical annulus at the beginning of the time interval, and then replacing it at the end of the time interval, so that the sweepout returns to its initial configuration. We want to do this generically, using moves, so first use a cut move to pinch the vertical annulus into two discs, one of which is disjoint from the horizontal surface, and the other intersects it in the double curve. Now use a double curve death move to remove the double curve. We will call this special pair of moves which removes a double curve a compound double curve death, and its time reverse a compound double curve birth. This removes the double curve, and also splits one of the spheres into two. We can also do this equivariantly, as the disc Δ is disjoint from its images, so in fact we will remove the orbit of a double curve, and split three spheres apart. This is illustrated in Figure 11.

Removing bigons Suppose we have a bigon A, which is double curve free for a time interval I. We can reduce the number of triple points in the sweepout for a subinterval of I by "undoing" the bigon. Think of the bigon as being "horizontal", and the sweepout surfaces which intersect it in the double arcs of the boundary of A as being "vertical". We can push the two vertical surfaces sideways through each other to remove the bigon, and reduce the number of triple points. As before, we wish to do this generically, and this in fact requires a saddle move followed by a triple point death move. We will call the pair of moves that remove the bigon a compound triple point death, and we will call their time reverse a compound triple point birth. We can also also do this equivariantly, as a double curve free bigon is disjoint from its images. This is illustrated in Figure 13.

Both of these modifications are in fact guaranteed to produce new sweepouts. Removing bigons only changes the level sets up to isotopy, while removing a double curve splits one 2-sphere into two spheres, and then joins them back together again. The degree of the map into S^3 is also preserved, because as the modification neighbourhood N_t is not all of S^3, (perhaps after adjusting the sweepout by an isotopy) we can find a point in S^3 which does not lie in any of the N_t, so the pre-image of this point does not change.

We can start using these modifications to try to simplify the sweepout. For example, suppose there is a local maximum in the graph of complexity against time, for which there is a double curve free bigon which is disjoint from all the moves occurring during the local maximum. Then we can remove the bigon for the duration of the local maximum, which reduces the height of the local maximum.

1.5 The main argument

Suppose we have a time interval which contains a local maximum of complexity. This will contain a move which increases complexity, immediately followed by a move which decreases complexity. If these two moves have disjoint move neighbourhoods, then we can just swap the order in which they occur, reducing the height of the local maximum. In general, if we change the sweepout in any way so that the height of a local maximum is reduced, then we say that the local maximum has been *undermined*.

The main argument, in Section 3, shows how to undermine an arbitrary local maximum, assuming the following lemmas:

Lemma 4.1 *Every sweepout contains triple points.*

We say that a bigon is *vertex free* if it contains no triple points in its interior, and precisely two triple points in its boundary.

Lemma 5.1 *Suppose a non-triple point move occurs while the configuration contains triple points. Then there is a vertex free bigon orbit disjoint from the move neighbourhood of the move.*

We say that a local maximum which consists of a compound triple point birth, followed by a compound triple point death, is a *special case local maximum*. There are only finitely many different ways in which the two compound moves can intersect.

Lemma 6.1 *A special case local maximum can either be undermined, or else contains an invariant unknotted circle.*

The first step of the argument in Section 3, is to change the sweepout, so that the highest local maxima consist of compound moves only. Lemma 5.1 implies that for every non-triple point move there is a vertex free bigon which is disjoint from the move. If this contains simple closed double curves, then there is an innermost one which we can remove using a double curve removal modification. If there are no double curves inside the bigon, then we can remove the bigon using a bigon removal modification. Both of these modifications reduce the height at which the non-triple point move occurs, and insert extra compound moves into the sweepout. We can use Lemma 5.1 repeatedly until all non-triple point moves lie far below the height of the local maxima. We then show that we can replace a non-compound triple point move with a compound triple point move, possibly adding a non-triple point move at a lower height than the local maxima. In this way we may assume that all moves above a certain height are compound moves, so it suffices to show how to undermine local maxima in which all the moves are compound moves.

The next step is to show that we can undermine all local maxima consisting of compound moves only, which are not special case local maxima, ie, they contain at least one compound double curve move. If both compound moves are double curve moves which are disjoint, then we can swap the order of the moves to undermine the local maximum. If they are not disjoint, then we show that we can find a disjoint double curve or bigon to remove for the duration of the local maximum. If one of the compound moves is a double curve move and the other is a triple point move, then we can swap the order, as the bigon for the triple point move must be disjoint from the disc for the double curve move.

Lemma 6.1 now implies that either we can remove all the special case local maxima, or else there is an invariant unknotted curve. Lemma 4.1 ensures that we can not remove all the triple points from the sweepout, so we cannot undermine all the special case local maxima, so there must be at least one that contains an unknotted invariant curve.

In order to complete the proof it remains to prove the three lemmas, and we now give a brief summary of the arguments that we will use.

1.6 Every sweepout contains triple points

We can choose a continuously varying "inside" for the sweepout spheres S_t. As g is degree one, we can choose the inside so that it starts out small and ends up large. The three images of S_t under G are coloured, and we colour the inside of each image of S_t with the same colour as the surface. So each component of $S^3 - G \,.\, S_t$ may be coloured by some combination of the colours red, blue and green. If a region is outside of all of the spheres, then we say that it is a clear region.

Before any of the sweepout spheres have appeared, S^3 is a clear connected invariant region. After the sweepout, S^3 is coloured with all three colours, and there are no clear regions at all. So some move must break up the clear connected invariant region. We show that if there are no triple points, none of the non-triple point moves can do this, using a simple case by case argument.

1.7 Disjoint bigons

We show that if there are bigons, then there is always a bigon disjoint from a non-triple point move. We use an elementary combinatorial argument to show this.

1.8 Special cases

A special case local maximum consists of a compound triple point birth, followed by a compound triple point death. The compound birth move creates a bigon, A say, and the compound death destroys a bigon, B say. Between the compound moves, both bigons are present at the same time. If the two bigon orbits are disjoint, then we can undermine the local maximum by just swapping the order in which the compound moves occur. If the two bigon orbits are the same, then the configuration before the local maximum is isotopic to the configuration after the local maximum, so we can just replace the moves creating the local maximum with an isotopy, thus undermining the local maximum. If they are not disjoint, and not the same, then there are only finitely many different ways in which they can intersect, and we deal with each possibility in turn.

For example, if the bigon orbits share a single vertex orbit in common, then we explicitly construct a new sweepout with fewer triple points. Otherwise, the two bigon orbits share all six vertices in common. If the union of the bigon orbits is connected, then we show that there is an invariant unknotted curve. In the remaining cases, we prove a useful undermining lemma, which shows that if we have a 3-ball disjoint from its images under G, which intersects the sweepout in a "simple" way, then we can replace the sweepout inside the 3-ball with one which has no triple points. We then show how to find such a 3-ball which contains the compound moves, for each of the remaining special cases.

We now give a brief description of the undermining lemma, and explain the "simple" condition on the intersection of the sweepout surface with the 3-ball. Assume we have chosen a 3-ball B which is disjoint from its images under G, which contains all the compound moves, and contains no triple points before or after the local maximum. The intersection of the sweepout surfaces with the boundary of B is constant up to isotopy, and consists of a pattern of intersecting circles, coloured red, blue and green, which we shall call the boundary pattern. If there is a simple closed curve of intersection which bounds a double curve free disc in ∂B then we can use this disc as a cut disc for a cut move, to reduce the number of curves of intersection between the sweepout surfaces and ∂B. If the curves of intersection create a double curve free bigon in the boundary, then we can use the bigon as a saddle disc for a saddle move which reduces the number of intersections of the double curves with ∂B. If we can remove all the intersections of the sweepout surfaces with ∂B in this way, then we say that the boundary pattern is *saddle reducible*. This is what we mean by "simple" intersections.

We can construct a partial sweepout which agrees with the original one on

∂B. Start with the configuration in B before the local maximum, and now do cut moves and saddle moves in some order to remove all the intersections of the sweepout surface with a 2-sphere parallel to ∂B, but just inside B. The remaining components of the sweepout surfaces inside B can now be removed using cut and death moves, to remove the double curves, and vanish moves to remove 2-sphere components. We can construct a similar partial sweepout starting with the configuration after the local maximum, and working back in time, using the same sequence of cut and saddle moves near the boundary. These two partial sweepouts can then be patched together in the middle by an isotopy to create the desired replacement partial sweepout.

2 Preliminary definitions

2.1 Generalised sweepouts

Definition 2.1 *Free action*

We say that a *free action* of the group $G \cong \mathbb{Z}_3$ on S^3 is generated by the diffeomorphism $g\colon S^3 \to S^3$ if g has no fixed points, and g^3 is the identity. Therefore f generates a cyclic group of order three, $G = <g> \cong \mathbb{Z}_3$.

This means that g is orientation preserving.

Remark 2.2 Throughout this paper, we assume that we have been given some fixed g, which we do not change.

Notation 2.3 If N is a subset of S^3, then we write $G \,.\, N$ for the orbit of N under G. As g is free, S^3/G is a manifold, we will call this manifold L.

Definition 2.4 *Standard action*

We say that the action is *standard* if g is conjugate by a diffeomorphism to a linear map, ie, an element of $SO(4)$.

This is equivalent to L being diffeomorphic to the quotient of S^3 by a linear group, in this case a lens space. The following theorem is well known:

Theorem 2.5 *If there exists a smooth unknotted invariant circle in S^3, then the action of G is standard.*

Proof Given an unknotted invariant circle, we can find an invariant tubular neighbourhood, which is a solid torus. As the curve is unknotted, the complement of the tubular neighbourhood is also a solid torus. So if we can show that a free action on a solid torus gives a solid torus, then this gives a genus one Heegaard splitting of the quotient, which is therefore a Lens space. We now show that a free $\mathbb{Z}_3$ action on a solid torus is standard.

The quotient manifold has torus boundary, as the boundary of the solid torus is a torus, and the quotient of a torus by a free $\mathbb{Z}_3$ action is also a torus. The quotient manifold is also irreducible, as it is covered by a solid torus, which is irreducible. A meridional disc in a solid torus is a compressing disc for the boundary, and this projects down to an immersed disc in the quotient, so the quotient manifold also has compressible boundary. By the loop theorem, there is an embedded compressing disc in the quotient. If we cut the quotient along this disc, we get a manifold with 2-sphere boundary, which bounds a 3-ball, as the quotient is irreducible. Therefore the quotient manifold is a 3-ball with a single one-handle attached, ie, a solid torus. □

Definition 2.6 *A generalised sweepout*

A *generalised sweepout* is a triple (M, f, h), where

- M is a closed, orientable 3-manifold.
- The smooth map $h\colon M \to \mathbb{R}$ is a Morse function, such that for all but finitely many $t \in \mathbb{R}$, the inverse image, $h^{-1}(t)$ is a collection of 2-spheres.
- The smooth map $f\colon M \to S^3$ is degree one.
- The map $f|_{h^{-1}(t)}$ is an embedding on the level set $h^{-1}(t)$ for every t.

We will often write a sweepout as (M, ϕ), where ϕ denotes the map $(f \times h)\colon M \to S^3 \times \mathbb{R}$. We will think of $t \in \mathbb{R}$ as the time coordinate.

Remark 2.7 The map $\phi\colon M \to S^3 \times \mathbb{R}$ is a smooth embedding. We will write π_{S^3} and $\pi_{\mathbb{R}}$ for the projection maps from the product $S^3 \times \mathbb{R}$ to its factors. From now on, whenever we say "sweepout", we mean "generalised sweepout".

Notation 2.8 We will write M_t for $h^{-1}(t)$. The M_t form the leaves of a singular foliation $\mathcal{F}$ of M. We will write S_t for $f(M_t)$, and call these the sweepout surfaces at time t.

Remark 2.9 The manifold M need not be connected. In fact each component of M will be either S^3 or a connect sum of $S^2 \times S^1$'s. We prove this later on as Proposition 2.45.

Example 2.10 Let $M = S^3$ be the set of points a unit distance from the origin in $\mathbb{R}^4$, with the Morse function h given by the x-coordinate. Let $f\colon M \to S^3$ be the identity map. Then $(S^3, f \times h)$ is a sweepout.

This example gives a foliation of S^3 by 2-spheres, with two singular leaves consisting of points. We will think of the leaves of this foliation as being parameterised by a time interval. Time starts at $t = -1$ at one singular leaf, and ends at $t = 1$ at the other. So in the standard sweepout S_{-1} is a single point. As t increases, S_t becomes a sphere which moves away from the initial point, increasing in size, till at $t = 0$, S_t is an equatorial sphere. It then decreases in size, until at $t = 1$, S_1 is again a single point, in fact the antipodal point to S_{-1}.

Definition 2.11 *Morse function on a manifold with boundary*

Let M be a manifold with boundary, and let $h\colon M \to \mathbb{R}$ be a function, which extends to a Morse function on M union an open collar neighbourhood of ∂M. There should be no singularities on ∂M. Then h is a Morse function on M.

Definition 2.12 *Compatible product structures*

Let K be a compact equivariant submanifold of $S^3 \times \mathbb{R}$, for which there is a level preserving diffeomorphism $\psi\colon\ S^3 \times \mathbb{R} \to S^3 \times \mathbb{R}$, so that $\psi(K)$ is a product $N \times I \subset S^3 \times \mathbb{R}$. We will always choose N to be 3-dimensional. Then we say that K has a compatible product structure. We can think of K as a family of equivariant submanifolds of S^3 that varies continuously with time. We will write K as N_I, and we will write N_t to refer to $K \cap (S^3 \times t)$.

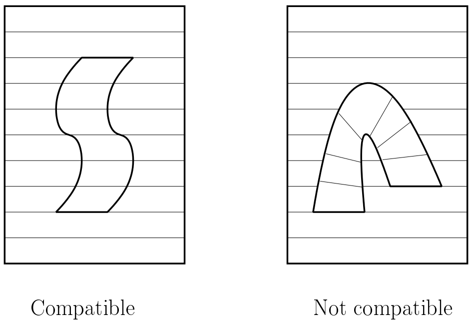


Figure 1: Compatible and incompatible product structures

The submanifold K need not be connected. Note that K need not be the product of a subset of S^3 with an interval in $\mathbb{R}$, it just has to be isotopic to one under a level-preserving isotopy of $S^3 \times \mathbb{R}$.

We will often be interested in how a compatible product submanifold intersects the sweepout surfaces.

Definition 2.13 *Constant boundary pattern*

Let (M, ϕ) be a sweepout, and let N_I be a compatible product submanifold of $S^3 \times \mathbb{R}$. If the intersection of the sweepout surfaces with ∂N_t only changes by an isotopy during the time interval I, then we say that N_I has *constant boundary pattern*, with respect to the sweepout (M, ϕ).

Definition 2.14 *A partial sweepout*

Let N_I be a compatible product submanifold of $S^3 \times \mathbb{R}$, where N is three dimensional. Let P be a 3-manifold with boundary, which need not be connected, and let $(f \times h)\colon P \to N_I$ be a proper embedding so that h is a Morse function and $h^{-1}(t)$ is a union of spheres and planar surfaces for all but finitely many t. If the surface $h^{-1}(t)$ has singularities, they should be disjoint from ∂P.

Then $(P, f \times h)$ is a *partial sweepout* with *image neighbourhood* N_I.

Remark 2.15 An image neighbourhood has constant boundary pattern, as there are no singularities on the boundary.

Definition 2.16 *A partial sweepout which is a product*

Let (P, ϕ) be a partial sweepout, in which P has a product structure, so that the image of the product structure on $\phi(P)$ is compatible with the product structure on $S^3 \times \mathbb{R}$. Then we say that (P, ϕ) is a *product partial sweepout.*

2.2 Diagram Conventions

Notation 2.17 The sweepout surface S_t is a union of 2-spheres, and we will label them red. The spheres in S_t have two sets of images under G, we will label the spheres in gS_t green, and the spheres in $g^2 S_t$ blue. If two spheres intersect in a double curve, we will label the double curve by the complementary colour, ie, the colour of the sphere not involved in the intersection. For example, if the red sphere intersects the green sphere in a double curve, we will label that double curve blue. To distinguish colours when printed in black and white (or viewed on a b/w monitor) we have used solid lines for green, long dashes for blue and short dashes for red.

Definition 2.18 *The outside of a sphere*

At a non-singular time t we can choose a normal vector for one of the disjoint red spheres. As each 2-sphere is separating in S^3, we can choose compatible normal vectors for the other spheres in S_t, so that in each complementary region the normal vectors all point either in or out. As t varies, we can choose a continuously varying family of such normal vectors defined on $S^3 - \{\text{singular points}\}$. We define the *outside* of the family of spheres to be the side that the normal vector points to, and we choose the normal vector so that the outside starts out large and ends up small.

Notation 2.19 The inside of the sphere will be labelled with the same colour as the sphere. Double curves occur in threes, one of each colour, and generically intersect spheres of the same colour transversely.

Definition 2.20 *A region*

A *region* (at time t) is the closure of a connected component of the complement of the orbit of S_t in S^3.

