Minimal and normal surfaces

There is a correspondence between the theory of minimal surfaces in differential geometry and the theory of normal surfaces. We will explore the correspondence and use it to derive:

Problem: 3-SPHERE RECOGNITION **INSTANCE:** A triangulated 3-dimensional manifold *M* **QUESTION:** Is *M* homeomorphic to the 3-sphere?

By Perelman's work, this is equivalent to:

Problem: SIMPLY CONNECTED 3-MANIFOLD **INSTANCE:** A triangulated 3-dimensional manifold *M* **QUESTION:** Is *M* simply connected?

We'll also see how this correspondence leads to new isoperimetric inequalities.

2-Sphere Recognition



Problem: 2-SPHERE RECOGNITION **INSTANCE:** A triangulated 2-dimensional manifold *M* **QUESTION:** Is *M* homeomorphic to the 2-sphere?

There is a simple and quick algorithm:

- 1. Compute the Euler characteristic of M, $\chi(M)$.
- 2. Check that *M* is connected.
- 3. If *M* is connected and $\chi(M) = 2$, output "Yes". Otherwise output "No".

This does not extend to 3-manifolds. All closed 3-manifolds have Euler characteristic zero. There is no known, simple, invariant that characterizes the 3-sphere. We give a different algorithm that does generalize from 2 to 3 dimensions.

2-Sphere Recognition with geodesics

Idea: Look at a maximal family G of disjoint separating geodesics on a surface. This family has certain properties on a 2-sphere that differ from its properties on any other surface.

These properties can be used to characterize, or recognize, the 2-sphere.



Stability of Geodesics

A geodesic on a surface is a curve that is *locally length minimizing*. Short curve segments minimize lengths among all curves connecting their endpoints. But longer segments may not be length minimizing.

A geodesic is *stable* if it cannot be homotoped to decrease its length, so that there is no shorter curve in some neighborhood.

Otherwise it is *unstable*.

(Similar ideas apply to minimal surfaces in a 3-manifold)



Stability of Geodesics

Unstable geodesics can be deformed to reduce length. Stable geodesics are length minimizing among nearby curves.





Lemma Every non-trivial homotopy class of curves contains a stable geodesic. **Proof**. Take the shortest curve in the class. This is embedded in any metric.





Gorodnik

Generic Metrics

We work with generic metrics, where

1. There are no families of parallel geodesics,

2. A geodesic is either stable or it can be pushed off to decrease length to either side.

Any Riemannian metric can be perturbed a little to make it generic (bumpy) [B. White 1991].





Properties of stable and unstable geodesics:

Theorem Suppose *F* is a surface with a generic Riemannian metric and *G* is a maximal family of disjoint separating geodesics.

- 1. If F is a 2-sphere then G contains an unstable geodesic.
- 2. No region of F-G has four or more boundary geodesics.
- 3. A region in F-G whose boundary is a single stable geodesic is a punctured torus.
- 4. A region in F-G whose boundary is a single unstable geodesic is a disk.
- 5. A region in F-G with two boundary geodesics is an annulus whose boundary consists of one stable and one unstable geodesic.
- 6. A region in F-G with three boundary geodesics is a "pair of pants" whose boundary consists of three stable geodesics.

These properties follow from the curve shortening flow (Gage, Hamilton, and Grayson.)

1. If *F* is a 2-sphere then *G* contains an unstable geodesic.

Goes back to an argument of Birkhoff 1917.



If F is a 2-sphere then G contains an unstable geodesic How can we find a geodesic on a 2-sphere?

Geodesics exist on a 2-sphere



Take a family of curves sweeping out the 2-sphere and shorten each curve in the family.

Geodesics exist on a 2-sphere



Apply the curve shortening flow. Some curves shorten in the direction of a_0 and others in the direction of a_1 . Some curve $a_{1/2}$ gets caught in the middle and converges to a geodesic.

Gage-Hamilton (1986), Grayson (1989)

2. No region of F-G has four or more boundary geodesics.

Assumes F is a surface and G is a maximal family of disjoint separating geodesics.





Proof: Join two boundary curves to form a new curve. Shrink the new curve to a geodesic or a point. The new curve is not homotopic to any of the four boundary curves, and not null-homotopic, and thus must flow to a new geodesic. This contradicts maximality of G.



Join two boundary curves to form a new curve. Shrink the new curve to a new geodesic.



Join two boundary curves to form a new curve. Shrink the new curve to a new geodesic.



Join two boundary curves to form a new curve. Shrink the new curve to a new geodesic.

