ALGORITHMIC RECOGNITION OF 3-MANIFOLDS

ABIGAIL THOMPSON

ABSTRACT. This article discusses recent progress in algorithmically classifying 3-manifolds by homeomorphism type.

INTRODUCTION

Perhaps the most important problem in 3-dimensional topology is the classification problem. In different mathematical settings precisely what is meant by a classification varies. The classification of surfaces is a “model” classification; the closed orientable surfaces are in one-to-one correspondence with the set of natural numbers (via the genus), and the correspondence is easily calculated from a simplicial description of the surface by computing the Euler characteristic. This suggests a definition for at least one type of mathematical classification: a classification of a set $S$ is a one-to-one correspondence between $S$ and a set $T$ which we understand well. In a useful classification of this type the set $T$ will provide illuminating information about the structure of the set $S$. For example, we could “classify” the set $S$ of all integer lattice points in the plane by putting them in one-to-one correspondence with the natural numbers, but this will not tell us much that is useful about $S$. The classification problem for 3-manifolds is to find a one-to-one correspondence between the set of (closed, orientable) 3-manifolds and the natural numbers; optimistically, the correspondence will provide interesting information about the set of 3-manifolds.

In topology there are some less-than-satisfactory classifications around; the classification of knots in the 3-sphere is a good example. Here is how it works: A piecewise linear (PL)-knot in the 3-sphere can be drawn in the plane with a finite number $n$ of distinct crossing points. We can think of drawing such a knot by starting with an $n$-vertex 4-valent planar graph and choosing a sign for each vertex to indicate which strand of the knot goes over and which under. There are only a finite number of imbeddings of $n$-vertex 4-valent planar graphs in the plane, up to

Received by the editors October 1, 1997.
1991 Mathematics Subject Classification. Primary 57M40.
isotopy in the plane. We can construct a list $K_1, K_2, K_3, \ldots$ containing all knots by starting with the unique knot that can be drawn with no crossings, then listing all the knots that can be drawn with one crossing, two crossings, etc. The list will be highly redundant. We need some kind of a sieve to winnow out redundancies. This is provided by an algorithm of Haken’s [H1], with contributions from Hemion, Johannson, Schubert and Waldhausen [He, J, S, W], which decides whether two knots are isotopic. The idea behind this algorithm, in a different setting, is discussed in section 1. (By an algorithm to answer a question $Q$ we mean a program by which a (hypothetical) computer can determine the answer to $Q$ in a finite amount of time. For a more precise definition, see [HU]). Armed with Haken’s algorithm, we return to our list and use it to construct a master list, beginning with $K_1$, and adding the next knot to the master list only after using the algorithm to check that it has not already appeared. So we can construct algorithmically a 1:1 correspondence between isotopy classes of PL-knots and the natural numbers. Unfortunately it is a rather unenlightening listing of knots. No natural properties of a knot (e.g., the crossing number) are known to correspond to the knot’s position in this list. Still, this is the current state of the art for knots, and it is better than the situation for 3-manifolds.

For simplicity we restrict the discussion to closed orientable 3-manifolds. We begin as we did for knots, using the fact that the set of all compact 3-manifolds is countable. One way to show this is to use a theorem due to Moise [Mo] which says that any compact 3-manifold has a PL-structure and can be constructed by gluing together a finite number of tetrahedra. Since there are only a finite number of ways to glue these together, we can construct a list containing all 3-manifolds analogous to our initial listing of knots. This list, too, is highly redundant. Indeed, the 3-sphere appears with depressing regularity. To construct a master list of 3-manifolds, we need a sieve, an analog of Haken’s algorithm. Such a sieve has not yet been found, but there are some classes of 3-manifolds for which a sieve is known. In particular Haken’s algorithm for knots is a variant on an algorithm for a certain class of 3-manifolds, discussed in the next section.

1. NORMAL SURFACES AND HAKEN MANIFOLDS

One effective way to study 3-manifolds is to study “interesting” surfaces inside them. Haken manifolds are irreducible (every 2-sphere bounds a 3-ball) manifolds which contain an essential ($\pi_1$-injective) imbedded surface. They are common, but not universal, among 3-manifolds. Knot complements are Haken manifolds, but the 3-sphere is not. There is an algorithm to decide if a 3-manifold is Haken and to decide if two Haken manifolds are homeomorphic. Both depend on normal surface theory in the manifold, developed primarily by Haken in the 1960s.