A region may be coloured by any combination of the colours red, green and blue, so there are eight different ways in which a region may be coloured.

Definition 2.21 *A clear region*

If a region is not coloured by any colour, we say that it is a *clear region*.

In order to draw diagrams of the intersections of the spheres, it is helpful to think of S^3 as $\mathbb{R}^3 \cup \{\infty\}$. There is an isotopy of S^3 that takes one of the red spheres to the xy-plane, so that the normal vector points upwards in the direction of the positive z-axis.

Definition 2.22 *The blue-green diagram and graph*

A *blue-green diagram* consists of finitely many red spheres, which may contain finitely many blue and green circles. The union of the green circles is embedded, as is the union of blue circles. The green circles may intersect the blue circles transversely to form a four-valent graph, which we call the *blue-green graph.*

Definition 2.23 *The configuration*

We will refer to the image of the three families of spheres in S^3, or some subset of S^3, as a *configuration.*

Different configurations may give rise to the same blue-green diagram. The blue-green graph may have many connected components in each red sphere. The vertices of the blue-green graph correspond to triple points, and are four-valent in the blue-green graph. They are six-valent in the graph of double curves in the 3-dimensional configuration.

Definition 2.24 *A red bigon*

A *red bigon* is a closed disc in a red sweepout sphere S_t, whose boundary consists of a pair of simple arcs, one green and one blue, which contain no triple points in their interiors. The endpoints of the arcs are a pair of triple points.

A green bigon is the image of a red bigon under g, and a blue bigon is the image of a red bigon under g^2.

Definition 2.25 *A bigon orbit*

A *bigon orbit* $G\,.\,A$ is the orbit of a red bigon A.

A bigon orbit is equivariant, and consists of a red bigon, a green bigon and a blue bigon. The interior of a red bigon may intersect its images if it contains double curves in its interior. However the boundary of a red bigon is disjoint from its images under G, see Corollary 2.51. There may be triple points in the interior of the bigon orbit, if the red bigon contains more than one component of the blue-green graph.

Definition 2.26 *Vertex free*

We say a subsurface of the sweepout spheres is *vertex free* if it contains no triple points in its interior, and there are no double arcs that cross the boundary transversely.

A vertex free subsurface of the sweepout spheres may still contain double curves without triple points in its interior.

Definition 2.27 *Double curve free*

We say a subsurface of the sweepout spheres is *double curve free* if its interior is disjoint from the double curves.

Remark 2.28 Double curve free implies vertex free.

2.3 General position

Definition 2.29 *Singular sets*

Let $(M, f \times h)$ be a generalised sweepout, and let $g\colon S^3 \to S^3$ be a diffeomorphism of period three. Let $\bar{f} = f/G$, and let $\overline{\phi} = \bar{f} \times h\colon M \to L \times \mathbb{R} = (S^3/G) \times \mathbb{R}$. The *double set*, Σ_2, of M is $\{x \in M \mid \overline{\phi}^{-1}(\overline{\phi}(x))$ contains at least two points $\}$. The *triple set*, Σ_3, of M is $\{x \in M \mid \overline{\phi}^{-1}(\overline{\phi}(x))$ contains at least three points$\}$.

As $\overline{\phi}$ is a map from a 3-manifold to a 4-manifold, Σ_2 will be a 2-dimensional manifold and Σ_3 will be a 1-dimensional manifold, if $\overline{\phi}$ is self-transverse.

Definition 2.30 *General position for a generalised sweepout with respect to g*

Let $(M, f \times h)$ be a generalised sweepout, and let $g\colon S^3 \to S^3$ be a period three diffeomorphism.

We say that the sweepout $(M, f \times h)$ is in *general position with respect to* g, if h is a Morse function when restricted to $\overline{\phi}(\Sigma_2)$ and $\overline{\phi}(\Sigma_3)$. Furthermore, we require the critical times of the Morse functions on M and the singular sets to all be distinct.

Remark 2.31 We require h/G to be a Morse function on the *image* of the double set, as critical heights on the image will always have pairs of singularities in the pre-image in M.

Theorem 2.32 *Let $(M, f \times h)$ be a generalised sweepout, and let g be a period three diffeomorphism of S^3. Then we can alter the map $f \times h$ by a small homotopy so that $(M, f \times h)$ is a sweepout in general position with respect to g.*

Proof The map $\overline{\phi}$ is a smooth immersion, so we can change $\overline{\phi}$ to $\overline{\phi}'$ by a small homotopy so that $\overline{\phi}'$ is a self-transverse immersion. The singular sets Σ_2 and Σ_3 will now be nested submanifolds of M. A small homotopy of $\overline{\phi} = \bar{f} \times h$ is the same as small homotopies of $\bar{f}$ and h, so we may assume that h' is a Morse function, as Morse functions are dense in $C^\infty(M, \mathbb{R})$.

We can now perturb the product structure on $L \times \mathbb{R}$ to make coordinate projection onto $\mathbb{R}$ a Morse function, not just on M, but on the singular sets as well. Note that perturbing the product structure changes the height function on M, without changing the singular sets. One way to construct such a perturbation is to choose an embedding of L in $\mathbb{R}^k$, for some k, and then use this to define an embedding $e\colon L \times \mathbb{R} \to \mathbb{R}^k \times \mathbb{R}$. Let π_i be projection on to the i-th coordinate of $\mathbb{R}^{k+1}$, and consider the functions e_a on M defined by $e_a = \pi_{k+1} \circ e \circ \overline{\phi}' + a_1\pi_1 + \cdots + a_{k+1}\pi_{k+1}$, for $a \in \mathbb{R}^{k+1}$. The set of $a \in \mathbb{R}^{k+1}$ for which e_a is a Morse function on the images of M and the singular sets is an open dense set of $\mathbb{R}^{k+1}$, as they are immersed manifolds in $\mathbb{R}^{k+1}$. Therefore we can choose a to be close to zero, and furthermore $\bar{f}' \times h'$ is homotopic to $\overline{\phi}'' = \bar{f}' \times (h' + e_a)$ using the homotopy $\bar{f}' \times (h' + e_{ta})$ for $t \in [0,1]$.

We have changed $\overline{\phi}$ to $\overline{\phi}''$ by a small homotopy, so that $\overline{\phi}''$ is transverse, and h'' is a Morse function on M and the singular sets. As f'' is homotopic to f, f'' is degree one. It remains to check that the level sets are still 2-spheres. We changed h to h'' by a small C^∞ perturbation, so not only will ${h''}^{-1}(t)$ lie in a small product neighbourhood of $h^{-1}(t)$, but we have also changed the first derivative by only a small amount. This means that the projection map ${h''}^{-1}(t) \to h^{-1}(t)$ coming from the product structure on a small tubular neighbourhood of $h^{-1}(t)$, is a local diffeomorphism, hence a covering map. Therefore ${h''}^{-1}(t)$ is also a 2-sphere. □

For each critical point, we can choose a small partial sweepout which contains it, which is isotopic to a "standard model" for that type of critical point. We will call such a small partial sweepout a *move*. We now give a precise definition of a move, and then list all the critical points that may arise, and describe their "standard models".

Definition 2.33 *Moves*

A move is a partial sweepout whose image neighbourhood is a tubular neighbourhood for the orbit of a critical point. We may assume that the image neighbourhood has a compatible product structure N_I, where each N_t is the orbit of a 3-ball which is disjoint from its images. Furthermore, we require the intersection of the sweepout with the image neighbourhood to be "as simple as possible". This means that the sweepout surfaces in N_I must be isotopic to the explicit descriptions of the configurations in N_I which we give below.

We call N_I the *move neighbourhood*. We call I the *move interval*.

The following is a complete list of the moves that may occur, and a description of the images of the sweepout surfaces in the move neighbourhood. In each case we draw pictures of the sweepout surfaces in one of the components of N_t. The other components are the disjoint images of this under G.

Critical points of the height function on M

(1) **Appear and vanish moves** (Index 0 and 3)

In an appear move, a sphere orbit appears. In a vanish move a sphere orbit disappears. At the critical time the singular set consists of the orbit of a single point. Passing through the critical time a 2-sphere either appears or vanishes.

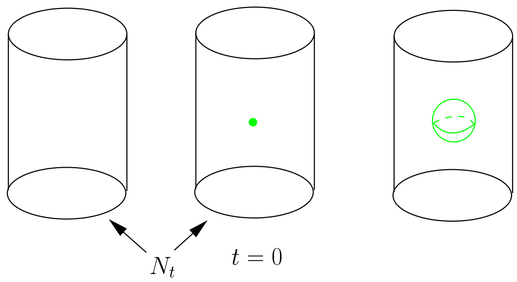


Figure 2: An appear move

The sweepout spheres are disjoint from ∂N. Before the singular time, N is disjoint from the sweepout spheres. After the singular time, each component of N_t contains a 2-sphere.

This can be modelled by $x^2 + y^2 + z^2 = t$, for $t \in [-1, 1]$. The singular time is $t = 0$. The time reverse of this is a vanish move.

(2) **Cut and paste moves** (Index 1 and 2)

In a cut move, a sphere orbit splits into two. In a paste move, two sphere orbits join together. At the singular time two sphere orbits share a single point in common. Each of the images of this pair of spheres also has a common point.

The green sphere intersects ∂N in a pair of circles. Before the singular time, the pair of circles bound a properly embedded annulus in N. After the singular time the circle bound a pair of disjoint discs.

This can be modelled by $z = x^2 + y^2 + t$, for $t \in [-1, 1]$. The singular time is $t = 0$. The time reverse of this is a paste move.

Definition 2.34 *Cut disc*

A cut move can be specified by giving an embedded disc in S^3, whose boundary is contained in a single sweepout sphere, whose interior is disjoint from the sweepout spheres, and which is disjoint from its images under G. We call this the *cut disc*.

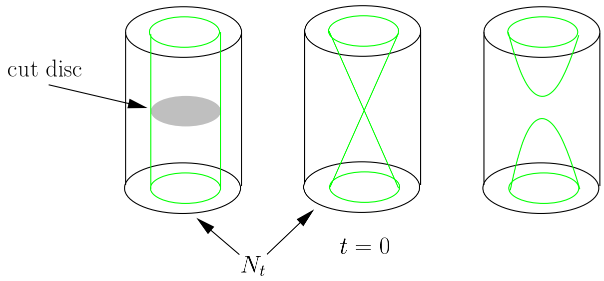


Figure 3: A cut move

Critical points of the height function on the double set The double set is 2-dimensional, so the height function has three sorts of critical points, births, deaths and saddles.

(1) **Birth/death of double curves** (index 0 and 2)

At the critical time, two spheres of different colours intersect in a single point in each of the components of N_t.

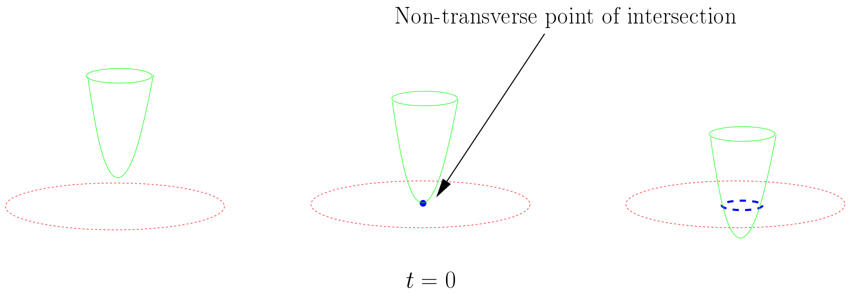


Figure 4: Birth of a blue double curve

The sweepout surfaces intersect one of the components of ∂N_t in a red circle and a green circle, which bound a red and a green disc inside that component of N_t, respectively. Before the singular time, the two discs are disjoint. At the singular time the two discs intersect at a single point. After the singular time the two discs intersect in a single (blue) double curve.

This can be modelled by choosing the green surface to be $z = x^2+y^2$, and the red surface to be $z = t$, for $t \in [-1, 1]$. The singular time is $t = 0$. A death move is the time reverse of this.

(2) **Saddle moves** (index 1)
During a saddle move, either a single double curve orbit splits into two double curve orbits, or two double curve orbits join together to form a single double curve orbit. At the critical time, each of the components of N_t contains the singular point of a figure eight double curve of intersection between two spheres of different colours.

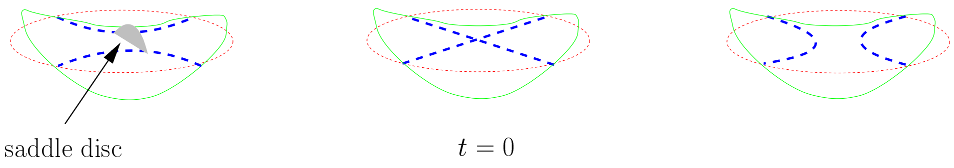


Figure 5: A saddle move

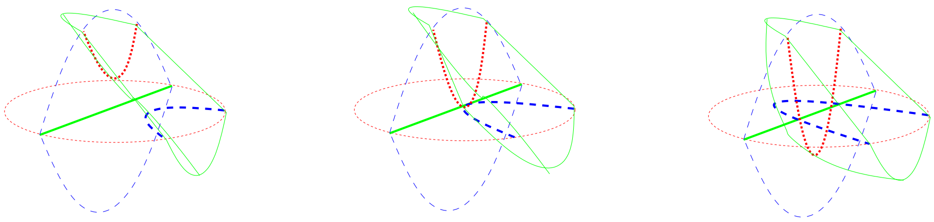
The component of N_t shown above contains two properly embedded discs, one of which is red and the other is green. The boundaries of the discs are two circles in ∂N_t which intersect at four distinct points. The points of intersection occur in the same order on each circle, so each point is adjacent to the same two points, whether you travel along the red circle or the green circle.
Before the singular time, the green disc intersects the red discs in a pair of double arcs, which connect two pairs of adjacent points in this component of ∂N_t.
After the singular time, the discs intersect in a pair of double arcs that connect each point to the other adjacent point in ∂N_t.
This can be modelled by choosing the green surface to be $z = x^2 - y^2$, and the red surface to be $z = t$, for $t \in [-1, 1]$. The singular time is $t = 0$. The time reverse of this is also a saddle move.

Definition 2.35 *Saddle disc*

The saddle move can be specified by giving the orbit of a disc which is disjoint from its images, properly embedded in $(S^3, G \,.\, S_t)$, whose interior is disjoint from the sweepout surfaces, and whose boundary consists of two arcs, one of which lies in the red sphere, and the other lies in the green sphere, which meet at the blue double arcs. This is illustrated in Figure 5. The saddle move can be thought of as pushing the green sphere through the red sphere along this disc.

Critical points of the Morse function on the triple point set

Triple point births and deaths (index 0 and 1) In a triple point move, a pair of triple points is either created or destroyed. The triple point set is one dimensional, so there are two sorts of singularities, births and deaths of triple points.



$t = 0$

Figure 6: A triple point birth

The component of N_t shown above contains three properly embedded disc, one of each colour. The boundary of each disc is a circle in ∂N_t. Each pair of discs intersects in a single arc, and the three arcs are disjoint, so that each disc contains two disjoint double arcs.
Before the triple point birth, each disc intersects the other discs in a pair of disjoint double arcs. At the singular time the three double arcs intersect in a common point. After the singular time each double arc contains two triple points.
This can be modelled by choosing the red surface to be $z = 0$, the blue surface to be $x = 0$, and the green surface to be $z = y^2 - x + t$, where $t \in [-1, 1]$. The singular time is at $t = 0$. A triple point death is the time reverse of this.

Definition 2.36 *Football*

A *football* is a region homeomorphic to a 3-ball in S^3, whose boundary consists of one double curve free bigons of each colour.

The images of a football under G are disjoint, this is an immediate consequence of Corollary 2.51. A triple point birth move creates a football orbit, and a triple point death move destroys a football orbit.

Definition 2.37 *Isotopy layer*

Let (M, ϕ) be a sweepout, let I be a closed subinterval of $\mathbb{R}$, and let $(h^{-1}(I), \phi)$ be the corresponding partial sweepout.

We say that $(h^{-1}(I), \phi)$ is an *isotopy layer*, if there are no moves during the time interval I. We say that I is an *isotopy interval.*

Definition 2.38 *Move layer*

Let (M, ϕ) be a sweepout, and let I be a closed subinterval of $\mathbb{R}$. Suppose the partial sweepout $(h^{-1}(I), \phi)$ is a union of two partial sweepouts with disjoint interiors, one of which is a move, and the other is a partial sweepout which is a product. Then we say that $(h^{-1}(I), \phi)$ is a *move layer.*

Definition 2.39 *Move position*

Let (M, ϕ) be a sweepout in general position, together with a choice of move neighbourhood for each move, so that the sweepout is a sequence of isotopy layers and move layers. Then we say that the sweepout is in *move position.*

Lemma 2.40 *Given a sweepout in general position, we can choose move neighbourhoods for every move, so that the sweepout is in move position.*

Proof The critical points of the sweepout occur at distinct times, so we can choose move neighbourhoods which are disjoint in time. $\square$

We will often wish to choose tubular neighbourhoods for subsets of the sweepout surfaces that intersect the sweepout spheres in a way which is "as simple as possible". We now make this precise.