But we assumed the original family was maximal. So this type of region does not occur.



This works even if there is some genus in the region. The original family was not maximal.

3. A region in *F*-*G* whose boundary is a single stable geodesic is a punctured torus.

Proof: It can't be a disk since Birkhoff's argument implies there would be an extra unstable geodesic.

It can't have genus greater than one, since there is a geodesic in every homotopy class, and so there would be a geodesic separating two handles.

It can be, so it must be a torus with one boundary curve.







4. A region in *F*-*G* whose boundary is a single unstable geodesic is a disk

Proof: Push the curve to one side, decreasing its length. Keep pushing until it shrinks to a point or to a geodesic. It must be to a point since the family is maximal, so this region is a disk.



5. Regions with two boundary curves are annuli that have one stable and one unstable boundary geodesics.



Regions with two boundary curves are annuli that have one stable and one unstable boundary geodesics.



Regions with two boundary curves



There is only one possibility for two boundary curves.

6. A region in *F*-*G* with three boundary geodesics is a "pair of pants" whose boundary consists of three stable geodesics.





There must be additional geodesics in this component of *F*-*G* if it has genus ≥ 1 .

Join two boundary curves. Must shrink to the 3rd geodesic.



Properties of stable and unstable geodesics:

Suppose F is a surface with a generic Riemannian metric and G is a maximal family of disjoint separating geodesics.

- 1. If F is a 2-sphere then G contains an unstable geodesic.
- 2. No region of F-G has four or more boundary geodesics.
- 3. A region in F-G whose boundary is a single stable geodesic is a punctured torus.
- 4. A region in F-G whose boundary is a single unstable geodesic is a disk.
- 5. A region in F-G with two boundary geodesics is an annulus whose boundary consists of one stable and one unstable geodesic.
- 6. A region in *F*-*G* with three boundary geodesics is a ``pair of pants" whose boundary consists of three stable geodesics.

Which of these can occur on a 2-sphere?

Geodesics on a 2-Sphere

Properties of stable and unstable geodesics:

Suppose F is a surface with a generic Riemannian metric and G is a maximal family of disjoint separating geodesics.

- 1. If F is a 2-sphere then G contains an unstable geodesic.
- 2. No region of F-G has four or more boundary geodesics.
- 3. A region in F-G whose boundary is a single stable geodesic is a punctured torus.
- 4. A region in F-G whose boundary is a single unstable geodesic is a disk.
- 5. A region in F-G with two boundary geodesics is an annulus whose boundary consists of one stable and one unstable geodesic.
- 6. A region in F-G with three boundary geodesics is a ``pair of pants" whose boundary consists of three stable geodesics.

Note: All regions are disks, annuli, or pairs of pants except Case (3).

So F is gotten by gluing together disks, annuli, and pairs of pants along separating curves unless some region has boundary that is a single stable geodesic.

What surface can be gotten by gluing together disks, annuli, and pairs of pants along separating curves?

Geodesics on a 2-Sphere

What surface can be gotten by gluing together disks, annuli, and pairs of pants along separating curves?



Geodesics on a 2-Sphere

What surface can be gotten by gluing together disks, annuli, and pairs of pants along separating curves?



Geometric 2-Sphere Characterization

Theorem *F* is a 2-sphere if and only if *G* satisfies:

1. There is at least one unstable geodesic in G.

2. No complementary region of F-G has boundary consisting of a single stable geodesic. **Proof**. Push the unstable geodesic to either side, decreasing its length. Either it flows to a stable geodesic and gets stuck, or it flows to a point. In the first case it is a boundary component of an annulus on that side, and in the second case it bounds a disk.



Look at adjacent regions. These glue together to form a tree of regions. The surface F is a 2-sphere if and only if each of these regions is a punctured sphere (disk with holes). This happens exactly when no complementary region has boundary consisting of a single stable geodesic



What about the 3-sphere?

Lemma: Suppose we cut a manifold *M* open along a collection of separating 2-spheres.

Then *M* is homeomorphic to a 3-sphere if and only if every component is homeomorphic to a "punctured" 3-ball (a 3-ball with some 3-balls removed).





Geometric 3-Sphere Characterization

Given a possible 3-sphere M with a generic Riemannian metric: 1. Find a maximal family of disjoint, stable, separating minimal 2spheres.

3. Find a maximal family of disjoint, unstable, separating minimal 2-spheres in the complement of the first family.

Let *G* be the resulting family of minimal 2-spheres. M is a 3-sphere if and only if *G* satisfies the following conditions:

1. There is at least one unstable minimal 2-sphere in G.

2. No complementary region of M - G has boundary consisting of a single stable minimal 2-sphere.

Proof. Similar to 2-sphere case, using results of Pitts, Simon, Smith, Meeks Yau on minimal 2-spheres.



