PRIME DECOMPOSITIONS AND NORMAL SURFACES

SIDDHARTHA GADGIL

CONTENTS

1. Connected sums and prime decompositions

1.1. **Connected sums.** One can construct a surface of genus 2 from two copies of the torus T^2 by deleting the interior of a disc from each copy and then gluing together the circles that bound these discs. This operation is known as the *connected sum*. Thus, the surface of genus 2 is the connected sum of two tori.

Definition 1.1. Let M_1 and M_2 be two (oriented) n-manifolds. Let $D_i \subset M_i$, $i = 1, 2$ be n-discs embedded in M_i. The connected sum $M = M_1 \# M_2$ of M_1 and M_2 is the manifold $M = (M_1 \setminus int(D_1) \coprod_{\partial D_1 = \partial B_2} (M_2 \setminus int(D_2))$, where the identification is made using the antipodal map.

Remark 1.1. This operation is well-defined because given two embeddings of an n-disc in an n-manifold M, there is a diffeomorphism of M that sends one to the other, by a theorem of Palais and Cerf.

Remark 1.2. Using the results mentioned above, it is easy to see that the operation of taking connected sums is associative, and $M \# S^n = M$

Date: January 10, 2005.

Example 1.1. Let F_g be the surface of genus g. Then $F_g = \#_{i=1}^{g} T^2$ is the g-fold connected sum of tori.

Thus, the 2-torus generates all oriented 2-manifolds as a monoid. We shall see that oriented 3-manifolds, under connected sum, are generated by so-called *prime manifolds*. We are, however, very far from understanding or classifying prime 3-manifolds.

1.2. **Prime decompositions.** Under connected sum, 3-manifolds have much the same structure as the multiplicative semi-group of positive integers. Thus S^3 is the identity, analogous to 1. Further, just as numbers are prime or composite according to whether they have a non-trivial factorisation, we may define prime manifolds.

Definition 1.2. A 3-manifold M is said to be prime if $M = M_1 \# M_2$ implies that at least one of M_1 and M_2 is prime.

The above definition makes sense in any dimension. In the case of 3-manifolds, however, one can go further and *factorise* 3-manifolds *uniquely*.

Given a 3-manifold that is not prime, we can express it non-trivially as a connected sum. If the resulting manifolds are not prime, we can factorise them in turn. A theorem of Knesser asserts that this process stops after finitely many steps.

Theorem 1.3 (Knesser)**.** *Any 3-manifold M can be expressed as a connected sum $M = M_1 \# M_2 \# \dots \# M_n$ of prime manifolds M_i.*

Knesser's proof of this theorem introduced the notion of *normal surfaces* into 3-manifold topology, which have since had a major role in other contexts as well. We prove this theorem after introducing normal surfaces.

We now restrict our attention to orientable 3-manifolds. In this case, a theorem of Milnor asserts that the prime decomposition is unique.

Theorem 1.4. *Suppose $M = M_1 \# M_2 \# \dots \# M_n$ and $M = M_1' \# M_2' \# \dots \# M_m'$, with all the M_i's and M_j''s prime. Then $m = n$ and, after re-ordering $M_i = M - i'$.*

Before proving this theorem, we shall study the prime pieces further.

1.3. **Prime manifolds.** A prime decomposition $M = M_1 \# M_2$ corresponds to a 2-sphere $S^2 \subset M$ that separates M. This is the 2-sphere along which the pieces have been glued. The prime decomposition is trivial iff the 2-sphere bounds a 3-ball.

Example 1.2 (S^3). By the Schoenflies theorem in dimension 3, any 2-sphere in S^3 bounds a ball. Thus, S^3 is prime.

Example 1.3 ($S^2 \times S^1$). Any separating sphere in $S^2 \times S^1$ lifts to the universal cover $S^2 \times \mathbb{R}$, where it is homologically trivial. Thus the sphere bounds a ball in the universal cover, and hence also in $S^2 \times \mathbb{R}$. That this is embedded in $S^2 \times S^1$ follows from the fact that its boundary is embedded.