Definition 2.41 *Thin regular neighbourhood*

Suppose that Σ is a subset of the sweepout spheres at some time t.

Let $N \subset S^3$ be a regular neighbourhood of Σ. If the intersection of N with each of the sweepout spheres forms a regular neighbourhood in the sweepout spheres for Σ, then we say that N is a thin regular neighbourhood for Σ.

Remark 2.42 If Σ is a manifold, then we may choose a thin regular neighbourhood which is a tubular neighbourhood of Σ. We call this a *thin tubular neighbourhood* for Σ.

Lemma 2.43 *Let $(M_I, f\times h)$ be a partial sweepout with image neighbourhood N_I, which contains no moves, and let K_0 be an equivariant submanifold of N_0. Then we can extend K_0 to an equivariant continuously varying family of submanifolds K_I contained in N_I, such that the intersection of the sweepout surfaces with K_t only changes up to isotopy for $t \in I$.*

The idea is to think of the partial sweepout $(M, f \times h)$ as an isotopy between the images of M_0 and M_1, and then use the isotopy extension theorem.

Theorem 2.44 *Let B be a manifold, which may have boundary, and let A be a properly embedded compact submanifold of B. Let $f\colon A \times I \to B$ be an isotopy for which each f_t is a proper embedding. Then f extends to an ambient isotopy $F\colon B \times I \to B$.*

We now prove Lemma 2.43.

Proof The image neighbourhood N_I has a compatible product structure, so we can identify N_I with $N_0 \times I$, and think of $f\colon M_I \to N_0$ as an isotopy between $f(M_0) = S_0$ and $f(M_1) = S_1$ which are properly embedded. We first show how to extend this isotopy to an equivariant ambient isotopy of N_0.

By the isotopy extension theorem, the isotopy f extends to an isotopy $F'\colon N_0 \times I \to N_0$, which need not be equivariant. We can choose a continuously varying family of tubular neighbourhoods U_t of S_t. As there are no moves, we can choose this neighbourhood to be sufficiently small, so that $G\,.\,U_t$ is a tubular neighbourhood for $G\,.\,S_t$.

Then $G\,.\,U_t$ is a 3-manifold on which G acts freely, so $G\,.\,U_t/G$ is also a 3-manifold. Furthermore $G\,.\,U_I/G$ is an isotopy between $G\,.\,U_0/G$ and $G\,.\,U_1/G$ in N_0/G, so by the isotopy extension theorem this extends to an isotopy F'' on N_0/G. Let F be the lift of F'' to N_0. The F is an equivariant ambient isotopy that extends f.

Now let K_I be the image of K_0 under the equivariant ambient isotopy F. The intersection of K_t with the sweepout surfaces only changes by isotopy, by construction. □

Finally we show that every component of M is either S^3 or a connect sum of $S^2 \times S^1$'s.

Proposition 2.45 *Let M be a closed 3-manifold, and let $h\colon M \to S^3$ be a Morse function, which has the property that at non-singular times the level sets are unions of 2-spheres.*

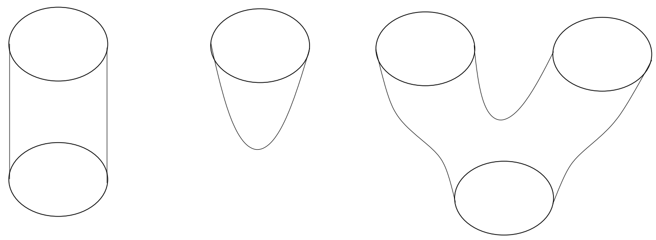
Then each component of M is either S^3 or a connect sum of $S^1 \times S^2$'s.

Proof Choose times t_i between each of the critical times of the height function. The pre-images of these times is a collection of embedded 2-spheres in M. They divide M into pieces which contain at most one singular point of a singular level set. So $M - \Sigma$ consists of pieces of the following three types:

(1) If there is no critical point of M in a component of $M - \Sigma$, then the component is a product $S^2 \times I$.

(2) If a component of $M - \Sigma$ contains a critical point of index 0 or 3, in which a sphere either appears or disappears, then it is a 3-ball.

(3) If a component of $M - \Sigma$ contains a critical point of index 1 or 2, in which a single sphere splits into two, or two spheres join together, then it is a 3-ball with two 3-balls removed.



No critical point   Index 0 or 3 critical point   Index 1 or 2 critical point

Figure 7: Components of $M - \Sigma$

In each case, capping off the 2-sphere boundaries of the piece with 3-balls produces a copy of S^3, so M is either S^3, or a connect sum of $S^2 \times S^1$'s. □

2.4 Complexity

Definition 2.46 *Complexity*

We say that the *complexity* of the sweepout at the generic time t is the ordered pair (n, d), where n is the number of triple points in the sweepout at time t, and d is the number of simple closed double curves with no triple points at time t. We order the pairs lexicographically.

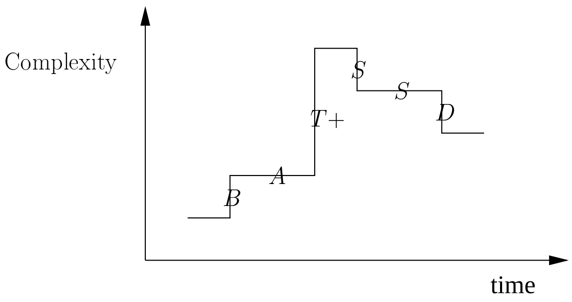
Triple points come in multiples of six, double curves come in multiples of three.

Definition 2.47 *The graphic*

The graphic is the graph of complexity against time, with labels added to show which move occurs. We will mark every move, even if it does not change complexity, so an unmarked interval in the graphic corresponds to a time interval in which no moves occur.

Move	Symbol
Appear	A
Vanish	V
Cut	C
Paste	P
Birth	B
Death	D
Saddle	S
Triple point birth	T+
Triple point death	T-

(a) Key to symbols in the graphic



(b) The graphic

Figure 8: The graphic

2.5 Orientations

We have chosen a continuously varying normal vector to S_t, which defines an inside and outside of the red spheres. We think of the inside as coloured red. This also gives us normal vectors for the images gS_t and g^2S_t, and we have coloured their insides green and blue respectively.

Each triple point lies at the intersection of three surfaces, one of each colour, so the triple point lies in the boundary of regions shaded with all possible combinations of colours. In particular, each triple point lies in the boundary of a "clear" region, ie, a region on the "outside" of all of the spheres.

Choose tangent vectors to the double curves at the triple point, so that the tangent vectors point toward the clear region. The vectors have a cyclic ordering (an orientation) coming from the map g ie (red, green, blue).

Definition 2.48 *Positive and negative triple points*

If this orientation agrees with the orientation of S^3, we call the triple point *positive*, if it disagrees we call it *negative*.

Every triple point is therefore labelled either positive or negative, and triple points that are connected by double arcs containing no triple points in their interiors have opposite sign.

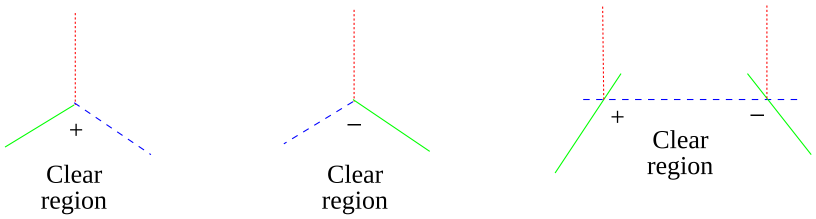


Figure 9: Signs of triple points

Definition 2.49 *Adjacent*

We say that two triple points are *adjacent* if they are connected by a double arc which contains no triple points in its interior.

Triple points are adjacent if they are connected by a red arc, so triple points may still be adjacent, even if they are not adjacent in the blue-green graph.

Lemma 2.50 *Adjacent triple points cannot lie in the same orbit under G.*

Proof The map f preserves the sign of the triple point. Adjacent triple points have opposite sign, so they can not be images of each other. □

Corollary 2.51 *A bigon orbit consists of a red bigon, a blue bigon and a green bigon, which have disjoint boundaries.*

Proof The vertices of a red bigon are adjacent triple points, so must have distinct orbits under G. Therefore the images of the arcs in the boundary of the red bigon have distinct endpoints, so they must be distinct also. □

2.6 Modifying sweepouts

Definition 2.52 *A modification neighbourhood*

Let (M,ϕ) be a sweepout in move position. Let N_I be a compatible product neighbourhood with constant boundary pattern. If $\partial(N_I)$ is disjoint from the chosen move neighbourhoods of the sweepout, then we say N_I is a *modification neighbourhood.*

Notation 2.53 If $P \subset M$, and ϕ is a map defined on M, then we will abuse notation and write ϕ to denote the restriction map $\phi|_P$.

Definition 2.54 *A pre-image sweepout*

Let (M,ϕ) be a sweepout in move position, and let N_I be a modification neighbourhood. Then $(\phi^{-1}(N_I),\phi)$ is a partial sweepout which we call the *pre-image partial sweepout for* N_I.

Every modification neighbourhood is the image neighbourhood of its pre-image sweepout.

Remark 2.55 We can construct a partial sweepout by drawing a sequence of configurations in N which differ by moves.

Definition 2.56 *Partial sweepouts with the same boundary*

Let (P,θ) and (P',θ') be two partial sweepouts with the same image neighbourhoods. If $\theta^{-1}\circ\theta'$ is a diffeomorphism between ∂P and $\partial P'$ then we say that the two partial sweepouts *have the same boundary.*

Theorem 2.57 *Let (M_1,ϕ_1) be a sweepout in move position, let N_I be a modification neighbourhood with pre-image sweepout (P_1,ϕ_1), and let (P_2,ϕ_2) be a partial sweepout with the same boundary as the pre-image sweepout. Assume that N_0 is not all of S^3. Define M' to be the manifold $(M_1-P_1)\cup P_2$, using the gluing map $\phi_1^{-1}\circ\phi_2$, and define ϕ' as follows:*

$$\phi' = g'\times h' = \begin{cases} \phi_2 \text{ on } P_2 \\ \phi_1 \text{ on } M_1 - P_1 \end{cases}$$

If all but finitely many level sets ${h'}^{-1}(t)$ are a union of 2-spheres, then (M',ϕ') is a sweepout.

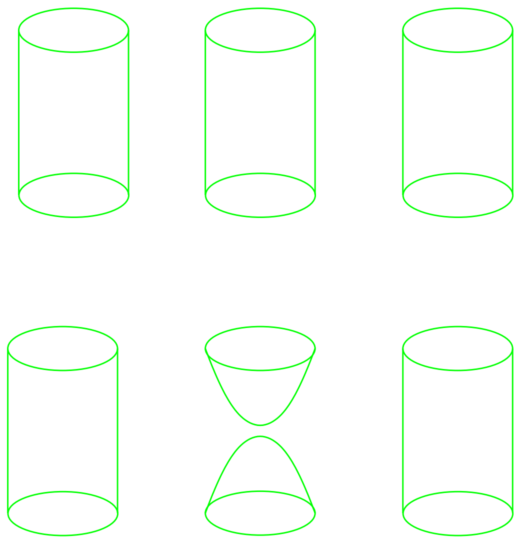


Figure 10: Two partial sweepouts with the same boundary

Proof It remains to show that $f' = \pi_{S^3} \circ \phi'$ is degree one. Let $f = \pi_{S^3} \circ \phi_1$. As N_t is not all of S^3 we may assume (perhaps after adjusting by an isotopy) that there is a point $p \in S^3$ which is disjoint from N_t for all $t \in I$. Then $f^{-1}(p)$ is the same as ${f'}^{-1}(p)$, so both maps have the same degree. □

Definition 2.58 *Modification*

Using the notation from Theorem 2.57 above, we say that (M', ϕ') is a *modification* of the sweepout (M, ϕ).

The following two modifications will be useful:

Removing a simple closed double curve which bounds a disc

Let (M, ϕ) be a sweepout in move position. Suppose for some time interval I we can choose a continuous family of double curves $\{\gamma_t \mid t \in I\}$, which bound a continuous family of discs $\{\Delta_t \mid t \in I\}$, which contain no double curves in their interiors, and which are disjoint from all the move neighbourhoods in the sweepout. By Lemma 2.43, we can choose a compatible product neighbourhood N_I so that each N_t is a thin tubular neighbourhood for the orbit of the disc Δ_t, and N_I is disjoint from the move neighbourhoods. This means that the intersection of each N_t with the sweepout surfaces consists of the orbit of a disc and an annulus, which intersect transversely in the orbit of the simple closed double curve γ_t.

Let (P, ϕ) be the pre-image sweepout for N_I. This is illustrated by the top row of pictures in Figure 11.

We can remove the double curve by doing a cut move parallel to Δ_t, and then a double curve death move. If we follow this pair of moves by its time reverse, a birth followed by a paste, we return the configuration to its initial state. This creates a new partial sweepout with the same boundary as (P, ϕ), but which does not contain any double curves for a subinterval of I. This is illustrated by the bottom row of pictures in Figure 11.

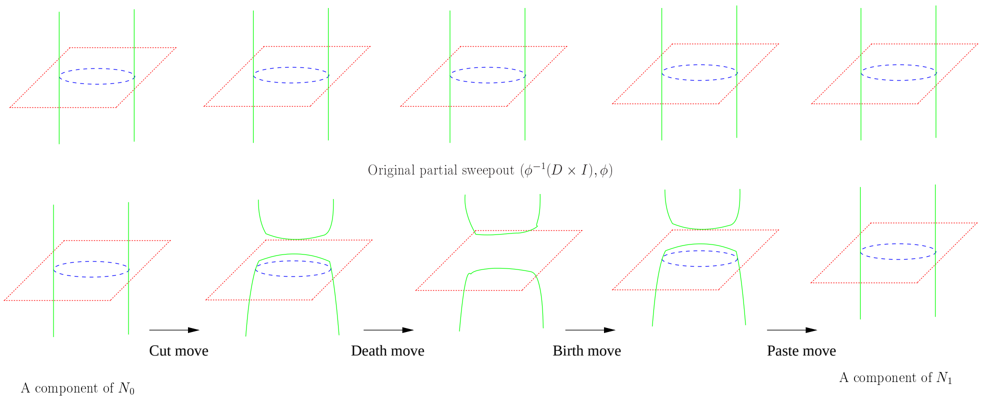


Figure 11: Removing a double curve which bounds a double curve free disc

We will call the pair of moves which removes the double curve a *compound double curve death*, and we call the pair of moves which replaces the double curve a *compound double curve birth*. No other moves may occur during a compound double curve move, but moves may occur in between the compound double curve birth and the compound double curve death.

Removing a double curve free bigon Let (M, ϕ) be a sweepout in move position. Suppose there is a time interval I for which we can choose a continuous family of bigons $\{A_t \mid t \in I\}$, so that A_t is always double curve free, and disjoint from the move neighbourhoods of the sweepout. By Corollary 2.51 such a bigon is disjoint from its images under G. By Lemma 2.43, we can choose a compatible product neighbourhood N_I disjoint from the move neighbourhoods, so that each N_t is a thin tubular neighbourhood for the bigon orbit $G.A_t$. As the three images of the bigon A_t are disjoint, N_t will be the orbit of a 3-ball, which is also disjoint from its images. So it will suffice to describe how to change the sweepout in a single component of N_I, as the changes can be extended equivariantly to

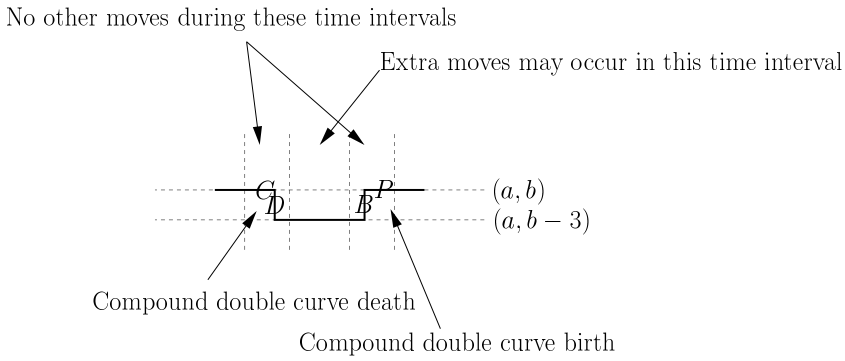


Figure 12: Removing a double curve changes the graphic

the other components.
Consider the component of N_t which contains the bigon A_t. This intersects the sweepout surfaces in three discs, one of each colour. One of these discs, which we shall think of as being "horizontal" is a tubular neighbourhood of the bigon A_t in the sweepout sphere which contains A_t. Each of the other discs, which we shall think of as "vertical", intersect the horizontal disc transversely in a single double arc. These double arcs contain the boundary of the bigon A_t. The vertical discs intersect each other in a pair of vertical double arcs. This is illustrated in the top row of pictures in Figure 13.
We can remove the bigon by pushing the vertical surfaces through each other sideways. In order to do this in general position, we first do a saddle move, using a saddle disc parallel to A_t. This saddles the vertical double arcs and creates a football containing A_t in its boundary. We can then do a triple point death move to destroy the football, removing the bigon, and reducing the number of triple points in this component by two, and by six in total. If we then do the time reverse of these two moves, we have created a partial sweepout with the same boundary as the original one, but which has fewer triple points for a subinterval of I. This is illustrated in the bottom row of pictures in Figure 13.
We will call the pair of moves that removes the bigon a *compound triple point death*, and the pair of moves that replaces the bigon a *compound triple point birth*. No other moves may occur during the pair of moves that make up a compound move. We will call triple point moves that are not part of compound triple point moves *non-compound* triple point moves.