Geometric 3-Sphere Characterization

Rubinstein's idea: Make this an algorithm by replacing minimal surfaces with normal surfaces.



Normal surfaces play the role of minimal surfaces. They are locally minimizes for weight.

What plays the role of an unstable minimal surface in the discrete setting?



Almost Normal Surfaces





An *almost normal surface* intersects one 3-simplex in one octagon. It is normal everywhere else. These surfaces play the role of unstable minimal surfaces in our setting, with weight replacing area.



Another type of *almost normal surface* intersects one 3-simplex in a pair of tubed elementary disks. These aren't needed in the 3-sphere recognition problem.

Almost Normal Surfaces



An *almost normal surface* contains an octagon in one 3-simplex. It is normal everywhere else. These surfaces play the role of unstable minimal surfaces in our setting, with weight replacing area.



There is a way to push this octagonal piece to either of its two sides, in such a way that the weight decreases by two. But can't push in both directions.



When we push an almost surface *S* off to one side or the other, we get a new surface of smaller weight that may not be normal. We can keep pushing, performing "normalization, until eventually we get to a normal surface or push *S* into a single tetrahedron.

During the normalization process, *S* never starts intersecting new normal surfaces. If *S* is initially disjoint from a normal surface N then it stays disjoint from N as it flows to a normal surface or a point. The disjoint normal surface N is a *barrier*.

An Algorithm for recognizing the 3-sphere

(Following Rubinstein and Thompson)

3-Sphere Recognition

Instance: A collection of 3-simplices M with faces paired. **Question**: Is M homeomorphic to the 3-sphere?

Find a maximal collection of non-parallel separating fundamental normal
 spheres S*.

2. Cut M open along S*. This gives three types of pieces.
Type a: A 3-ball neighborhood of a vertex.
(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, not of type a.

3. For each Type c piece , compute all the fundamental almost normal 2spheres inside it. If each type c piece contains a fundamental almost normal 2-sphere, then M is the 3-sphere. If some type c piece fails to contain a fundamental almost normal 2-sphere, then M is not the 3-sphere.

Why does this work?

Algorithm for recognizing the 3-sphere

Cut *M* open along a maximal collection of separating normal 2-spheres.

1. *M* is homeomorphic to a 3-sphere if and only if every component is homeomorphic to a "punctured" 3-ball (a 3-ball with some 3-balls removed).



Algorithm for recognizing the 3-sphere

Cut *M* along a maximal collection of separating 2-spheres.

M is homeomorphic to a 3-sphere if and only if every component is homeomorphic to a punctured 3-ball. Let's look at the three types of regions in M-S*.

Type a: A 3-ball neighborhood of a vertex. Every vertex is enclosed in such a piece and these are 3-balls.




Cut *M* along a maximal collection of separating 2-spheres.

Type b: A piece X with more than one boundary component. These regions also are always 3-balls.



Cut *M* along a maximal collection of separating 2-spheres.

Type b: A piece X with more than one boundary component. These regions also are always 3-balls.



Cut *M* along a maximal collection of separating 2-spheres.

Type b: A piece X with more than one boundary component. These regions also are always 3-balls.



Cut *M* along a maximal collection of separating 2-spheres.

Type b: A piece with more than one boundary component. These regions also are always 3-balls.



Type c: A region with one boundary component, not a vertex linking ball. If the region is a 3-ball, then we can foliate it with 2-spheres shrinking down to a point. Each leaf of this foliation intersects the edges of the triangulation. We move these edges by an isotopy to minimize the maximum weight within the family of 2-sphere leaves. This maximum weight is realized by a 2-sphere *S* that is in *thin position*.



A. Thompson gave an argument that used thin position to prove that such a 2-sphere is normal except in one tetrahedron, which it intersects in an octogonal piece.



Type c: A region with one boundary component, not a vertex linking ball.

If the region contains an almost normal 2-sphere *S* then we can push this 2-sphere to either of its two sides, in such a way that the weight decreases by two.



Keep pushing to decrease the weight. This is the process of "normalization" that we have seen before. The process continues until *S* shrinks to a point or until *S* shrinks to a normal surface.

If *S* shrinks to a point when it is pushed off to one side, then it bounds a ball on that side. If it shrinks to a normal sphere, then it must shrink to the unique normal 2-sphere on the boundary of the Type c region. Thus the almost normal 2-sphere must be boundary parallel on one side and bound a ball on the other side. therefore the Type c region must be a 3-ball.