Observe that $S^2 \times S^1$ contains a 2-sphere that does not bound a 3-ball. We see below that this is the only such prime 3-manifold.

Definition 1.3. A 2-sphere $S^2 \subset M^3$ in a 3-manifold is said to be essential if it does not bound a 3-ball.

Observe that any non-separating 3-sphere is automatically essential. On the other hand, a separating essential 2-sphere corresponds to a non-trivial connected-sum decomposition.

Proposition 1.5. *Suppose M contains a non-separating 2-sphere F. Then $M = M_1 \# S^2 \times S^1$ for some M_1.*

Proof. Suppose F is a non-separating 2-sphere. Consider a neighbourhood $F \times [0,1]$ of F, with boundary components F_0 and F_1. As F is non-separating, there is an arc γ joining F_0 to F_1 that is disjoint from $F \times [0,1]$. The boundary of a tubular neighbourhood $\mathcal{N}(\gamma)$ of this arc is a tube joining F_0 and F_1. The discs $F_i \setminus \mathcal{N}(\gamma)$ together with this tube form a separating 2-sphere. Thus, we get a corresponding connected-sum decomposition consisting of the pieces obtained by cutting along this sphere and gluing in balls.

It is easy to see that the piece containing F is $S^2 \times S^1$. This proves our claim. □

Definition 1.4. A 3-manifold M is said to be irreducible if every 2-sphere in M bounds a 3-ball.

Corollary 1.6. *Any prime manifold M is either $S^2 \times S^1$ or irreducible.*

Proof. As we have seen above, M cannot contain a separating essential sphere. Thus, M is either irreducible or contains a non-separating sphere. In the latter case, the decomposition given by the above proposition must be trivial, hence $M = S^2 \times S^1$. □

1.4. **Uniqueness of Prime decompositions.** We are now in a position to prove uniqueness. Recall that prime manifolds must be either $S^2 \times S^1$ or irreducible. We first show that the complement of the $S^2 \times S^1$ pieces is canonical.

Proposition 1.7. *Suppose $M = M_1 \# S^2 \times S^1 = M_2 \# S^2 \times S^1$, then $M_1 = M_2$.*

Proof. As prime components $S^2 \times S^1$ correspond to non-separating spheres, the result follows from the following lemma.

Lemma 1.8. *Let S_1 and S_2 be two non-separating spheres in M. Then there is a homeomorphism of M that exchanges S_1 and S_2.*

Proof. First consider the case when S and S' are disjoint. Delete the interior of neighbourhoods $S_i \times [0,1]$ from M to get $\hat{M}$, where the neighbourhoods are chosen in such a manner that the $S_i \times \{1\}$ are in the same component, if $\hat{M}$ is not connected. It is a well known result in PL-topology that there is a homeomorphism of $\hat{M}$ exchanging $S_1 \times \{0\}$ and $S_2 \times \{0\}$ and also $S_1 \times \{1\}$ and $S_2 \times \{1\}$. This induces the required homomorphism on M.

In general, we can assume that S_1 and S_2 intersect transversely in a finite number of arcs. The proof proceeds by induction on the number of arcs of intersection. We have proved the claim in the case where this is 0.

If there is at least one component of intersection, then we can find an innermost disc $D \subset S_1$ such that $\partial D \subset S_2$ but $int(D) \cap S_2 = \phi$. Let E and E' be the two discs in S_2 bounded by ∂D. Then, by an elementary homological argument, at least one of the spheres $D \cup E$ and $D \cup E'$, say $D \cup E$ separates. Let $S'' = D \cup E$ pushed off slightly so as to intersect S and S' in fewer components than those of $S \cap S'$. By applying the induction hypothesis, the result follows. □

Observe that if $S_i \subset S^2 \times S^1$ are the non-separating spheres in the above decompositions, then $M - i$ is obtained by splitting along S_i and then capping off the two components by balls. The proposition is now an immediate consequence of the lemma.

□

By a repeated application of this result, we can reduce to the situation where each prime component of M is irreducible. We shall now prove the result in this case.