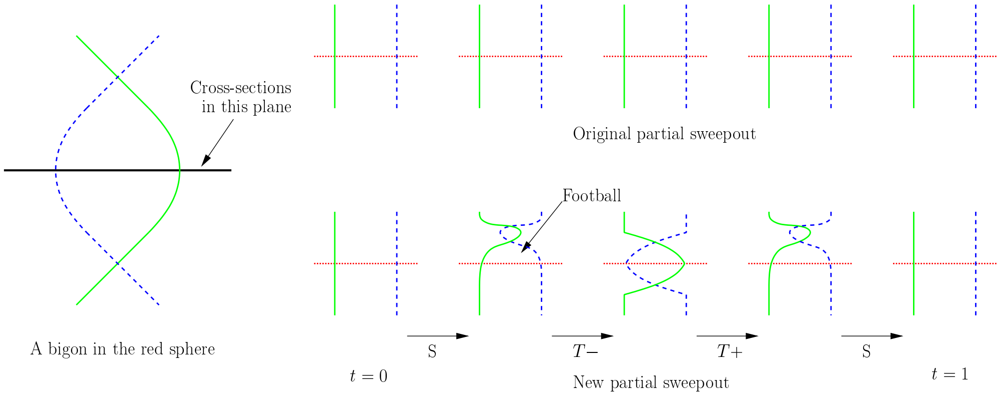


Figure 13: Removing a double curve free bigon

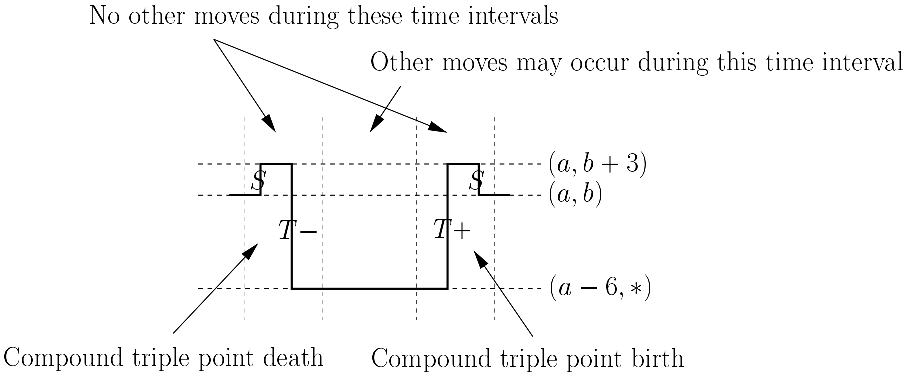


Figure 14: Removing a double curve free bigon changes the graphic

> The saddle move may leave complexity unchanged, or may increase it by $(0,3)$. We have illustrated the latter case above.

We can do one of these modifications whenever we can find a simple closed curve that bounds a disc, or a double curve free bigon, that is disjoint from the move neighbourhoods of the sweepout for some time interval. If this time interval contains a local maximum in the graphic, the effect of the modification will be to reduce the height of the local maximum. We say that we have *undermined* the local maximum.

In the discussion above, we always wrote $G.A_t$ to refer to a continuously varying family of subsets of the sweepout spheres defined for some time interval. In future, we will just write $G.A$ without the subscript to refer to such a family, if it is clear from context that we mean a family defined for a time interval.

3 Reduction to special cases

In this section we will prove the main theorem, assuming the following three lemmas, which we will prove in the remaining sections.

Lemma 4.1 *Every sweepout contains triple points.*

Lemma 5.1 *Suppose a non-triple point move occurs while the configuration contains triple points. Then there is a vertex free bigon orbit disjoint from the move neighbourhood of the move.*

Definition 3.1 *A special case local maximum*

We say that a local maximum in the graphic is a *special case local maximum* if it consists of a compound triple point birth, followed by a compound triple point death, with no other moves occurring in between.

Lemma 6.1 *A special case local maximum can either be undermined, or else contains an invariant unknotted circle.*

Lemma 5.1 enables us to use the two modifications described in the previous section to change all the local maxima into special case local maxima, so that we can apply Lemma 6.1. If we could undermine all of the local maxima, then we would have a sweepout with no triple points, contradicting Lemma 4.1, so there must be a special case local maximum we cannot undermine, which therefore contains an unknotted invariant curve.

So in order to prove the main theorem, it suffices to show that given a sweepout of complexity $\geqslant (6,0)$, we can change it to produce a new sweepout with lower complexity. Therefore the following lemma proves the main theorem, assuming the three lemmas above.

Lemma 3.2 *Suppose (M,ϕ) is a sweepout with maximum complexity $(p,q) \geqslant (6,0)$. Then either we can find a new sweepout with lower maximum complexity, or else there is an invariant unknotted curve.*

First we deal with the case where the complexity of the sweepout is (p,q), with $q>0$, and show that if there are no invariant unknotted curves we can reduce the complexity to $(p,q-3)$. We divide the proof into three steps, which we describe below. Finally, we show how to deal with the case $q=0$.

Step 1 *Undermine non-triple point moves*

Lemma 5.1 implies that we can repeatedly undermine every non-triple point move, until they all lie below $(p,0)$. This produces a sweepout with complexity at most $(p,q+3)$, but now all moves at complexities at least $(p,0)$ are either compound moves, or triple point births and deaths. $\boxed{\text{Step 1}}$

Step 2 *Replace all non-compound triple point moves with compound triple point moves*

We can replace a non-compound triple point death with the following sequence of moves. Choose one of the bigons of the football, call it A, and do a compound triple point death move using it. We can choose to do the saddle move of the compound triple point death move using a saddle disc which lies inside the football. This splits the football into a smaller football close to A, and a 3-ball region bounded by a pair of discs which meet in a simple closed double curve. The triple point death move of the compound move removes the football close to A, and we can now do a double curve death move to remove the simple closed double curve. This is illustrated below.

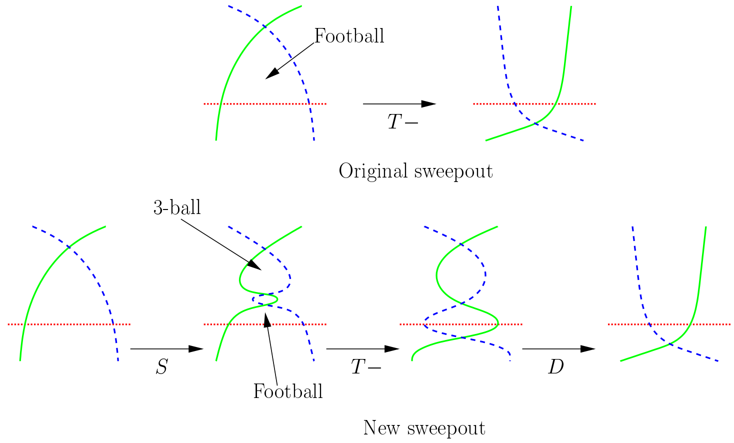


Figure 15: Replacing a non-compound triple point move with a compound triple point move

This does not increase complexity above $(p,q+3)$, as non-compound triple point moves have heights at most (p, q). This introduces non-triple point moves at heights strictly less than $(p, 0)$, but now all moves above $(p, q-3)$ are compound moves.

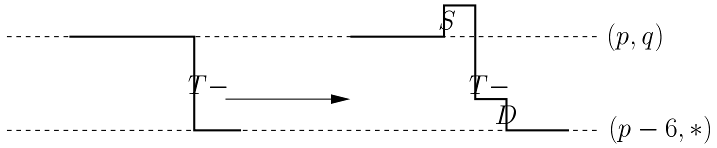


Figure 16: Replacing a non-compound triple point move with a compound triple point move

$\boxed{\text{Step 2}}$

Step 3 *Reducing compound local maxima*

We now consider the local maxima of complexity (p, q) or higher, and show we can change the sweepout to reduce them all to complexity at most $(p, q-3)$. First, in Case 1, we show how to undermine the compound local maxima of complexity $(p, q+3)$, so that they have complexity at most (p, q). In Case 2, we then show how to undermine compound local maxima of height (p, q), with $q > 0$. Finally, in Case 3, we show how to undermine compound local maxima of height $(p, 0)$.

Case 1 *A local maximum of complexity* $(p, q+3)$

A compound move always changes complexity, so the local maximum consists of one compound move, immediately followed by another compound move. If a local maximum has complexity $(p, q+3)$, then it must contain at least one compound triple point move, as these are the only compound moves at this height. The other move may be either another compound triple point move, or a compound double curve move. If the local maximum contains a compound double curve move, then there are two cases, depending on whether the compound double curve move occurs first, followed by the triple point move, or whether the compound triple point move occurs first, followed by the compound double curve move. The picture below shows all three possibilities.

The first case shown is a special case local maximum, which by Lemma 6.1, can either be undermined, or contains an invariant unknotted curve. It now suffices to show how to undermine the second case, as the third local maximum is just the time reverse of the second one.

Let $G.A$ be the bigon orbit of the bigon in the compound triple point move, and let Δ be the disc orbit of the disc in the compound double curve move.

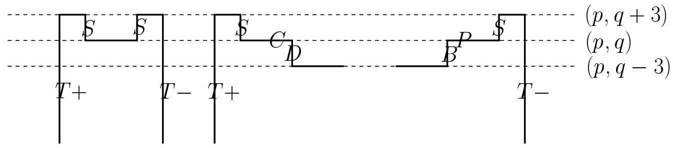


Figure 17: Local maxima of complexity $(p, q+3)$

The bigon orbit is double curve free, so $G \,.\, A$ must be disjoint from the disc orbit Δ, so we can just swap the order of the compound moves.

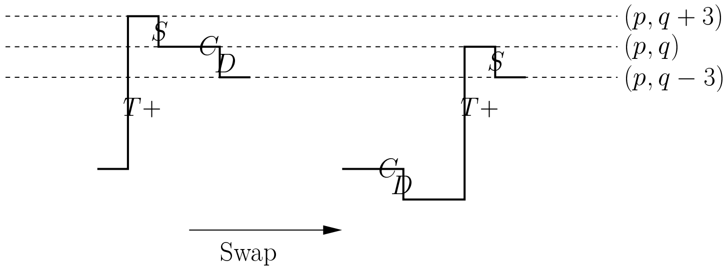


Figure 18: A compound triple point birth followed by a compound double curve death

We may now assume that we have reduced the complexities of all the local maxima to at most (p, q). $\boxed{\text{Case 1}}$

Case 2 *A local maximum at height* (p, q)*, with* $q > 0$

A compound move always changes complexity, so a local maximum of height (p, q) consists of one compound move, followed by another compound move. Each compound move may be either a compound double curve move, or a compound triple point move. The picture below shows all possible local maxima of complexity (p, q), up to time reverse. The time reverse of the second picture may also occur.

By Lemma 6.1, the first case, Figure 19(a), can either be undermined, or else the configuration contains an unknotted invariant curve. The second case, Figure 19(b), and its time reverse can be undermined by the argument given above in Case 1. For the remaining case, Figure 19(c), we show that it can either be undermined, or reduced to one of the first two cases.

The remaining case consists of a compound double curve birth followed by a compound double curve death. Let Δ_1 be the orbit of the double curve free disc in the first compound double curve move, and let Δ_2 be the orbit of the double curve free disc in the second compound double curve move.

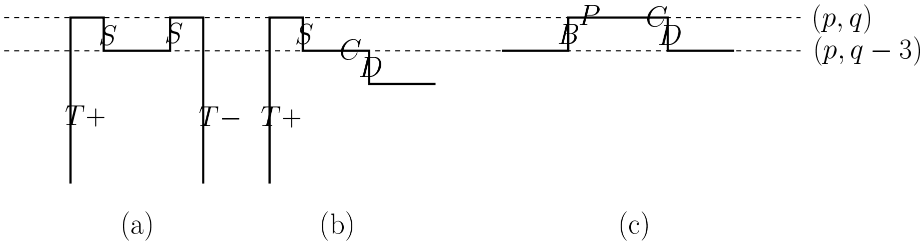


Figure 19: Local maxima of complexity (p, q)

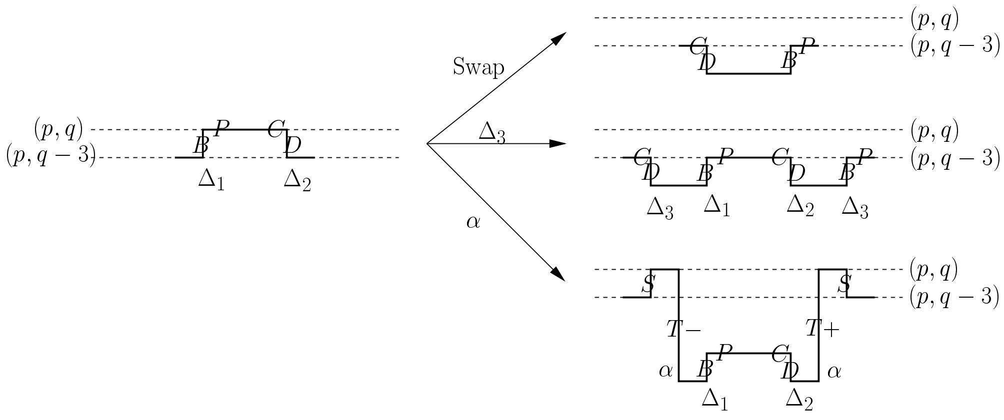


Figure 20: A compound double curve birth followed by a compound double curve death

If Δ_1 and Δ_2 are disjoint, then we can just swap the order of the compound moves. This undermines the local maximum.

If Δ_1 and Δ_2 are not disjoint, then they must share a common boundary, which is the orbit of a simple closed curve γ. The green and blue components of γ are contained in the red spheres, while the red component is disjoint from the red spheres. As there are triple points, by Lemma 5.5 there are at least four vertex free bigons with disjoint interiors, so there is a vertex free bigon orbit $G \centerdot A$ disjoint from the orbits of Δ_1 and Δ_2. So either $G \centerdot A$ is double curve free, or else it contains an innermost simple closed curve, which bounds a double curve free disc Δ_3.

If there is a simple closed double curve which bounds a double curve free disc Δ_3, which is disjoint from $\Delta_1 \cup \Delta_2$, then we can undermine the local maximum by removing the double curve for the duration of the local maximum, by applying a double curve removal modification.

If there is a double curve free bigon $G \centerdot A$, which is disjoint from $\Delta_1 \cup \Delta_2$, then we can modify the sweepout by inserting a compound triple point death, followed by a compound triple point birth, which removes the bigon for the duration of the local maximum. This reduces the complexity of the compound double curve local maximum, but may not reduce the overall complexity of the sweepout, as the extra compound triple point moves may still have maximum complexity (p, q). However, we can remove all local maxima consisting only of compound double curve moves in this way, and then the remaining local maxima of height (p, q) will have at least one compound triple point move. This means that they are all covered by the previous two cases, and so can be undermined. $\boxed{\text{Case 2}}$

Case 3 *A local maximum of height* $(p, 0)$

Finally, we explain how to deal with the case when the sweepout has complexity $(p, 0)$. In this case, the local maxima of greatest complexity do not contain any simple closed double curves without triple points. We can again use Lemma 4.1 to undermine the non-triple point moves of greatest complexity, but as there are no simple closed double curves without triple points (because $q = 0$), all the compound moves will be compound triple point moves. This means that the resulting sweepout may have local maxima of complexity $(p, 0)$ and $(p, 3)$, but all the local maxima at these heights will consist either of compound triple point moves or of non-compound triple point moves.

We can now apply Step 2 to replace all the non-compound triple point moves with compound triple point moves, and so now all local maxima are special case local maxima, and we can apply Lemma 6.1 directly. $\boxed{\text{Case 3}}$

This completes Step 3, the final step in the proof of Lemma 3.2. $\boxed{\text{Step 3}}$

We have now proved the main result, assuming Lemmas 4.1, 5.1 and 6.1 above. The remaining sections are devoted to proving these three lemmas.

4 Every sweepout contains triple points

In this section we prove the following lemma:

Lemma 4.1 *Every sweepout contains triple points.*

At the beginning of the sweepout there is a clear connected invariant region, which we will call an initial region. We show that if a sweepout has no triple points, then there is always an initial region. However, this gives a contradiction, as at the end of the sweepout there are no sweepout spheres, and S^3 is a single region coloured with all three colours, so there are no clear regions at all.