Suppose $M$ has a triangulation $T$, with $t$ tetrahedra. A normal surface in $M$ is a closed surface $F$ which intersects every tetrahedron in $T$ in a collection of disks, called elementary disks. These come in two flavors: triangles parallel to the corners and quadrilaterals separating the vertices in pairs; see Figure 1.

There are four types of triangles and three types of quadrilaterals in each tetrahedron. If $F$ is an imbedded surface, then at most one of the quadrilateral types can occur in each tetrahedron.

Let $F$ be a normal surface in $M$, and let $J_1$ and $J_2$ be two tetrahedra in $T$. Suppose $L_1$ is a face of $J_1$ and $L_2$ is a face of $J_2$ and these two faces are identified
in $M$. $F$ intersects $L_1$ and $L_2$ in arcs. Each face contains three types of arcs, one cutting off each vertex of the triangle. The number of each type of arc occurring on $L_1$ and $L_2$ must be the same in order for the pieces of $F$ to match up when $J_1$ and $J_2$ are glued together along their common faces.

Haken realized that this can be expressed algebraically. Assign a variable to every normal disk type in each tetrahedra of $T$, obtaining $7t$ variables ($x_1, x_2, \ldots, x_{7t}$). Call the disk type corresponding to $x_i$ the $i$th disk type. Figure 2 illustrates how one of the matching equations, $x_1 + x_2 = x_3 + x_4$, arises.

There are $2t$ 2-simplices in $T$, so the set of matching equations is a linear system of $6t$ equations (three equations for each 2-simplex in $T$) in $7t$ variables. A normal surface $F$ in $M$ yields a non-negative integral solution $x = (x_1, x_2, \ldots, x_{7t})$ to the system, where the coordinates of $x$ are the number of copies of each disk type contained in $F$. By imposing extra conditions to ensure that all quadrilaterals in a given tetrahedron are of the same type, we obtain a more restricted linear system. Given a non-negative integral solution $x$ to this restricted system, we can piece together a unique imbedded normal surface $F$ by taking $x_i$ copies of the $i$th disk type and gluing them together across the faces. So there is a 1:1 correspondence between imbedded normal surfaces in $M$ and non-negative integral solutions to this linear system.

Non-negative integral solutions to such a system are linear combinations of a finite set of non-negative integral solutions, $x_1, x_2, \ldots, x_n$, called fundamental solutions, which can be found in a finite amount of time. Since each fundamental
solution corresponds to an imbedded normal surface, we obtain a corresponding
finite set of imbedded normal surfaces in $M, F_1, F_2, \ldots, F_n$, called fundamental
surfaces. Any imbedded normal surface in $M$ can be written algebraically as a
non-negative integral linear combination of the $F_i$'s. Wonderfully, this algebraic
fact carries over to the geometry, where it means, roughly, that any normal surface
$F$ can be constructed by taking some finite number of copies of each $F_i$, cutting
these apart along curves where they intersect, and then re-gluing carefully. Call
this operation surface addition.

Suppose now we are given two triangulated 3-manifolds, $M$ and $N$, and we want
to decide if they are homeomorphic. Normal surfaces give a combinatorial way to
analyze the process of decomposing a manifold along essential surfaces. If both $M$
and $N$ are Haken, the work of Haken et al. [H1, He, J, S, W] uses this description to
give an algorithm to decide if $M$ and $N$ are homeomorphic. This algorithm, slightly
modified, also gives the algorithm, mentioned in the introduction, that distinguishes
knots. Later work of Jaco and Oertel [JO] showed that if a triangulated manifold
is Haken, then there is an essential surface among the fundamental surfaces. This
gives an algorithm to tell if a 3-manifold is Haken. So normal surfaces translate a
difficult geometric problem—finding an interesting surface in a 3-manifold—into a
problem in linear algebra—finding a generating set of solutions for a certain type
of linear system. This linear algebra problem can be solved in a finite amount of
time, though not, as yet, in any practical sense (see [HLP] for a discussion of the
complexity of a closely related algorithm). Given our original $M$ and $N$, then, we
can first apply Jaco-Oertel to see if they are both Haken and, if they are, then apply
Haken’s algorithm to see if they are homeomorphic. This yields a classification of
Haken manifolds. The remaining case is when $M$ and $N$ are both non-Haken, and
recognizing the 3-sphere, the simplest non-Haken manifold, is the first step.

2. Recognizing the 3-sphere

In 1992, in a series of lectures at the Technion, Haifa, J. Hyam Rubinstein [R2]
described an algorithm to decide whether or not a triangulated, closed, orientable 3-
manifold is homeomorphic to the 3-sphere. His argument that the algorithm works
used a mixture of normal surface theory and ideas from minimal surface theory. We
will describe the algorithm and give a brief description of the proof from a different
point of view, using thin position for knots.