Proof of the uniqueness theorem. We may assume that $M = M_1 \# M_2 \# \dots \# M_n = M_1' \# M_2' \# \dots \# M_m'$, with each M_i irreducible and not S^3. Let S_i be the spheres along which M has been decomposed to give $M = M_1 \# M_2 \# \dots \# M_n$, and let S be one of the spheres corresponding to the second decomposition.

Suppose first that S is disjoint from each S_i. Then $S \subset M_k$ for some k. As M_k is irreducible, S bounds a ball. This means that the splitting corresponding to S is a splitting of M into two collections of M_i. If this is so for each of the spheres corresponding to the second decomposition, then the two decompositions must agree. We see that, after isotopy, this is indeed so.

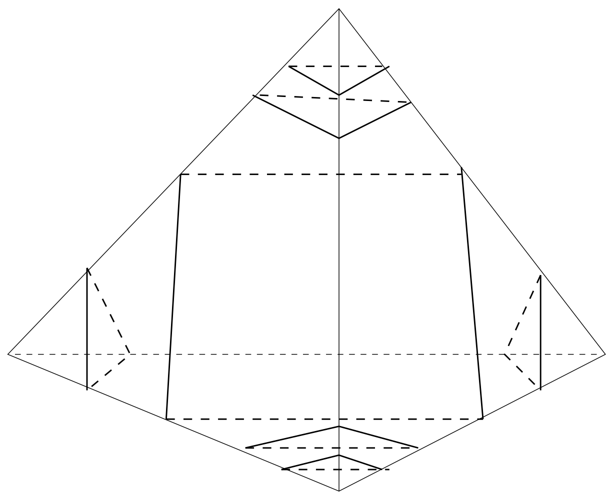

FIGURE 1

Suppose in general that S intersects some S_i's transversely. We can find an innermost disc $D \subset S$ whose interior does not intersect any S_i and whose boundary lies in S_k for some k. Let E and E' be the discs in S_k bounded by ∂D. Then $D \cup E$ and $D \cup E'$ pushed off slightly are spheres in some M_j. Thus, these must bound balls. Using this, it is easy to isotope S to have fewer components of intersection with S_i's. By iterating this, we are in the earlier case. □

2. Normal surfaces

Suppose M is a 3-manifold with a fixed triangulation $\mathcal{K}$. A normal surface $F \subset M$ is a surface in M that intersects each tetrahedron in the triangulation in a collection of *elementary discs*. An elementary disc is either a *triangle*, which is the link of a vertex, or a *quadrilateral* that separates a pair of edges of the tetrahedron from which it is disjoint. There can be at most one type of quadrilateral in each tetrahedron. The various possible elementary discs are illustrated in figure 1

A normal surface is the piecewise-linear analogue of a minimal surface. Given a surface, one can try to isotope it to a minimal surface. In this process, either one obtains a minimal surface, or the surface collapses in some way. In the case of 3-manifolds, we shall see that *essential surfaces* can be isotoped to normal surfaces in a certain sense.

Normal surfaces were introduced by Knesser to prove finiteness of prime decompositions. Subsequently, they have come to play at least two other roles - in decision problems for 3-manifolds and in a manner similar to minimal surfaces.

2.1. **Normal form for spheres.** Recall that a decomposition into connected sums corresponds to a collection of disjoint spheres such that no component of their complement is a punctured ball. We shall call such a collection *essential.* Knesser showed that given such a collection, we can find another such collection with the same number of spheres that is in normal form with respect to the given triangulation.

This may not be the original collection of spheres, but differ from this by the following move. Suppose S is an essential sphere and D is a disc whose boundary lies in S and whose interior is disjoint from all the spheres. Then S can be compressed along D to give two spheres. We can replace the original sphere with at least one of these so that the new collection continues to be essential. The new collection is obtained by making this replacement.

Theorem 2.1 (Knesser). *Given an essential collection of spheres in M, there is an essential collection of normal spheres with the same number of spheres.*

Proof. Assume that the collection of spheres is transversal to the triangulation. We minimize, over all collections of essential spheres with a given cardinality, the pair (α, β) lexicographically ordered, where α is the cardinality of the intersection with the 1-skeleton and β is the sum over all simplices of the number of components of intersection with the spheres. First, we see some simple properties of this collection.