Definition 4.2 *An initial region*

We say that a region is *initial* if it is clear, connected and invariant under f.

Remark 4.3 The union of the clear regions gets mapped to itself by f, but need not be connected. Individual connected clear regions need not be invariant.

Lemma 4.4 *A sweepout without triple points contains initial regions at all times.*

Proof At the beginning of the sweepout, there is an initial region. This is because before the first appear move, there are no sweepout surfaces, and so all of S^3 is a clear connected invariant region, and so is initial.

We now complete the proof of the lemma by considering each type of move in turn, and showing that if there is an initial region before the move, then there must also be an initial region after the move, assuming that there are no triple points.

A single move is supported in the orbit of a 3-ball which is disjoint from its images, so a single move cannot eliminate an initial region by shrinking it down to a point. However a single move might eliminate an initial region by disconnecting it.

Case 1 *Appear and vanish moves*

If a double curve free 2-sphere orbit appears inside an initial region, then the interior of the 2-sphere orbit is a new region, namely a 3-ball orbit coloured by the same colour as the sphere. The region on the outside of the 2-sphere orbit is still connected, and so is still an initial region. $\boxed{\text{Case 1}}$

Case 2 *Cut and paste moves*

A paste move does not disconnect any region, so we need only consider cut moves. If a cut move disconnects an initial region C, then the cut disc D must be contained inside C. The images of D must also lie inside C, so we can find a path γ from D to gD which does not intersect any other sweepout surface. There are two cases, depending on whether or not the curve $\gamma \cup g\gamma$ intersects the cut disc transversely.

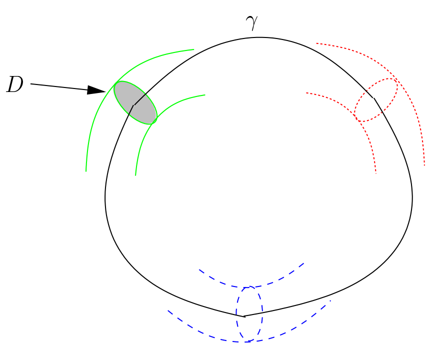


Figure 21: A cut move

In the right hand case the region containing $G\,.\,\gamma$ is still an initial region after the cut move. In the left hand case, the disc D divides the green sphere into two discs. The union of D with one of these discs is a sphere. As the simple closed curve $G\,.\,\gamma$ lies in the clear region it is disjoint from the green sphere, so it intersects this 2-sphere precisely once transversely, a contradiction, as $H_1(S^3) = 0$, so the left hand case cannot occur. $\boxed{\text{Case 2}}$

Case 3 *Double curve births and deaths*

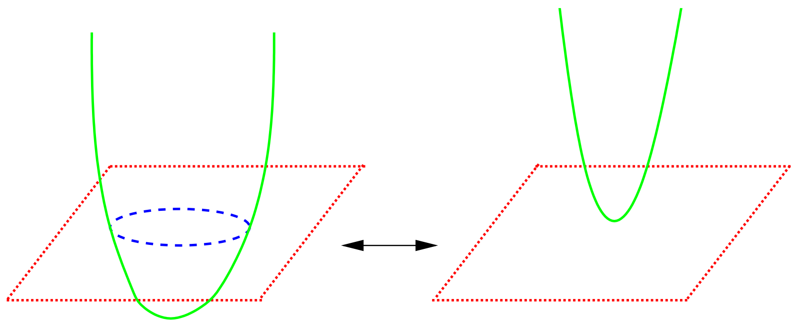


Figure 22: Birth/death of a blue double curve

The region that is created or destroyed in the move cannot be invariant. The other regions that intersect the move neighbourhood are not disconnected by the move. $\boxed{\text{Case 3}}$

Case 4 *Saddle moves*

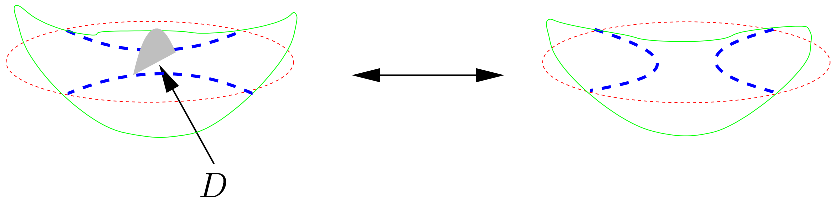


Figure 23: A saddle move

The only region which could become disconnected as a result of a saddle move is the region containing the saddle disc D in the diagram above. If the disc D intersects two distinct blue circles, then as there are no triple points, we can choose a path from one side of D to the other, in the interior of the initial region, parallel to either one of the blue double curves. So in this case, the saddle move does not disconnect the initial region.

Now suppose that the disc D intersects a single blue double curve. The disc D has three disjoint images under G. The region is connected, so we can find a path γ from D to gD, which only intersects the discs in its endpoints, and is disjoint from all the sweepout surfaces. Then $G\,.\,\gamma$ is a simple closed curve that connects the three discs in one of the following two ways:

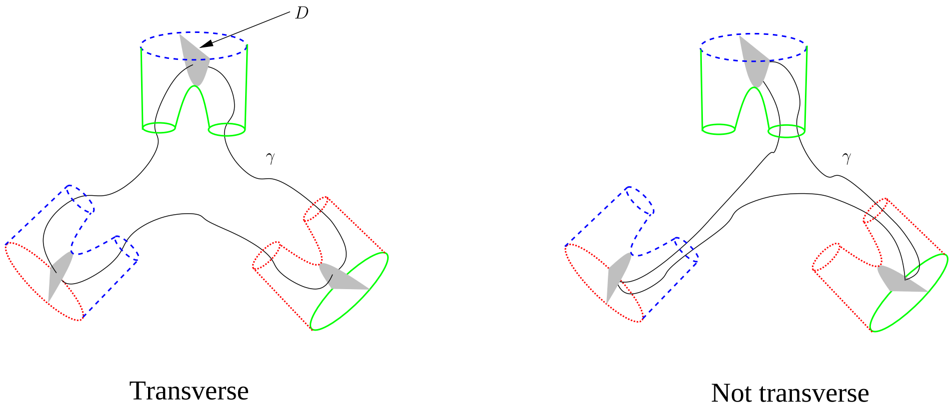


Figure 24: The invariant curve $G\,.\,\gamma$

In the one-sided case, the region containing $G\,.\,\gamma$ is still an initial region after the saddle move. We now show that the transverse case cannot arise.

The green double curve, which we will call α, divides one of the red spheres into two discs. Consider the disc which contains part of the boundary of the saddle disc D. This disc is divided into two subdiscs by the saddle arc. As the orbit of the saddle disc $G\,.\,D$ intersects the red spheres in precisely two arcs, at most one of the subdiscs of the red disc intersects an image of the saddle disc D. Let D' be one of the red subdiscs whose interior is disjoint from $G\,.\,D$. This is illustrated below. The red sphere is drawn as horizontal, and the blue sphere intersects it transversely in the double curve α.

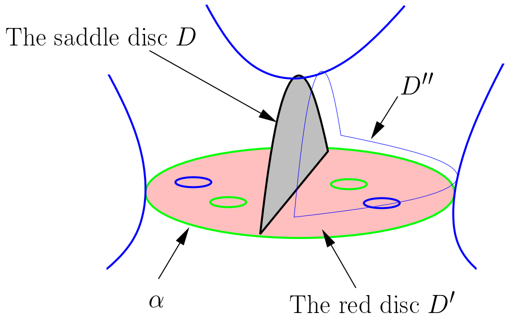


Figure 25: The disc D'' is parallel to $D \cup D'$

The union of D and D' form a disc whose boundary lies in one of the blue spheres. Let D'' be a disc parallel to $D \cup D'$. We may choose this disc so that it is disjoint from the double curves, and from $G\,.\,D$, so it intersects the sweepout surfaces in simple closed curves only. Cut move do not disconnect the initial region, by Case 2, so we may remove the intersections of the sweepout surfaces with D''. We may need to change the path γ if it passes through a cut disc, however we may always choose a new path from D to gD, which we shall also call γ, as cut moves do not disconnect the initial region. Now the boundary of $D \cup D'$ bounds a blue disc with no double curves in its interior. However, the union of this blue disc with the red disc D', and the saddle disc D, is a 2-sphere, which the curve $G\,.\,\gamma$ intersects precisely once transversely, a contradiction. This shows that the transverse case cannot arise. $\boxed{\text{Case 4}}$

Therefore there is always an initial region after a move in a sweepout without triple points. This completes the proof of Lemma 4.4, which in turn completes the proof of Lemma 4.1, that every sweepout must contain triple points. $\square$

5 Disjoint bigons for non-triple point moves

In this section we prove the following lemma:

Lemma 5.1 *Let N_I be a move neighbourhood for a non-triple point move in a sweepout. Then there is a vertex free red bigon disjoint from N during the time interval I.*

This means that the orbit of this red bigon will also be disjoint from the move neighbourhood of the move. The configuration outside the move neighbourhood only changes by an isotopy during the move interval, so it suffices to find a bigon disjoint from the move neighbourhood at a given time during the move interval. We show that a configuration with triple points has at least four red bigons, and that a move neighbourhood can't intersect all of these at once.

Definition 5.2 *A red pseudo-bigon*

A *red pseudo-bigon* is a closed disc in the red sphere whose boundary consists of one green arc and one blue arc, which may contain triple points in their interiors.

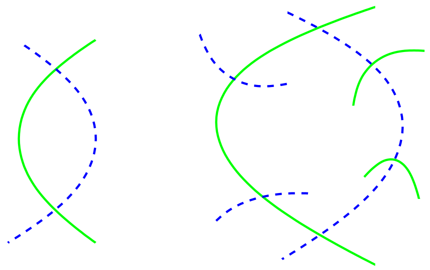


Figure 26: Red pseudo-bigons

Proposition 5.3 *If there are triple points, then the blue-green diagram contains at least four pseudo-bigons with disjoint interiors.*

Proof Consider a green curve which contains triple points. This divides a red sphere into two discs with disjoint interiors. Each disc must contain at least one blue arc, which divides each disc into a pair of pseudo-bigons with disjoint interiors. □

Lemma 5.4 *An innermost pseudo-bigon is a vertex free bigon.*

Proof Choose an innermost pseudo-bigon. If a boundary arc of this pseudo-bigon contains triple points in its interior, then we can find a double curve arc inside the pseudo-bigon with both endpoints on the same boundary arc which forms a smaller pseudo-bigon, a contradiction. So the boundary arcs of an innermost pseudo-bigon contain no triple points in their interiors.

If an innermost pseudo-bigon has no triple points in the interiors of its boundary arcs, but triple points in its interior, then the interior contains a pair of intersecting circles, which create a pseudo-bigon, again giving a contradiction. Therefore an innermost pseudo-bigon has no triple points in its interior, and has boundary arcs with no triple points in their interiors, so is a vertex free bigon. □

Corollary 5.5 *If a red sphere contains triple points, then the red sphere contains at least four vertex free red bigons with disjoint interiors.*

Remark 5.6 Two vertex free disjoint red bigons may have translates which intersect in simple closed double curves.

We now prove Lemma 5.1.

Proof of Lemma 5.1 We deal with the possible non-triple point moves one by one.

Case 1 *There is always a bigon orbit disjoint from a vanish or appear move*

At the singular time t the move neighbourhood for a vanish or appear move contains only the orbit of a singular sphere, consisting of the orbit of a single point. The bigon orbits must all be contained in different sweepout spheres, so they are all disjoint from the move neighbourhood. $\boxed{\text{Case 1}}$

Case 2 *There is always a bigon orbit disjoint from a cut or paste move*

At the singular time t the move neighbourhood for a cut or paste move contains two discs of the same colour, which meet at a single point in their interiors. These discs are contained in the interiors of at most two vertex free bigon orbits, so there is a disjoint bigon orbit, as there are at least four vertex free bigon orbits by Corollary 5.5. $\boxed{\text{Case 2}}$

Case 3 *There is always a bigon orbit disjoint from a birth or death double curve move*

At the singular time t the move neighbourhood for a birth or death move contains two discs of different colours, which intersect at a single point. The orbit of these discs contains two red discs, which can be contained in at most two different vertex free red bigons. Therefore there is a red bigon disjoint from N, and so the orbit of this red bigon is a bigon orbit which is also disjoint from N. $\boxed{\text{Case 3}}$

Case 4 *There is always a bigon orbit disjoint from a saddle move*

At the move time t, the red spheres contain blue and green double curve circles, and precisely one figure eight double curve of each colour, which are in the same orbit under G. We will call the green figure eight curve γ. The blue figure eight curve is then $g\gamma$. All of these double curves may contain triple points.

The orbit of the move neighbourhood for the saddle move intersects the red sphere in a pair of discs, one of which contains the singular point of γ, and one of which contains the singular point of $g\gamma$. Furthermore, the intersection of the blue and green double curves with the discs consists only of a pair of green arcs which cross at the green singular point, and a pair of blue arcs which cross at the blue singular point, so it suffices to find a red bigon which is disjoint from the singular points of γ and $g\gamma$.

The green figure eight curve γ divides the red sphere in which it lies into three surfaces. Precisely two of these surfaces have closures which are discs, call these discs D and D'. Let D'' be the closure of the complement of $D \cup D'$.

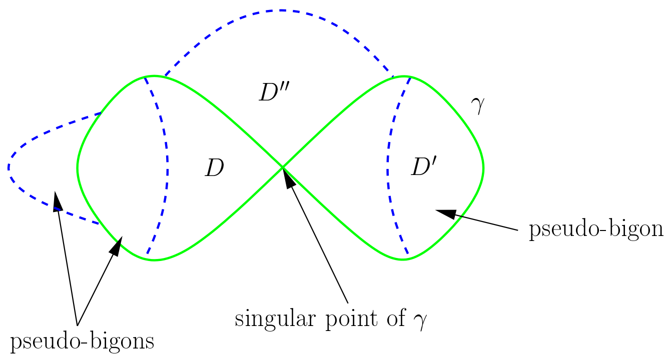


Figure 27: The figure-eight curve

We consider the following three cases in turn: there is a blue arc in D'' with both endpoints on the same component of $\gamma - \{\text{singular point}\}$, there is a blue arc in D'' with endpoints on different components of $\gamma - \{\text{singular point}\}$, or else there are no blue arcs in D''.

If there is an arc in D'' with both endpoints on the same component of γ, then this arc creates a pseudo-bigon in D''. However, there must also be a blue arc in one of the discs D or D', creating a pseudo-bigon whose interior is disjoint from the one in D''. At most one of these pseudo-bigons can contain the blue singular point of $g\gamma$, so there is a pseudo-bigon disjoint from the move, and hence a vertex free bigon disjoint from the move.

If there is an arc in D'' with endpoints on both D and D', then there must be blue double arcs in both D and D', so each disc contains a pseudo-bigon. At most one of these pseudo-bigons can contain the blue singular point of $g\gamma$, so there is a pseudo-bigon disjoint from the move, and hence a vertex free bigon disjoint from the move.

If there are no blue arcs in D'', then γ is triple point free, so the green and blue figure eight curves do not intersect any other double curves, and so may be contained in at most two vertex free red bigons. As there are at least four vertex free red bigons, there is a vertex free red bigon disjoint from the move neighbourhood. $\boxed{\text{Case 4}}$

This completes the proof of Lemma 5.1, that there is always a vertex free red bigon disjoint from a non-triple point move. □

6 Undermining the special cases

Recall that a special case local maximum consists of a compound triple point birth followed by a compound triple point death. In this final section we complete the proof of the main result by proving the following theorem:

Theorem 6.1 *Given a special case local maximum, we can either undermine the local maximum, or find an invariant unknotted curve.*

This means that if there are no invariant unknotted curves, we can remove every local maximum, so the minimax sweepout contains no triple points, a contradiction, by Lemma 4.1.

Consider a special case local maximum. Let $G \centerdot A$ be the bigon orbit which is created by the compound triple point birth move, and let $G \centerdot B$ be the bigon destroyed in the compound triple point death move. At a time during the local maximum in between the compound moves, both bigons are present in the sweepout, and are double curve free. There are only finitely many ways in which such a pair of bigons may intersect, and we deal with each possibility in turn.

There is the trivial case when $G \,.\, A$ and $G \,.\, B$ are the same bigon orbit. In this case the configurations before and after the local maximum are isotopic, so we can replace the partial sweepout containing the moves that occur during the local maximum with a partial sweepout that is an isotopy and contains no moves, thus undermining the local maximum. There are three remaining cases to consider depending on how many vertices the bigon orbits $G \,.\, A$ and $G \,.\, B$ have in common. If $G \,.\, A$ and $G \,.\, B$ have no vertices in common, then they must be disjoint, so we can just postpone the first compound move till after the second compound move, thus undermining the local maximum. The remaining two cases to consider are when $G \,.\, A$ and $G \,.\, B$ have either three or six vertices in common.