We have shown that Type a and Type b regions are always 3-balls and that a Type c region is a 3-ball if and only if it contains an almost normal 2-sphere.

We have reduced the question of whether a 3-manifold is homeomorphic to a 3sphere to questions about normal and almost normal 2-spheres in the manifold, which we can answer using Haken's methods.

Minimal Surface - Normal Surface Correspondence

Smooth Riemannian Manifolds	Combinatorial Triangulated Manifolds
Geodesic	Normal curve
Length or Area	Weight
Stable minimal surface	Normal surface
Unstable minimal surface	Almost normal surface
Flow by mean curvature	Normalization
A smooth S ³ contains an unstable minimal S ²	A PL S3 contains an almost normal S2
If ∂X is a stable S ² and int(X) contains an unstable S ² and no stable S ² \Rightarrow X = B ³	∂X a normal S ² and int(X) contains an almost normal S ² and no normal S ² \Rightarrow X = B ³

Normal and minimal surfaces

The correspondence between normal surfaces and minimal surfaces has more applications. It can be used to investigate classical problems in Differential Geometry.

Classical Isoperimetric Inequality:

A curve γ in \mathbb{R}^2 bounds a disk *D* with $4\pi A \leq L^2$.

Equality holds if and only if γ is a circle.



What about if we are given an unknotted curve γ in \mathbb{R}^3 ? Is there a disk with $A \leq f(L)$ for some function f?



Normal and minimal surfaces

The correspondence between normal surfaces and minimal surfaces has many more applications. It can be used to investigate classical problems in Differential Geometry.

Classical Isoperimetric Inequality:

A curve γ in **R**² bounds a disk *D* with $4\pi A \leq L^2$.

Equality holds if and only if γ is a circle.

Suppose we are given an unknotted curve γ in \mathbb{R}^3 ? Is there a disk with $A \le f(L)$ for some function f?

- 1. There is an immersed disk with $4\pi A \le L^2$. (Andre Weil, 1926)
- 2. There is an embedded surface with $4\pi A \le L^2$. (W. Blaschke, 1930)

What bounds can we get for an embedded disk?







Normal and minimal surfaces

What bounds can we get on an embedded disk?



Theorem 1. (H-Lagarias-Thurston, 2004) There is a constant C > 1 and a sequence of unknotted, smooth curves γ_n embedded in R³, each having length L = 1, such that the area of any embedded disk spanning γ_n is greater than *n*.

Theorem 2.

For any embedded closed unknotted smooth curve γ in R³ having length *L* and thickness *r*, there exists a smooth embedded disk of area *A*, having γ as boundary with

$$A \le (C_0)^{(L/r)^2} L^2$$

where $C_0 > 1$ is a constant independent of γ , *L* and *r*.

(The thickness of a curve r is the radius of its tubular neighborhood.)

These results may not seem to connect to algorithms or normal surfaces. But Theorem 2 falls out of Haken's Normal surface theory.



Lower Bounds Is UNKNOTTING really hard?

Is the Haken algorithm an efficient approach to UNKNOTTING?

Are Fundamental Surfaces really complicated?

Is UNKNOTTING really hard?



Is the Haken algorithm an efficient approach to UNKNOTTING?

Are Fundamental Surfaces really complicated?

Yes.

Spanning disks for some unknots cannot be less than exponentially complicated.

There are unknotted polygons in R³ that have *n* edges and that cannot be spanned by disks having fewer than cⁿ triangles.

This is an example of a **Lower Bound** for a computational problem.

Spanning Disks can be Exponentially Complicated

Theorem (H-Snoeyink-Thurston)

There exists a sequence of unknotted polygons K_n with 11n edges such that any disk spanning the unknot K_n contains at least 2^n triangular faces.





Proof:







Spanning Disks can be Exponentially Complicated

Theorem (H-Snoeyink-Thurston)

There exists a sequence of unknotted polygons K_n with 11n edges such that any disk spanning the unknot K_n contains at least 2^n triangular faces.





Three curves in the sequence of unknots K_n



How to construct K_n





To construct K, start with this braid

$$\alpha = \sigma_1 \sigma_2^{-1}$$

Obtained by iterating the braid α n times, followed by n iterations of α^{-1} . $\alpha = \left| \begin{array}{c} & & \\$

How to construct K_n







 $\alpha = \sigma_1 \sigma_2^{-1}$



3 K-1

Obtained by iterating the braid α n times, followed by n iterations of α^{-1} . $\alpha = \alpha^{-1} = \alpha^{-$



Spanning Disks for K_n



standard disks

Each K_n is the boundary of a *standard* embedded disk in \mathbb{R}^3 . We will show that

- 1. This disk cannot be constructed with less than 2ⁿ flat triangles.
- 2. No other disk can do better.