Claim 1. *The intersection of the spheres with each 2-simplex does not contain any closed loop.*

For, one can find an innermost such loop. This bounds a disc disjoint from all the spheres. Compressing along this disc as above gives an essential collection that contradicts minimality.

Claim 2. *The intersection of each spheres with each 3-simplex consists of discs.*

Clearly there cannot be any closed components of intersection for an essential collection. If there is a non-disc component, we can find an innermost one, and hence a compressing disc, as above.

Thus, the intersection of the spheres with each tetrahedron is a collection of discs.

Claim 3. *Each of these discs intersects each edge in at most one point.*

If not, after finding an innermost disc, we can perform an isotopy to reduce the intersection with the 1-skeleton.

It is easy to see now that the components of intersection are either triangles or quadrilaterals of the above form. ∎

2.2. **Normal form for incompressible surfaces.** The above methods readily generalise to collections of incompressible surfaces.

Theorem 2.2 (Haken)**.** *Any collection of disjoint incompressible surfaces in an irreducible manifold M can be isotoped to a collection of normal surfaces.*

Proof. The proof is similar to Knesser's theorem, except that we have to re-examine the case when a surface F bounds a disc D with interior disjoint from it. Since the surface is incompressible, ∂D bounds a disc $E \subset F$. D and E together bound a sphere. By incompressibility, this sphere bounds a ball in M. Thus we can isotope F across this ball so that it is unchanged outside E but E is isotoped onto D.

Using this isotopy in place of the sphere replacement, we can again show that a minimal collection, in the same sense as earlier, is in normal form. □

3. Finiteness theorems

We are now in a position to prove the basic finiteness theorems.

Theorem 3.1 (Knesser's theorem)**.** *Let M be a 3-manifold with a triangulation having t tetrahedra. Then any collection of disjoint non-parallel separating spheres has at most $6t + dim(H_1(M, \mathbb{Z}_2))$ spheres.*

Proof. By what was shown earlier, we may assume that the collection is normal.

Consider each complementary component $W_i \subset M$ of the spheres. Now the intersection of the spheres with each tetrahedron is a collection of triangles and tetrahedra. Of the complementary components $N_i \cap \tau$ in a tetrahedron τ, up to 4 contain vertices and up to two more lie between pieces of different type. If some N_i does not intersect some tetrahedron in one of these pieces, it must be an I-bundle over a surface.

Thus, at most $6t$ components N_i are not I-bundles. Further, each of the I-bundles bounds a collection of sphere, and hence must be either a trivial bundle over a sphere or a twisted bundle over a projective plane. In the former case, we get a contradiction as the two spheres it bounds must be parallel. In the latter case, each such I-bundle contributes to $H_1(M, \mathbb{Z}_2)$. The claim follows. □

Corollary 3.2. *Any 3-manifold has a finite prime decomposition.*

Proof. Decompose M into non-trivial connected sums as long as possible. The above result says that if $M = M_1 \# \dots \# M_{6t+dim(H_1(M,\mathbb{Z}_2))+1}$, then some M_i is trivial. Thus, the process must terminate to give a prime decomposition. □

The proof of Knesser's theorem, together with the theorem of Haken quoted earlier, immediately yields the following.

Theorem 3.3 (Knesser-Haken finiteness). *Suppose M is an irreducible 3-manifold and F_i is a collection of disjoint, non-parallel embedded incompressible surfaces, then the collection has at most $6t + dim(H_1(M, \mathbb{Z}_2)) + b_1(M)$ surfaces.*

Notice that the k disjoint surfaces separate M into at least $k - b_1(M) + 1$ components. The rest of the proof is as before.

References

[1] Hempel, J. P. (1976) 3-*manifolds*, Ann. of Math. Stud. **86**, Princeton University Press.

[2] Jaco, W. H. (1980). *Lectures on Three-manifold Topology*, CBMS Regional Conf. Ser. in Math, American Math. Soc., Volume **43**

Department of mathematics, SUNY at Stony Brook, Stony Brook, NY 11794

E-mail address: gadgil@math.sunysb.edu