Section 6.1 deals with the case in which $G \,.\, A$ and $G \,.\, B$ have three vertices in common. In this case we give explicit constructions of modifications which undermine the local maxima.

Section 6.2 deals with the case in which $G \,.\, A$ and $G \,.\, B$ have six vertices in common. If $G \,.\, A \cup G \,.\, B$ is connected, we produce an invariant unknotted curve. If $G \,.\, A \cup G \,.\, B$ is not connected, then we show how to undermine the local maximum.

We now fix some notation which we will use for the rest of this section.

Definition 6.2 *Special case modification neighbourhood*

Suppose (M, ϕ) is a sweepout containing a special case local maximum. Suppose that N_I is a modification neighbourhood with the following properties:

- N_I contains the move neighbourhoods for the compound triple point moves in its interior, and is disjoint from all the other move neighbourhoods of the sweepout.
- N_t is a tubular neighbourhood for $G \,.\, A_t \cup G \,.\, B_t$ for the times t in between the compound triple point moves.

Then we say that N_I is a *special case modification neighbourhood* for the special case local maximum.

The special case modification neighbourhood may be connected or disconnected, and may have components that are not 3-balls. The sweepout only changes by an isotopy outside the special case modification neighbourhood during the local maximum.

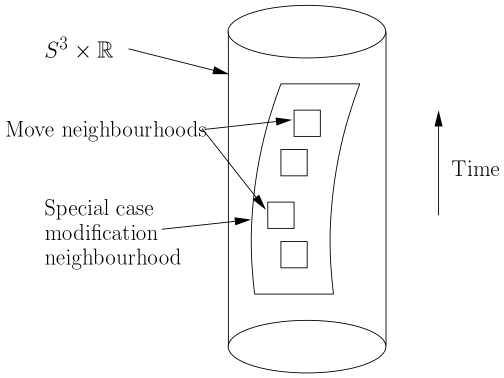


Figure 28: Schematic picture of the special case modification neighbourhood

Lemma 6.3 *Let* (M, ϕ) *be a sweepout containing a special case local maximum. Then we can choose a special case modification neighbourhood for the given special case local maximum.*

Proof Choose move neighbourhoods for the local maximum, and label the corresponding time intervals as shown below.

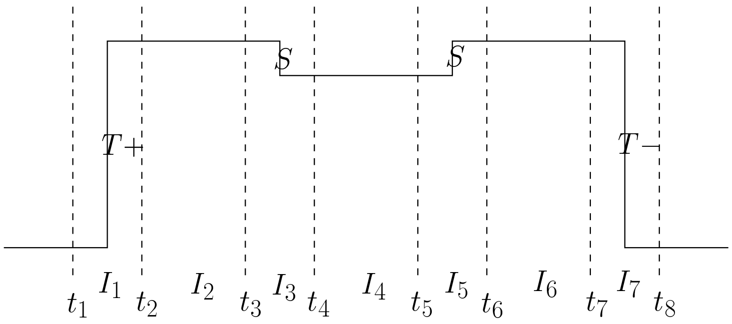


Figure 29: The graphic for a special case local maximum

We will show how to construct the special case modification neighbourhood going forward in time. A similar construction will work going back in time. Let J_{I_5} be the move neighbourhood for the second saddle move. Let K_{I_7} be the move neighbourhood for the triple point death move.

Choose a thin tubular neighbourhood for $G.A \cup G.B$ at time t_4. The pre-image sweepout for $S^3_{I_4}$ is a product, so by Lemma 2.43 we can choose a compatible product neighbourhood N_{I_4} so that N_t is a thin tubular neighbourhood for $G\,.\,A \cup G\,.\,B$ for each $t \in I_4$.

The saddle move occurring during the time interval I_5 is determined by a choice of saddle disc D at time t_5. The saddle move neighbourhood J_{t_5} is a tubular neighbourhood of this disc D at time t_5. The saddle disc D is parallel to the bigon $G\,.\,B$, which is contained in N_{t_5}, so at time t_5 we can choose the thin tubular neighbourhood N_{t_5} of $G\,.\,A \cup G\,.\,B$ to be large enough to contain J_{t_5}.

Let $N'_{t_5} = N_{t_5} - J_{t_5}$. The partial sweepout $(S^3 \times I_5) - J_{I_5}$ is a product, so by Lemma 2.43 we can extend N'_{t_5} to a continuously varying family N'_{I_5}, so that the intersection of each N_t with the sweepout surfaces is isotopic to the intersection of the sweepout surfaces with N_t. In particular, if we define N_{I_5} to be $N'_{I_5} \cup J_{I_5}$, then as ∂N_t is contained in $S^3 - J_t$, the boundary pattern only changes up to isotopy in ∂N_t for the time interval I_5.

There are no moves during the time interval I_6, so by Lemma 2.43, we can extend N_I across this interval, so that the sweepout in N_t only changes up to isotopy.

As N_{t_6} contains the football created by the saddle move, N_{t_7} will also contain this football. At time t_7, the move neighbourhood for K_{t_7} is a tubular neighbourhood of this football, so we can choose N_{t_7} to be large enough to contain K_{t_7}. Now by using Lemma 2.43 in the same way as for the saddle move, we can extend N_I across the time interval I_7, so that N_I has constant boundary pattern. □

It will be convenient to fix notation for the different components of the orbit of each bigon.

Notation 6.4 Let A be the red bigon component of $G\,.\,A$, and B the red bigon component of $G\,.\,B$. We will write N^A for the component of N that contains the red bigon A.

6.1 Three vertices in common

In this section we show:

Lemma 6.5 *A special case local maximum in which $G\,.\,A$ and $G\,.\,B$ have exactly three vertices in common can be undermined.*

We shall prove this by giving explicit constructions of partial sweepouts, with the same boundary as the pre-image sweepout for the special case modification neighbourhood, but which contain fewer triple points.

The red bigon A will be disjoint from its images under g, though it will intersect images of the red bigon B. There are two different ways in which the bigon-orbits can share three vertices. Case 1 is if the red bigons A and B have a vertex in common, we will call this the *coplanar* case. Case 2 is if the red bigon A shares a vertex with the green bigon gB. We will call this the *orthogonal* case. If the red bigon A shares a vertex with the blue bigon g^2B, then the red bigon B shares a vertex with the green bigon gA, so this is just the time reverse of Case 2. In both of these cases we will show that we can modify the sweepout to reduce the local maximum.

Case 1 *Coplanar*

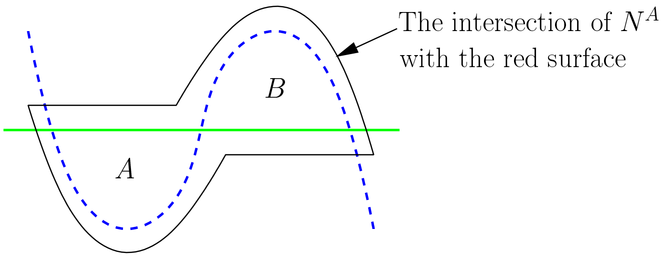


Figure 30: Coplanar red bigons in the red sphere

In this case N_I is the orbit of a 3-ball which is disjoint from its images. Each component of N_I contains an image of $A \cup B$ under G. It suffices to describe the modification in the component N_I^A of N_I that contains the red bigons $A \cup B$, we do the corresponding equivariant modifications in the other components of N_I. During the local maximum, we may assume that the red and blue surfaces remain fixed, and only the green surface moves. The red and blue spheres intersects N^A in a pair of discs, which intersect in a single double arc. It is convenient to give N^A a $\{\text{disc}\} \times I$ product structure, so that the red disc is horizontal, and the blue disc is vertical. This is shown below:

We can choose the product structure so that during the central product interval, the green surface is also vertical. This is shown below:

In order to describe the partial sweepout for the modification, it will be convenient to draw pictures of N^A split into four parts, along the red and blue surfaces. We will call these pieces *quadrants*.

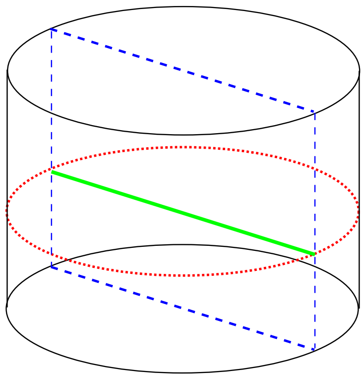


Figure 31: Case 1: The red and blue surfaces in N^A

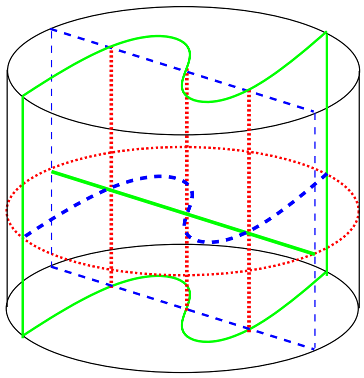


Figure 32: Case 1: The green surface in N^A during the local maximum

We can now construct the configurations before and after the local maximum. Figure 34(a) shows the configuration before the local maximum. It is produced from the configuration in Figure 33 by removing the bigon A. Figure 34(b) shows the configuration after the local maximum. It is produced from the configuration in Figure 33 by removing the bigon B.

We now describe a new partial sweepout with the same boundary as the pre-image sweepout for the special case modification neighbourhood, which we will use to modify the sweepout. Notice that the configurations in Figure 34(a) and Figure 34(b) are not isotopic, as for example, the top right quadrant contains different numbers of discs.

During the original partial sweepout, the components of the green surface in each quadrant are discs, and so the configuration is determined (up to isotopy)

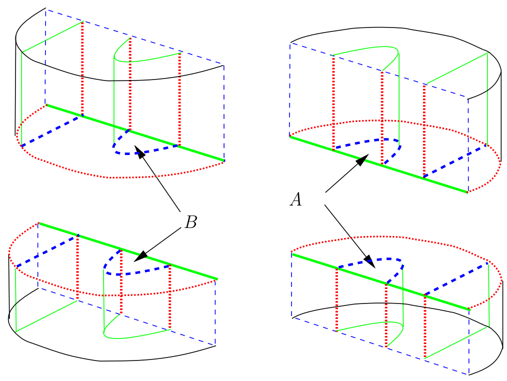
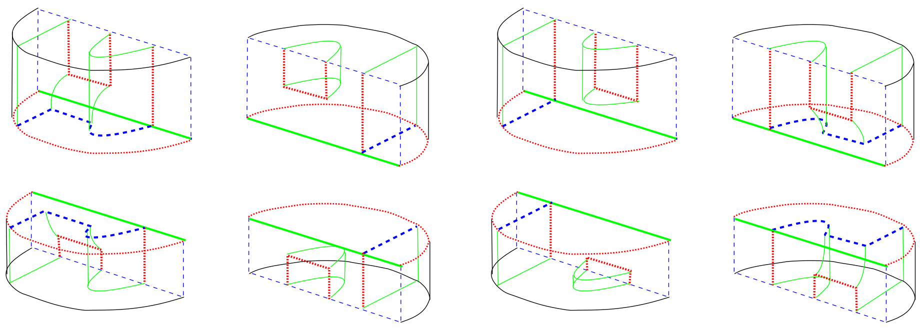


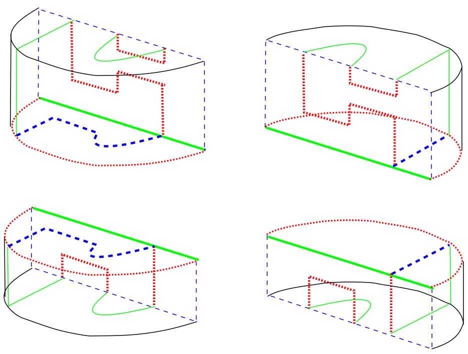
Figure 33: The configuration between the compound moves



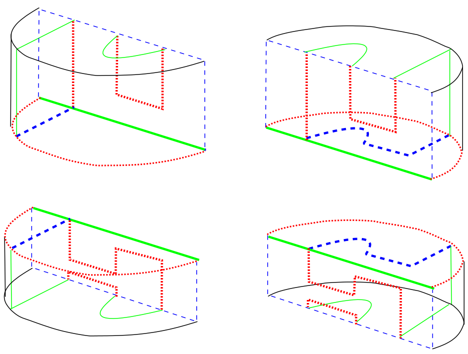
(a) Before the local maximum (b) After the local maximum

Figure 34: The original partial sweepout in N^A

by the boundaries of these discs. We can make the boundary curves isotopic by applying appropriate saddle moves to the configuration. Figure 34(a) shows the result of applying a saddle move parallel to A to Figure 33. Figure 34(b) shows the result of applying a saddle move parallel to B to Figure 33. This preserves the fact that the green surfaces in each quadrant are discs, so it suffices to check that the configurations are isotopic on the boundaries of the quadrants. The configurations on the boundaries of each quadrant are clearly isotopic, and furthermore, we can do these isotopies so that they agree on the surfaces which are identified. Also, the curves consist of the same combination of coloured arcs, so we can do an isotopy without changing the number of coloured arcs, so there will always be exactly one triple point.



(a) After a saddle move parallel to B



(b) Before a saddle move parallel to A

Figure 35: The new partial sweepout in N^A

So we can construct a partial sweepout starting at the configuration in 34(a), and then applying a saddle move to create the configuration in Figure 35(a). This is then isotopic to Figure 35(b), and we can then apply another saddle move to create the final configuration Figure 34(b). This partial sweepout has the same boundary as the initial partial sweepout in N^A, but has only one triple point, so we can use it to modify the sweepout to undermine the local maximum. $\boxed{\text{Case 1}}$

Case 2 *Orthogonal*

We now consider the case in which the images of A and B which share a vertex have different colours. We may assume that A shares a vertex in common with gB. If A shares a vertex with g^2B, then gA shares a vertex with B, so we could just swap the labels on A and B. The neighbourhood N is then the orbit of a 3-ball which is disjoint from its images, each component of which contains an image of $A \cup gB$.

Let N^A be the component of N which contains $A \cup gB$. It suffices to describe how to modify the sweepout inside N^A only. Figure 37 shows the configuration in N^A at a time in between the compound moves. We have drawn the red disc as horizontal, and the green disc as vertical. The red and green discs intersect in a single double arc.

In the picture above we have N^A as a thin tubular neighbourhood of a horizontal bigon and a vertical bigon. The picture below is isotopic to the one above,

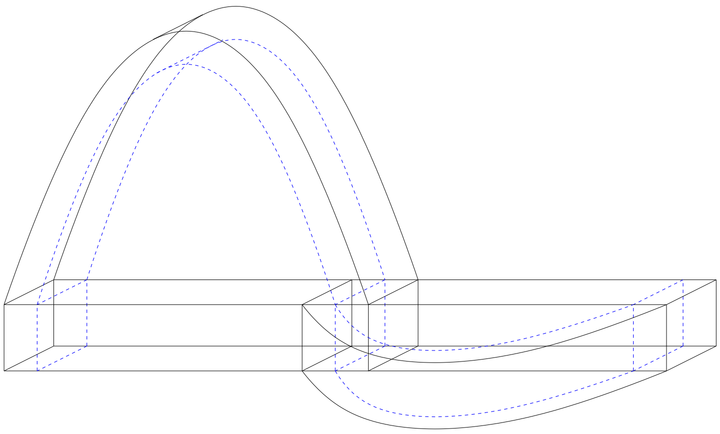


Figure 36: Case 2: The blue surface inside N^A

except we have drawn N^A as $\{\text{disc}\} \times I$, with the red surface as horizontal, and the green surface as vertical.

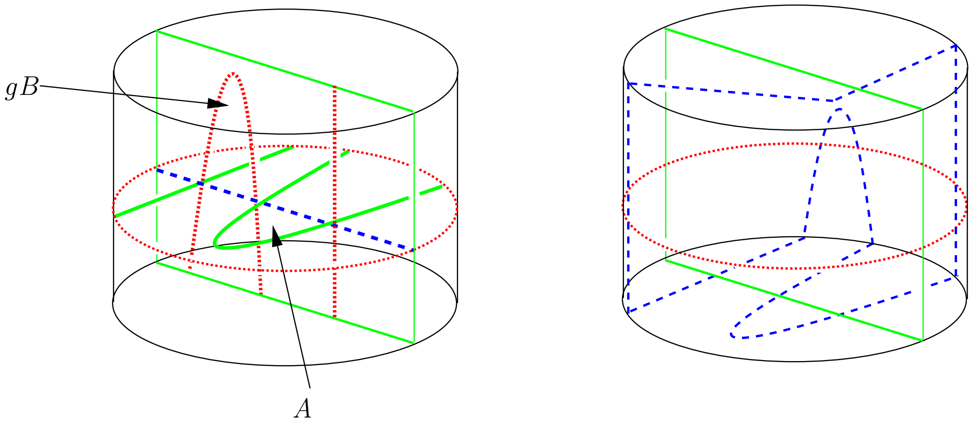


Figure 37: Case 2: N^A. This picture is isotopic to the one above.