The starting point for the algorithm to decide if $(M, T)$ is the 3-sphere is this: if
there is a normal 2-sphere in $(M, T)$, we can find it in finite time. This is because the
Euler characteristics of surfaces add under the surface addition operation described
above. So if $A$, $B$ and $C$ are closed, connected, orientable normal surfaces, $A$
is a 2-sphere and $A = B + C$, then exactly one of $B$ or $C$ is a 2-sphere and is
combinatorially “simpler” than $A$. Repeating this process, we eventually arrive at
a normal 2-sphere which cannot be written as a sum. This 2-sphere must be a
fundamental surface, so we can restrict our search for 2-spheres to the fundamental
surfaces, which can be searched in finite time.

A slightly souped-up version of this argument allows us to find a maximal col-
clection $S^*$ of disjoint normal 2-spheres in $(M, T)$ so that any normal 2-sphere in
$M$ which is disjoint from every element of $S^*$ is parallel to one of the elements of
$S^*$ (see [JO, JT]). $S^*$ is not unique.
At this juncture, Rubinstein introduced a new concept. Normal surfaces in $(M, T)$ are the piecewise-linear analog of stable minimal surfaces in $M$, where the metric on $M$ used to measure area is concentrated along the 1-skeleton, so that the area of a surface is just the number of times it intersects the 1-skeleton [JR]. But minimal surfaces also come in an unstable variety. Rubinstein developed the piecewise-linear analog of an unstable minimal surface in $M$, which he called an almost normal surface. Such a surface is normal except in one tetrahedron, where it contains a disk component whose boundary crosses the edges of the 1-skeleton in eight points (an octagonal disk); see Figure 3.

A way to get an intuitive feeling for the difference between normal and almost normal surfaces is to drop down one dimension and to consider simple closed geodesics on a wiggly (and slippery) 2-sphere; see Figure 4.

The curve $w$ around the waist of the 2-sphere is a stable geodesic. If it were a taut rubber band and you pushed it slightly, it would snap back into position. $w$ is analogous to a normal surface. The left-hand equator $e$ is a geodesic, but it is unstable: if you push it slightly, it will collapse down to $w$ or shrink to a point on the left side. The curve $e$ is analogous to an almost normal surface. If the 2-sphere were round, there would be no simple closed stable geodesics, but there would still be unstable ones, the great circles. In fact, at least three simple closed geodesics exist for any smooth metric on $S^2$ [LS, Gr]. Up one dimension, in any smooth metric on the 3-sphere there is an unstable minimal $S^2$ [SS]. The PL analog of this unstable minimal 2-sphere is an almost normal 2-sphere.

Rubinstein realized that algorithms for finding fundamental normal surfaces work equally well for classes of surfaces with a finite number of disk types intersecting the tetrahedra, in particular for the class of almost normal surfaces. The introduction of almost normal surfaces, together with the observation that they can be searched
for in a finite amount of time, provides the key to the 3-sphere recognition problem.
Here is an outline of Rubinstein’s algorithm:

**Given.** A closed, orientable, triangulated 3-manifold $M$, with fixed triangulation $T$.

**Question.** Is $M$ homeomorphic to the 3-sphere?

**Step 0.** Calculate the first homology group of $M$ with $\mathbb{Z}_2$ coefficients and determine whether it is trivial or not. This can be done straightforwardly using the triangulation $T$. If it is non-trivial, $M$ is not the 3-sphere, and we are done. So we will now assume that it is trivial, implying that $M$ contains no non-separating surfaces.

**Step 1.** Find $S^*$, a maximal collection of disjoint normal 2-spheres in $(M, T)$.

**Step 2.** Cut $M$ open along $S^*$. This splits $M$ into three different types of pieces:

- **Type a:** a 3-ball neighborhood of a vertex of $T$. (Every vertex is enclosed in such a piece.)
- **Type b:** a piece with more than one boundary component.
- **Type c:** a piece with exactly one boundary component which is not of type a.

**Step 3.** Search each type c piece for almost normal 2-spheres. If each type c piece has one, then $M$ is the 3-sphere. If some type c piece fails to contain an almost normal 2-sphere, then $M$ is not the 3-sphere.