Braids and surface diffeomorphisms



Associated to a braid is a diffeomorphism of a punctured disk.







How can we understand the long term behavior of the sequence

$$\phi, \phi^2, \dots$$





The level o "standard" disk are iterated by P. Y: S2-(4pts) -> S2-(4pts)





















Each iteration of φ more than doubles the number of times that the standard disk spanning the curve intersects B₀.

$$2a + 2b \longrightarrow 6a + 8b > 2(2a + 2b)$$

Each iteration of φ more than doubles the number of times that a disk spanning the curve intersects B₀.

 $2a + 2b \longrightarrow 6a + 8b > 2(2a + 2b)$





The level o "standard" disk are iterated by P. Y: S2-(4pts) -> S2-(4pts)



What if we looked at some other disk spanning K_{n} , rather than the standard disk. Could it intersect B_0 in less points?



Look at the level sets of a Morse function for some disk. Type 1 and 2 critical points don't affect the number of intersections with B_0 . Type 3 do change this number, perhaps drastically. But only one type 3 can occur. So the argument applies below or above this critical point, which suffices for the estimate





From Complexity Theory to Differential Geometry





L =length of KA =area of a disk with boundary K

Theorem. Curves in the plane satisfy the inequality

$$A \le \frac{L^2}{4\pi}$$

Is there a similar inequality for the area of disks spanning unknotted curves in R³?



Isoperimetric inequality



L = length of K A = area of a disk with boundary K

Theorem. Curves in the plane satisfy the inequality

$$A \le \frac{L^2}{4\pi}$$

Is there a similar inequality for the area of disks spanning unknotted curves in R³?



Theorem (H-Lagarias-Thurston) (2005) There is no isoperimetric inequality for disks spanning embedded curves in R³
There is no isoperimetric inequality for disks spanning embedded curves in R³.



There is a sequence K_n of length-one, unknotted curves such that K_n does not bound a disk of area less than n.

What bounds can we get on an embedded disk?



Theorem 1. (H-Lagarias-Thurston, 2004) There is a constant C > 1 and a sequence of unknotted, smooth curves γ_n embedded in R³, each having length L = 1, such that the area of any embedded disk spanning γ_n is greater than *n*.

Theorem 1 holds for the curves below if they are normalized to have length one. Any disk spanning these curves crosses the cylinder below exponentially often.



Question.

Can we control the area of a spanning disk by adding some additional geometric condition?

What bounds can we get on an embedded disk?

Theorem 2.

For any embedded closed unknotted smooth curve γ in R³ having length *L* and thickness *r*, there exists a smooth embedded disk of area *A*, having γ as boundary with

$$A \le (C_0)^{(L/r)^2} L^2$$

where $C_0 > 1$ is a constant independent of γ , *L* and *r*.

For a curve with length one:

$$A \le (C_0)^{(1/r)^2}$$





Theorem 2.

For any embedded closed unknotted smooth curve γ in R³ having length one and thickness *r*, there exists a smooth embedded disk of area *A*, having γ as boundary with

$$A \le (C_0)^{(1/r)^2}$$

Proof. Isotop γ within its (1/r) tubular neighborhood to a polygon K with *n* edges, where

 $n \le 32(1/r)$

Triangulate the complement of K in a ball B of radius 4. B contains less than *t* tetrahedra by an explicit construction, where

 $t = 290n^2 + 290n + 116$

Then construct a spanning disk for γ that is a *fundamental normal disk*. This requires at most C₂ disks, where

$$C_2 = 2^{10^8 t}$$

Each disk is a triangle in a ball of radius 2, and thus has area at most 8. Sum up the areas to get an upper bound.



What bounds can we get on an embedded disk?

Theorem 2.

For any embedded closed unknotted smooth curve γ in R³ having length *L* and thickness *r*, there exists a smooth embedded disk of area *A*, having γ as boundary with

where $C_0 > 1$ is a constant independent of γ , *L* and *r*.

For a curve with length one: $A \leq (C_0)^{(1/r)^2}$

This is a result in classical differential geometry that falls out of Haken's Normal surface theory.

Lesson:

Area or length and curve or surface complexity are closely related.





$$A \le (C_0)^{(L/r)^2} L^2$$