As in the previous case, it will be convenient to draw pictures of N^A divided into four quadrants by cutting it along the red and green surfaces, as shown in Figure 38.

We may assume that the red and green surfaces in N^A remain fixed, and only the blue sphere moves, so going forward in time, we remove the green bigon gB by pushing the blue surface down, as shown in Figure 39(b), and going back in time, we remove the red bigon A by pushing the blue surface sideways, as

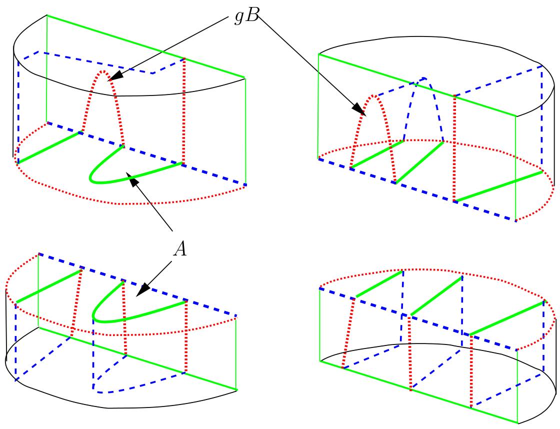
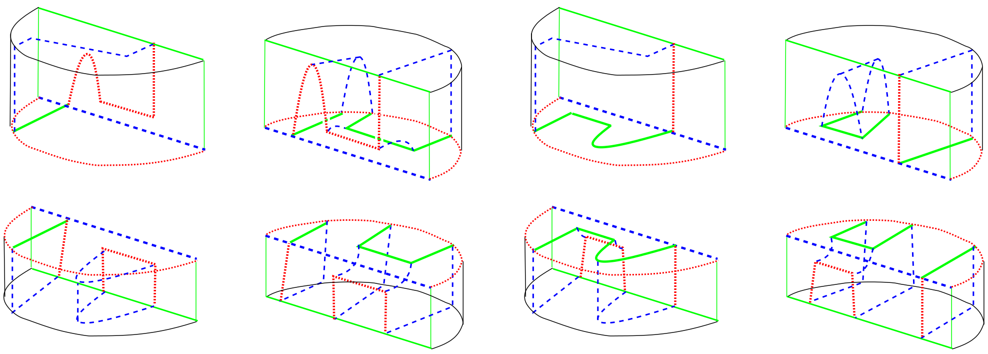


Figure 38: Case 2: N^A

shown in Figure 39(a).

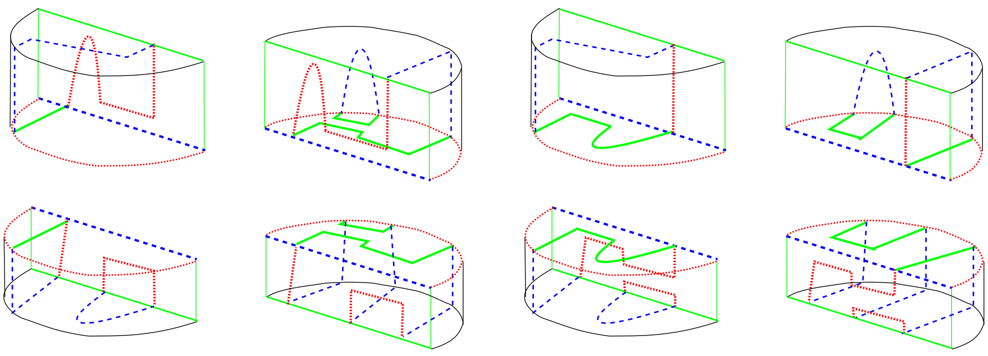


(a) Before the local maximum (b) After the local maximum

Figure 39: The original sweepout, after an isotopy

Figure 39(a) and Figure 39(b) are not isotopic, as for example, the top right quadrant contains different numbers of blue discs. However, by applying an appropriate saddle move to Figure 39(a) and Figure 39(b), we can create a pair of isotopic configurations.

We now describe the new partial sweepout. Start with the initial configuration shown in Figure 39(a). A saddle move is then applied to the green double curves, using a saddle disc parallel to the green disc gB. The resulting configuration is shown in Figure 40(a). As all the components of the blue surface in each quadrant are still discs, it suffices to draw their boundaries. Figure 40(a) is isotopic to Figure 40(b), as the boundaries of the blue discs in each quadrant are isotopic. Furthermore, we can do these isotopies so that they agree on the surfaces in the boundary of each quadrant which are identified, and without changing the number of arcs of each colour. In particular, this means that there is only one triple point in the new partial sweepout in N^A. Figure 40(b) is produced from Figure 39(b) by doing a saddle move in the bottom quadrants, using a saddle disc parallel to red disc A, so we end at the correct final configuration.



(a) After the first saddle (b) Before the second saddle

Figure 40: The new partial sweepout

Case 2

6.2 Six vertices in common

In this section we deal with the final case, when the bigon orbits $G.A$ and $G.B$ in the special case local maximum have all six vertices in common.

There are three different cases to consider depending on how many edge orbits the bigon orbits $G.A$ and $G.B$ have in common. If they have both edge orbits in common, then $G.A$ and $G.B$ are in fact the same bigon orbit, and the configurations before and after the local maximum are isotopic, so the local maximum can be undermined by replacing the sweepout in the special case modification neighbourhood by an isotopy. The remaining two cases are when

$G\,.\,A$ and $G\,.\,B$ share one edge orbit in common, which we discuss in Section 6.2.1, or no edge orbits in common, which we discuss in Section 6.2.2.

In order to undermine these local maximum, we shall prove a general undermining lemma, which we now describe. Let N_I be a modification neighbourhood in which N is the orbit of a 3-ball which is disjoint from its images. We call an intersection of a double curve with ∂N a *double point* of the boundary pattern. A *boundary bigon* is a disc in ∂N whose boundary consists of exactly two arcs of intersection with the sweepout surfaces, which contain no double points in their interiors. Given a boundary bigon, we can use this disc as a saddle disc for a saddle move, which simplifies the boundary pattern by removing a pair of double points. We say a boundary pattern is *saddle reducible* if all double points can be removed by a sequence of saddle moves using boundary bigons.

Given a modification neighbourhood N_I which has a saddle reducible boundary pattern, and no triple points in N_0 and N_1 we show that we can construct a new partial sweepout with the same boundary as the original one, but which contains no triple points for the whole time interval.

In some cases, the special case modification neighbourhood will have saddle reducible boundary, and we can apply the lemma directly. In other cases, N will not be the orbit of a 3-ball, but we will show how to extend the neighbourhood to a larger one which is a 3-ball with a saddle reducible boundary pattern.

We now give a precise definition of what it means for a boundary pattern to be saddle reducible, and we show that if a modification neighbourhood has a saddle reducible boundary pattern then we can replace the partial sweepout in this neighbourhood by one which contains no triple points.

Definition 6.6 *Double points*

We say a transverse intersection of a double curve with ∂N is a *double point* in ∂N.

Definition 6.7 *A boundary-bigon*

A *boundary-bigon* is a disc in ∂N whose boundary consists of a pair of arcs of different colours, with no double points in their interiors.

Remark 6.8 Given a simple closed curve which bounds an innermost disc in the boundary pattern, we can use this as a cut disc for a cut move, which removes the curve of intersection. Similarly, given a boundary-bigon in the boundary pattern, we can use it as a saddle disc for a saddle move, producing a boundary pattern with two fewer double points.

Definition 6.9 *Saddle reducible*

Let N_I be a modification neighbourhood, where N is the orbit of a 3-ball which is disjoint from its images. We say that the boundary pattern of N_I is *saddle reducible* if there is a time t at which there is sequence of cut moves and saddle moves, using cut discs and saddle discs contained in ∂N_t, which remove all the curves in the boundary pattern of N_t.

Remark 6.10 The boundary pattern of a modification neighbourhood N_I only changes up to isotopy, so if it is saddle reducible at some time $t \in I$, it will be saddle reducible for all times $t \in I$.

Not all boundary patterns are saddle reducible, for example:

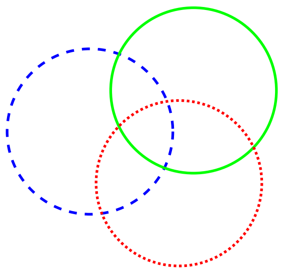


Figure 41: This boundary pattern is not saddle reducible

Lemma 6.11 *Let (M, ϕ) be a sweepout. Let N_I be a modification neighbourhood, where N_t is the orbit of a 3-ball which is disjoint from its images. Suppose that N_I contains no triple points at the beginning and the end of the time interval, and has a saddle reducible boundary pattern.*

Then there is a new sweepout (M, ϕ), which is the same as (M, ϕ) outside of N_I, and which has no triple points in N_I.

Proof We construct a partial sweepout with the same boundary as the pre-image sweepout for N_I, but which contains no triple points. Let C_I be a subset of N_I, such that C_I has a compatible product structure, and $C_t \cong G . S^2 \times [0,1]$ is a collar neighbourhood for ∂N_t, so that the sweepout surfaces intersect C_t in a product $\{\text{boundary pattern}\} \times [0,1]$. We may assume that ∂N_t is $G . S^2_t \times \{0\}$. Let C'_I be the "outer half" of C_I, so that $C'_t \cong G . S^2 \times [0, \frac{1}{2}]$.

As the intersection of $C_t \times \{s\}$ with the sweepout surfaces is the same for all $s \in [0,1]$, the closure of $N_t - C'_I$ also has saddle reducible boundary pattern, so there is a sequence of cut moves and saddle moves which make the sweepout surfaces disjoint from $G \,.\, S^2 \times \{\frac{1}{2}\}$. As these moves use cut discs and saddle discs contained in $G \,.\, S^2 \times \{\frac{1}{2}\}$, we may assume that the move neighbourhoods of the moves are contained in C_I. The sphere orbit $G \,.\, S^2 \times \{\frac{1}{2}\}$ is now the boundary of a 3-ball orbit containing no triple points. We can remove all the sweepout spheres inside the 3-ball orbit by first removing double curves using cut and death moves, and then removing the remaining disjoint spheres by vanish moves.

We can construct a similar partial sweepout starting with the final configuration instead. Use the same sequence of cut moves and saddle moves as before, to make the sweepout surfaces disjoint from $G \,.\, S^2 \times \{\frac{1}{2}\}$. The resulting collection of closed 2-spheres inside the three ball orbit bounded by $G \,.\, S^2 \times \{\frac{1}{2}\}$ may be different, but remove them all using cut, death and vanish moves. The resulting configuration will be the same as the one above, so we can use the first partial sweepout, followed by the time reverse of the second, to create a partial sweepout with the same boundary as $(\phi^{-1}(N_I), \phi)$, but which has no triple points. □

The lemma below shows that if there are not many double points, we do not need to explicitly construct the boundary pattern.

Lemma 6.12 *If the boundary pattern has at most four double points, then it is saddle reducible.*

Proof If the boundary pattern has only two double points, then all the curves in the boundary are simple, except for a single pair of curves that intersect each other twice, creating four bigons. We can remove any simple closed curves of intersection from the bigons by using cut moves. Then using any of these bigons as a saddle disc removes the double points.

Suppose there are four double points. If all the double points lie on a single pair of curves, as shown in Figure 6.2(a), then there is a bigon we can use to reduce the number of double points. This reduces us to the previous case, so this boundary pattern is saddle reducible. If the double points lie on two disjoint pairs of curves, Figure 6.2(b), then the boundary pattern is also saddle reducible. Otherwise, two of the double points lie on a single pair of curves, Figure 6.2(c), at least one of which contains another double point. However, curves must contain even numbers of double points, so this curve must in fact contain four double points, coming from two distinct intersecting curves, so there is a bigon we can use to reduce the number of double points. □

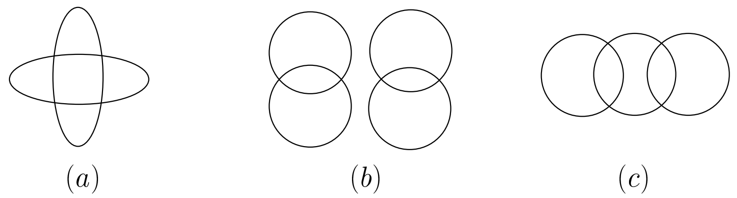


Figure 42: All possible boundary patterns with four double points

In some cases N_t will not be an orbit of a 3-ball, and in these cases we will show that we can extend the neighbourhood to produce a larger neighbourhood which is the orbit of a 3-ball. We will then examine the boundary pattern, and show that in each case it is saddle reducible.

The following lemmas will allow us to make some simplifying assumptions about the configuration outside N_t during the local maximum.

Lemma 6.13 *Suppose (M, ϕ) is a sweepout containing a special case local maximum, with bigon-orbits $G \cdot A$ and $G \cdot B$, and special case modification neighbourhood N_I. Then $N_t \cap S_t$ has at most three components.*

Proof The neighbourhood N_t is the union of two regular neighbourhoods of bigon-orbits. Each component of a regular neighbourhood of a bigon-orbit intersects the sweepout surfaces in three discs, one of each colour. So a regular neighbourhood of $G \cdot A$ intersects the red spheres S_t in three discs, as does a regular neighbourhood of $G \cdot B$. However, the bigon-orbits $G \cdot A$ and $G \cdot B$ have a common vertex-orbit, so each red disc must intersect at least one other red disc, so there are at most three components of $N_t \cap S_t$. □

Lemma 6.14 *Suppose (M, ϕ) is a sweepout containing a special case local maximum, with bigon-orbits $G \cdot A$ and $G \cdot B$, and special case modification neighbourhood N_I. If there is a component of the blue-green graph disjoint from N, then there is a vertex-free bigon-orbit disjoint from N_I.*

Proof Every component of the blue-green graph contains at least four red pseudo-bigons with disjoint interiors. By Lemma 6.13, $N_t \cap S_t$ has at most three components, and N_t may intersect at most three of these red pseudo-bigons, so there is a pseudo-bigon disjoint from N_t. This pseudo-bigon contains a vertex-free red bigon disjoint from N_t. As the configuration in $S^3 - N_t$ only changes by an isotopy, there is therefore a continuously varying family of disjoint bigons for the time interval I. □

Lemma 6.15 *Suppose (M, ϕ) is a sweepout containing a special case local maximum, with bigon-orbits $G \,.\, A$ and $G \,.\, B$. Suppose there is a vertex-free red bigon C at some time during the local maximum, which contains some simple closed double curves. Then we can modify the sweepout to remove the simple closed double curves during the corresponding local maximum in the new sweepout.*

Proof The double curves in the interior of C are disjoint from the move neighbourhoods of the moves occurring during the special case local maximum, so we can remove them for the duration of the local maximum by doing cut and death moves before the compound triple point birth, and then birth and paste moves after the compound triple point death. $\square$

Finally, we prove a lemma which will be useful in showing that certain sets are disjoint from their images under G.

Lemma 6.16 *Let $A_1, \ldots, A_n$ be double curve free bigons of any colours, such that each bigon A_i has the same vertices as A_1. Let A be the union of these bigons. Then A is disjoint from its images under G.*

Proof If any image of one of the A_i intersects A, then as the bigons are double curve free, it must intersect A in at least one vertex. But this means that there is a bigon which shares a vertex with one of its images, a contradiction, by Corollary 2.51. $\square$

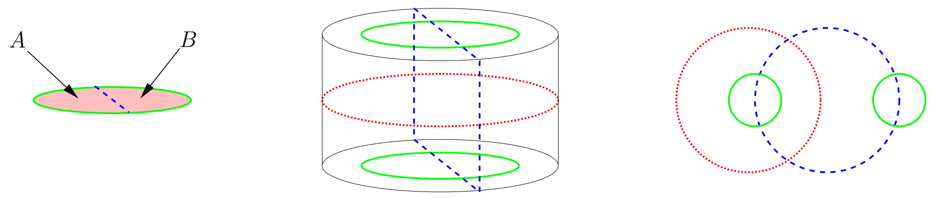
This lemma guarantees that we can choose a regular neighbourhood of such a set of bigons which is disjoint from its images.

6.2.1 One edge in common

There are two cases to consider, coplanar and orthogonal. In the coplanar case the red bigons A and B share a common edge. In the orthogonal case the red bigon A shares a common edge with either the green or blue image of B.

Case 1 *Coplanar*

In this case $A \cup B$ is a red disc which is disjoint from its images under G. In the picture below, A and B share a green edge in common. If they share a blue edge in common, then the picture is the same, but with the colours blue and green reversed.