**Note 1.** The boundary of a (small) neighborhood of each vertex in $T$ is a normal 2-sphere. Since $S^*$ is maximal it must contain all of these vertex 2-spheres. So every vertex is interior to a piece of type a, and there are no vertices interior to pieces of type b and c. This is convenient, as it means that the remnants of the 1-skeleton of $T$ interior to pieces of type b and c are properly imbedded arcs instead of properly imbedded graphs. Properly imbedded arcs in 3-manifolds lend themselves to traditional knot-theoretic approaches.

**Note 2.** I am using “calculate”, “find”, and “search” to mean that these operations can be done by a hypothetical computer (a Turing machine) in a finite amount of time.

Why does it work? The proof relies on two lemmas:

**Lemma 1.** A type b piece is a punctured 3-ball.

**Lemma 2.** A type c piece is a 3-ball if and only if it contains an almost normal 2-sphere.

By Lemma 2, if some type c piece fails to contain an almost normal 2-sphere, then it is not a 3-ball, and $M$ is not the 3-sphere. If every type c piece contains an almost normal 2-sphere, then (by Lemmas 1 and 2) $M$ is a collection of 3-balls and punctured 3-balls glued together. Since by Step 0 all 2-spheres separate, it follows that $M$ is the 3-sphere.

The difficult part of the argument is in the proof of Lemma 2. We discuss it briefly, from the point of view taken in [T1].

**Proof of Lemma 2.** Let $M'$ be a piece of type c, so $M'$ has a single 2-sphere boundary component and contains no vertices of $T$.

Assume $M'$ contains an almost normal 2-sphere $S$. We show that $M'$ is homeomorphic to a 3-ball. Let $\Lambda$ be the remnants of the 1-skeleton in $M'$. An almost
normal 2-sphere has the property that it can be pushed slightly in either direction to decrease its number of intersections with $A$ by two.

One can see this by concentrating on the octagonal disk, whose boundary intersects two of the edges in its tetrahedron twice each. We can push across either of these edges to remove these pairs of intersection points; see Figure 5.

This is similar to reducing the length of the equator $e$ on the wiggly 2-sphere by pushing it slightly either to the right or to the left. There is a straightforward method originally due to Haken (see [He, pp. 47–51] for a description) to continue this pushing operation on $S$, continually decreasing the number of points of intersection with $A$. Roughly speaking, the procedure gets stuck only if $S$ gets pushed to a normal 2-sphere or if it collapses to a point. $S$ splits $M'$ into two pieces, the outside, containing $\partial M'$, and the inside. Since the only normal 2-sphere in $M'$ is the boundary, when the initial push of the octagonal component is toward the outside, $S$ gets isotoped to the boundary of $M'$; when the initial push is toward the inside, $S$ collapses to a point. (During the pushing operation the sphere never recrosses the original almost normal 2-sphere.) This is enough to conclude that $M'$ is a 3-ball.

Now assume $M'$ is a 3-ball. We will find an almost normal 2-sphere. This is where knot theory provides a really useful tool. Since $M'$ is a 3-ball, we can assume it is round and can foliate it with concentric 2-spheres, except for a single point $x$ in the center. Label the leaves of this singular foliation $F_t$, $t$ in $[0, 1]$, with $F_1$ being the boundary of $M'$ and $F_0 = x$. Most spheres in $F$ will be transverse to the 1-skeleton $A$ in $M'$. As the 2-spheres of $F$ move inward from the boundary, they will pass through a finite number of (single) points of tangency with $A$, creating maxima and minima of $A$ with respect to $F$. Call the 2-spheres which are tangent to $A$ critical 2-spheres. A simple example to picture is a knotted trefoil arc inside a 3-ball; see Figure 6.

In Figure 6 the arc has a single maximum and two minima with respect to $F$. Notice that we could isotop $A$ slightly to obtain $A'$, introducing unnecessary critical points into the picture; see Figure 7.

We now use an idea due to D. Gabai [G], namely, thin position for $A$. Given a fixed imbedding of arcs $A$, Gabai assigned a complexity to $A$ with respect to $F$, called the width. To calculate the width of $A$ with respect to $F$, we choose
2-spheres $S_1, S_2, \ldots, S_m$ in $F$, where $S_1$ lies between the boundary 2-sphere and the first critical 2-sphere, and the remaining $S_j$’s are chosen to lie between each adjacent pair of critical 2-spheres. The width of $A$ with respect to $F$ is the number of times $A$ intersects $S_1 \cup S_2 \cup \cdots \cup S_m$. In Figure 6 the width of $A$ is $2 + 4 + 2 = 8$; in Figure 7 the width of $A'$ is $2 + 4 + 2 + 4 + 2 = 14$. By a slight abuse of notation, we say that the width of $A$ is what we get by minimizing width over all isotopic imbeddings. The imbedding which realizes the minimum is called thin position for $A$ with respect to the foliation $F$. Figure 6 is thin position for the knotted trefoil arc. Thin position is different from, but related to, bridge position for a knot; see [T2].