The red bigons A and B share an edge

The component of N_t containing A and B

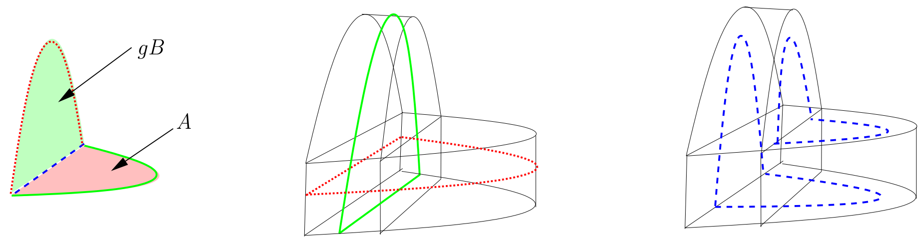
The boundary pattern

Figure 43: N^A

The orbit of N^A is a union of three disjoint 3-balls, with boundary pattern shown in Figure 43, which is saddle reducible, so the local maximum can be undermined by Lemma 6.11. $\boxed{\text{Case 1}}$

Case 2 *Orthogonal*

We may assume that the red bigon A shares an edge in common with the green bigon gB. If A shares an edge with the blue bigon g^2B, then we can reduce to the previous case by swapping the labels on A and B.



The red bigon A shares an edge with the green bigon gB

Intersection of the red and green spheres with ∂N^A

Intersection of the blue sphere with the boundary of N^A

Figure 44: N^A

The orbit of N^A is three disjoint 3-balls, and N^A has a saddle reducible boundary pattern, so we can undermine the local maximum by Lemma 6.11. $\boxed{\text{Case 2}}$

6.2.2 No edges in common

There are two different ways in which the bigon-orbits $G\,.\,A$ and $G\,.\,B$ may have six vertices in common, but no edges in common. This is illustrated in the

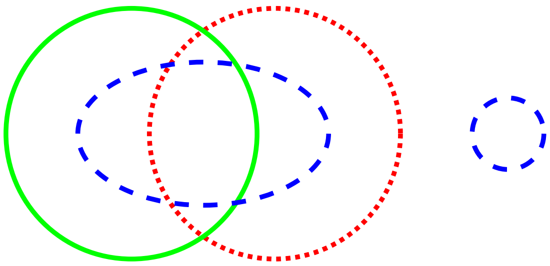


Figure 45: The boundary pattern

picture below. We have coloured tubular neighbourhoods of one bigon black, and tubular neighbourhoods of the other bigon pink.

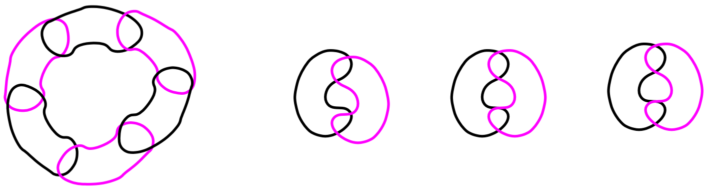


Figure 46: The bigon-orbits $G\,.\,A$ and $G\,.\,B$ may share six vertices in two different ways

If the union of the bigon-orbits is connected, we can show that the action of G is standard.

Lemma 6.17 *Let $G\,.\,A$ and $G\,.\,B$ be a pair of bigon-orbits which have all six vertices in common, and with $G\,.\,A \cup G\,.\,B$ connected. Then there is an invariant unknotted curve.*

Proof Let γ be the orbit of the green edges of A and B. Then γ is an invariant simple closed curve consisting of a pair of arcs of each colour. The blue and green arcs lie in the red sphere. The red arcs lie in the boundary of bigons whose other edge lies in the red sphere, so we can isotope them into the red sphere across these bigons, so γ is unknotted, as it is isotopic to a curve which lies in a sphere. □

So the remaining situations we need to analyse are when N_I has three connected components. There are two main cases to consider, the coplanar case, when the red bigons A and B have a pair of vertices in common, and the orthogonal

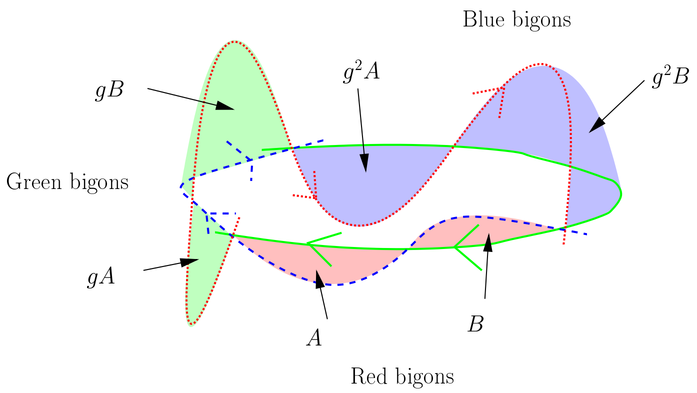


Figure 47: An example of an unknotted invariant curve

case, when the red bigon A has a pair of vertices in common with either the green bigon gB or the blue bigon g^2B.

Case 1 *Coplanar*

In this case the red bigons A and B have two vertices in common, but no edges in common. Then N^A contains a simple closed green curve, which we will call a, and a simple closed blue curve, which we shall call gb.

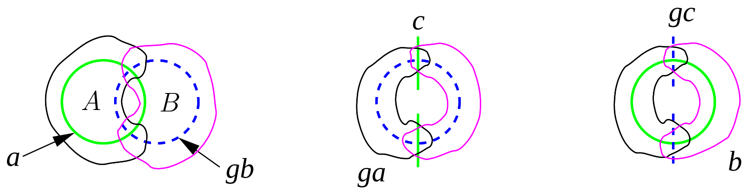


Figure 48: The intersection of N with the red sphere

In the picture above, we have drawn a thin tubular neighbourhood of $G\,.\,A$ in black, and a thin tubular neighbourhood of $G\,.\,B$ in pink.

The double curves $b \cup ga$ divides the red spheres into components, at least two of which are discs. Only one of these discs can intersect N^A, so one of the curves bounds a disc Δ in a red sphere which is disjoint from N^A. If there are vertex-free bigon-orbits disjoint from N^A, then we can use them to undermine the local maximum, so we may assume that this disc cannot contain triple points in its interior. This means that Δ contains a single simple double arc. By pinching off any simple closed double curves in Δ for the duration of the local maximum, we may change the sweepout so that Δ contains no simple closed double curves in its interior.

We may assume that b bounds the disc Δ, the case where ga bounds the disc Δ is the same, but with the colours blue and green swapped round.

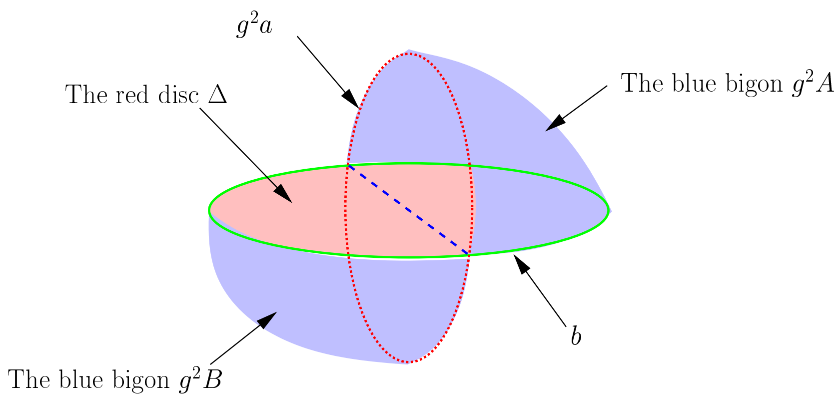


Figure 49: The red disc Δ, together with the blue bigons g^2A and g^2B.

The surface $\Delta \cup gA \cup gB$ shown above is disjoint from its images under G, by Lemma 6.16, as it is a union of double curve free bigons which share the same vertices. Therefore we can choose a thin tubular neighbourhood of the bigons which is a 3-ball containing N^A, which is also disjoint from its images under G. The boundary pattern for this 3-ball has exactly two double points, so will be saddle reducible, by Lemma 6.12. So the special case local maximum can be undermined by Lemma 6.11. $\boxed{\text{Case 1}}$

Case 2 *Orthogonal*

In this case the red bigon A shares both vertices in common with either the green or blue image of B. We may assume that A shares vertices with the green bigon gB. If A shares vertices with the blue bigon g^2B, then we can reduce to the previous case by swapping the labels on A and B.

The neighbourhood N^A contains a simple closed blue curve, call this gb. It also contains a green double arc, which we shall call a and a red double arc, which we shall call g^2c.

The tubular neighbourhood of $G \,.\, A \cup G \,.\, B$ intersects the red spheres in three annuli. These annuli divide the spheres into pieces, which we shall label $\Delta_1, \ldots, \Delta_6$, as shown in Figure 50. To be precise, note that each annulus in $N_t \cap S_t$ has a boundary component which is disjoint from the double curves, call

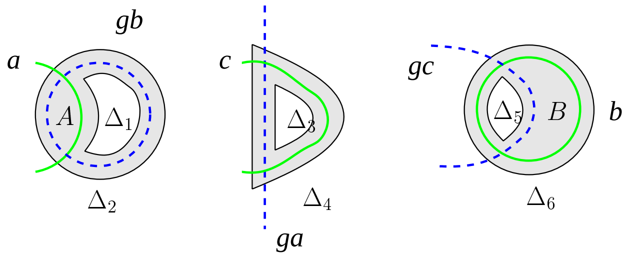


Figure 50: The intersection of N with the red sphere

this the inner boundary component, and call the other one the outer boundary component. Give the component of $S_t - N_t$ which borders the inner boundary component of N^A the label Δ_1. Similarly give the components of $S_t - N_t$ which border the inner components of gN^A and g^2N^A the labels Δ_2 and Δ_3. Now give the component of $S_t - N_t$ which borders the outer component of N^A the label Δ_4. Similarly, give the components of $S_t - N_t$ which border the outer components of gN^A and g^2N^A the labels Δ_5 and Δ_6. Some of the components of $S_t - N_t$ may have more than one label, but at least two of them must be innermost discs which have only one label.

At least one of the innermost discs is *not* labelled Δ_4. We will show how to extend N^A to a neighbourhood which is a 3-ball with saddle reducible boundary by adding an innermost disc labelled Δ_i, where $i \neq 4$. We now deal with these cases one by one. Note that the case that Δ_1 is innermost is the same as the case that Δ_5 is innermost, by swapping colours. Similarly the case for Δ_2 is the same as the case for Δ_6.

We may assume that there are no bigon-orbits disjoint from N_I, as if there were, we could use them to undermine the local maximum. Therefore the red bigons Δ_1, Δ_2 and Δ_3 may not contain components of the graph of double curves with triple points which are disjoint from the orbit of N_I.

Case 2.1 One of Δ_1 or Δ_5 is innermost

We may assume we are in the case Δ_1, as the other case is the same as this one, but with the colours green and blue reversed. We will assume that Δ_1 is vertex-free, otherwise there is a bigon disjoint from N_I which we can use to undermine the local maximum. Furthermore, we may assume that Δ_1 is double curve free, by removing the double curves.

The three bigons A, gB and Δ_1 share a common pair of vertices, so $A \cup gB \cup \Delta_1$ is disjoint from its images under G, by Lemma 6.16. We will use a thin tubular neighbourhood of $A \cup gB \cup \Delta_1$ as our modification neighbourhood, which we may assume contains N_I. This is a 3-ball which is disjoint from its images under G.

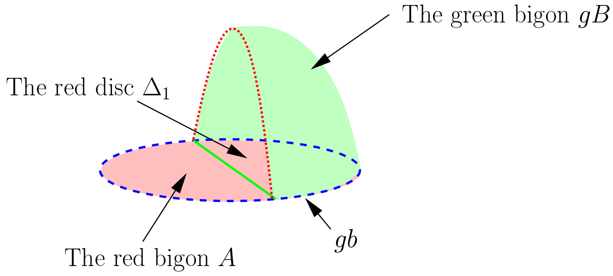


Figure 51: The red bigons A and Δ_1, together with the green bigon gB

The boundary pattern has four double points, and so is saddle reducible by Lemma 6.12.

Case 2.2 Δ_3 is innermost

Again, we may assume that Δ_3 is double curve free. Note that the bigons $gA \cup g^2B \cup \Delta_3$ have the same vertices, so the union of these bigons is disjoint from its images under G. We can choose a tubular neighbourhood of $gA \cup g^2B \cup \Delta_3$, which is disjoint from its images under G, and which contains N_I. This is a 3-ball, which has four double points, as there are two double arcs which intersect the boundary. Therefore it has a saddle reducible boundary pattern, by Lemma 6.12.

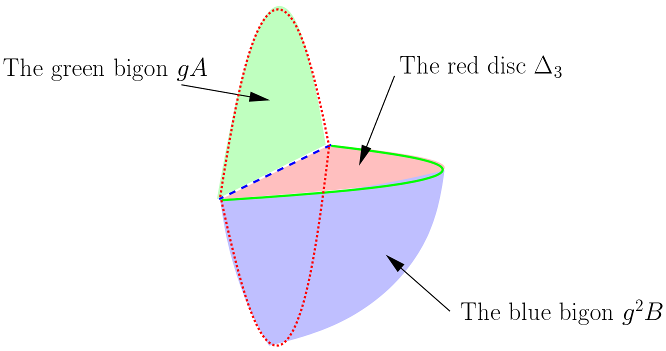


Figure 52: The red bigon Δ_3, together with the green bigon gA and the blue bigon g^2B

Case 2.3 Either Δ_2 or Δ_6 is innermost

We will assume Δ_2 is innermost, as the other case is the same, except with the colours blue and green reversed.

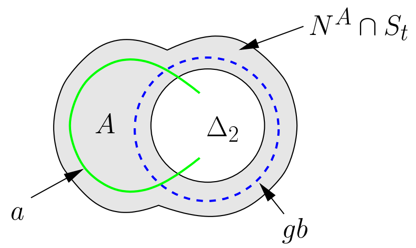


Figure 53: Δ_2 is innermost

Any extra triple points inside the innermost disc create disjoint bigons which we could use to undermine the local maximum, so we may assume that a has exactly two triple points, and divides the disc Δ_2 into two bigons. We can now choose a thin tubular neighbourhood of $A \cup gB \cup \Delta_2$ which contains N^A, and which is disjoint from its images.

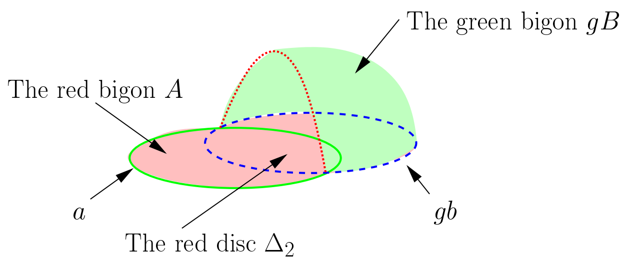


Figure 54: The red bigon A, togther with the red disc Δ_2 and the green bigon gB

A tubular neighbourhood of the bigons shown above is a 3-ball, whose boundary contains precisely two double points, coming from the red double arc. So the boundary pattern is saddle reducible, by Lemma 6.12. $\boxed{\text{Case 2}}$

We have shown that if there are no unknotted invariant circles, then every special case local maximum can be undermined, so there is a sweepout with no triple points, a contradiction.

This completes the proof that any $\mathbb{Z}_3$ action on S^3 is standard.

References

[1] **B Evans**, **J Maxwell**, *Quaternion actions on* S^3, Amer. J. Math. 101 (1979) 1123–1130

[2] **H Hopf**, *Zum Clifford-Kleinschen Raumproblem*, Math. Ann. 95 (1925-26) 313–319

[3] **R Lee**, *Semicharacteristic classes*, Topology 12 (1973) 183–199

[4] **G R Livesay**, *Fixed point free involutions on the* 3*-sphere*, Ann. of Math. 79 (1960) 603–611

[5] **J Milnor**, *Groups which act on* S^3 *without fixed points*, Am. J. Math. 79 (1957) 623–630

[6] **R Myers**, *Free involutions on lens spaces*, Topology 20 (1981) 313–318

[7] **P M Rice**, *Free actions of* $\mathbb{Z}_4$ *on* S^3, Duke Math. J. 36 (1969) 749–751

[8] **G Ritter**, *Free actions of* $\mathbb{Z}_8$ *on* S^3, Trans. Amer. Math. Soc. 181 (1973) 195–212

[9] **J H Rubinstein**, *On 3-manifolds that have finite fundamental group and contain Klein bottles*, Trans. Amer. Math. Soc. 251 (1979) 129–137

[10] **J H Rubinstein**, *Free actions of some finite groups on* S^3, Math. Ann. 240 (1979) 165–175

[11] **P Scott**, *The geometries of* 3*-manifolds*, Bull. London Math. Soc. 15 (1983) 401–487

[12] **H. Seifert**, **W. Threlfall**, *Topologische Untersunchung der Diskontinuitätsbereiche endlicher Bewegungsgruppen des dreidimensionalen sphärischen Raumes*, Math. Ann. 104 (1930-31) 1–70

[13] **W P Thurston**, *Three-dimensional geometry and topology*, Vol 1 (S Levy, editor) Princeton University Press, Princeton, NJ (1997)

[14] **C B Thomas**, *Elliptic Structures on* 3*-Manifolds*, London Mathematical Society Lecture Note Series 104 (1986)

[15] **J A Wolf**, *Spaces of Constant Curvature*, 5th Edition, Publish or Perish, Houston, 1984