If $A$ is in thin position, some of the 2-spheres in $F$ have very useful properties. These 2-spheres were a key element in Gabai’s solution of the property $R$ conjecture [G] as well as in Gordon-Luecke’s solution to the knot complement problem [GL]. In our case, if $A$ is in thin position with respect to $F$, one of the 2-spheres in $F$ provides the basis for proving that $M'$ contains an almost normal 2-sphere. A 2-sphere $S$
in \( F \), lying between a maximum and a minimum of \( A \), as in Figure 6, has the property that it can be pushed slightly toward the outside or the inside to remove pairs of intersection points with \( A \). One push removes the nearest maximum, one the nearest minimum. This property is shared by almost normal 2-spheres, and with further work, one can construct an almost normal 2-sphere from \( S \). So if \( M' \) is a 3-ball, it contains an almost normal 2-sphere.

What is the relation between the 3-sphere recognition algorithm and the Poincaré conjecture? A closed (compact, with no boundary) \( n \)-manifold is a homotopy \( n \)-sphere if it has the homotopy type of \( S^n \). The generalized Poincaré conjecture is that every homotopy \( n \)-sphere is homeomorphic to the \( n \)-sphere. The generalized Poincaré conjecture is true in dimensions four and above [Sm], [F] and is one of the great outstanding problems in dimension 3. A useful fact in three dimensions is that a closed 3-manifold is a homotopy 3-sphere if and only if it has trivial fundamental group (see [Hp]). It seems possible that a scheme similar to the 3-sphere recognition algorithm could be used at least to “recognize” a homotopy 3-sphere, that is, given a closed, triangulated 3-manifold \( M \), to determine if it has trivial fundamental group. This is closely related to the word problem for 3-manifold groups. A solution would not in itself answer the Poincaré conjecture, but it would be a significant step forward. If such a homotopy 3-sphere recognition algorithm existed, one could start searching for counterexamples to the Poincaré conjecture as follows: take the list containing all 3-spheres, and, using the homotopy 3-sphere recognition algorithm, start checking for homotopy 3-spheres. If a homotopy 3-sphere is found, use the 3-sphere recognition algorithm to see if it is homeomorphic to the 3-sphere. If there is a counterexample to the Poincaré conjecture, this procedure will eventually find it. Of course if there is no counterexample, the procedure will continue forever. Rubinstein has suggested that understanding immersed almost normal 2-spheres, that is, almost normal 2-spheres which are allowed self-intersections, would be a key part of a scheme to modify the 3-sphere recognition algorithm to make it work for homotopy 3-spheres. While there is some recent progress on understanding immersed normal surfaces in 3-manifolds [AMR], [Ra], there is as yet no homotopy 3-sphere recognition algorithm in sight.

References

[AMR] I. R. Aitchison, S. Matsumoto and J. H. Rubinstein, Immersed surfaces in the figure-8 knot complement, preprint.

[CGLS] M. Culler, C. McA. Gordon, J. Luecke, and P. Shalen, Dehn surgery on knots, Ann. of Math. 125 (1987), 237–300, MR 88a:57026

[F] M. Freedman, The topology of four-dimensional manifolds, J. Differential Geom. 17 (1982), 357–453. MR 84b:57006

[G] D. Gabai, Foliations and the topology of 3-manifolds III, J. Differential Geom. 26 (1987), 479–536. MR 89a:57014b

[GL] C. McA. Gordon and J. Luecke, Knots are determined by their complements, J. Amer. Math. Soc. 2 (1989), 371–415. MR 90a:57006a

[Gr] M. Grayson, Shortening embedded curves, Ann. of Math. 129 (1989), 71-111. MR 90a:53050

[H1] W. Haken, Theorie der Normalflächen, Acta Math. 105 (1961), 245–375. MR 25:4519a

[H2] ———, Some results on surfaces in 3-manifolds, Studies in Modern Topology, Math. Assoc. Amer., Prentice-Hall, Englewood Cliffs, NJ, 1968, pp. 39–98. MR 36:7118

[HLP] J. Hass, J. Lagarias, and N. Pippinger, The computational complexity of knot and link algorithms, preprint.
Department of Mathematics, University of California Davis, Davis, California 95616-5224

E-mail address: thompson@math.ucdavis.edu