Closed geodesics on incomplete surfaces.

Paul Norbury (`pnorbury@ms.unimelb.edu.au`) and J. Hyam
Rubinstein (`rubin@ms.unimelb.edu.au`) *
Department of Mathematics and Statistics, University of Melbourne, Australia 3010

Abstract. We consider the problem of finding embedded closed geodesics on the
two-sphere with an incomplete metric defined outside a point. Various techniques
including curve shortening methods are used.

1. Introduction.

The existence of closed geodesics on a Riemannian surface often de-
pends only on the topology of the surface. Having proven existence,
such a result usually falls short of giving basic properties like whether
the geodesics are embedded, or where the geodesics are located on the
surface. A simple case of the latter is whether a geodesic contains or
avoids a given point on the surface.

The general problem addressed in this paper is the existence of em-
bedded closed geodesics on incomplete Riemannian surfaces. We require
such a geodesic to *avoid the incomplete points.* Any incomplete point
is required to be contained in neighbourhoods with arbitrarily small
radius and area. We loosely say the metric is defined outside a finite
set of points.

We will consider the situation when a surface (in fact a two-sphere)
contains exactly one incomplete point. Topologically, one can think of
those embedded closed geodesics that miss the incomplete point as the
complement of the space of based loops. That is

$$\{\text{based loops}\} \subset \{\text{all loops}\}$$

and we wish to ask if all of the critical points of the energy functional
are contained inside the smaller space. In finite dimensions one can have
$S^2 \subset S^3$ with all critical points of a Morse function on S^3 contained
inside S^2, whereas for $\mathbb{RP}^2 \subset \mathbb{RP}^3$ any Morse function on $\mathbb{RP}^3$
must have a critical point outside of $\mathbb{RP}^2$ by a parity argument. On
the space of loops, one cannot use a similar argument to show that

---

the set of critical points of the energy functional cannot be contained inside a space of based loops, since there is an example of a two-sphere that does not possess an embedded closed geodesic avoiding a given point. We give this example in Section 6. It consists of a two-sphere with an incomplete metric so it would be interesting to know if there is a complete example.

There are two main approaches to proving the existence of closed geodesics on Riemannian surfaces. The first uses an infinite dimensional version of Morse theory. In finite dimensions, Morse theory enables one to use the topology of a manifold to deduce information about critical points of a smooth function on the manifold. By applying Morse theory to the length or energy functional on the loop space of a Riemannian manifold, one aims to use the topology of the loop space to deduce information about the existence of closed geodesics. The gradient flow, or path of steepest descent, of the length function is not well-defined so one usually looks to other curve-shortening methods to get existence and properties of closed geodesics. This is the main approach we take in this paper.

The study of periodic orbits of a Hamiltonian system on the tangent bundle associated to the geodesic equation is the second main approach for finding closed geodesics. Various tools are used to study Hamiltonian systems, such as methods of ordinary differential equations (ODEs), Poincare sections and Floer homology. Although we will not use Floer homology here, it is worth pointing out that it involves the study of an action functional on the loop space of the tangent bundle using another infinite dimensional version of Morse theory, where one replaces the gradient flow equation with a natural holomorphic-like partial differential equation. In Section 2 we use ODE methods to study the Hamiltonian system.

A model example of an incomplete two-sphere is an ellipsoid with exactly three embedded closed geodesics, each given by the intersection of the ellipsoid with a coordinate plane, with a point of intersection of two of the geodesics removed. Then there is precisely one embedded closed geodesic on the incomplete surface. More generally, the three geodesics theorem gives the existence of three embedded closed geodesics on a smooth Riemannian two-sphere. It uses the fact that the space of unparametrised loops on the two-sphere, modulo identification of all constant loops, is homotopy equivalent to $I\!\!R\!P^3$. Essentially, an embedded closed geodesic is associated to each primitive homology class of $I\!\!R\!P^3$, with the index of the geodesic given by the dimension of the homology class, assuming a bumpy metric. Inside the space of unparametrised loops is the set of loops containing a given point on the two-sphere. This set is homotopy equivalent to $I\!\!R\!P^2$ so one might

expect to get only two embedded closed geodesics containing the point, of index one and two. (The index of a closed geodesic might decrease when we hold fixed a point on the surface.) If one removes a point from a Riemannian two-sphere, leaving an incomplete Riemannian metric on a surface, then one would expect at least one geodesic to survive—avoid the incomplete point. Again, the counterexample of Section 6 means that one needs to be careful with such an argument.

The Gauss-Bonnet theorem—that the integral of the Gaussian curvature over a smooth closed Riemannian surface is 2π times the Euler characteristic—does not apply to incomplete metrics. One can use a local version of the Gauss-Bonnet theorem applied to a neighbourhood of a point, that includes the geodesic curvature of the boundary of the neighbourhood, to measure the incompleteness at a point. See (3) in Section 2.1.

The counterexample of Section 6 shows that it is difficult in general to prove the existence of an embedded closed geodesic on an incomplete surface. However there is a circumstance that arises naturally where existence can be proven. Consider an incomplete metric on a surface minus some points, that satisfies the two extra conditions:

(i) it has positive Gaussian curvature; and

(ii) the Gauss-Bonnet theorem fails in a strong way—the integral of the Gaussian curvature over the surface is infinite.

An example of a metric satisfying such conditions, defined in a neighbourhood of an incomplete point, is

$$ds^2 = dr^2 + rd\theta^2 \tag{1}$$

where r is the distance from the point. The curvature of this metric is $1/4r^2$ which blows up at the incomplete point.

THEOREM 1. *There exists an embedded closed geodesic on any two-sphere minus a point equipped with an incomplete positive curvature metric asymptotic to (1) and satisfying either of:*
   (i) it admits a finite cyclic symmetry of order greater than 2;
   (ii) its shortest geodesic is long.

The technical assumptions (i) or (ii) are necessary since the counterexample in Section 6 consists of an incomplete positive curvature metric asymptotic to (1). Condition (ii) is a condition on the "length" of a homology class of loops on the two-sphere stated precisely in Definition 2. It can be estimated in some cases.

One would expect Theorem 1 to remain true with the explicit asymptotic condition (1) replaced with the weaker, more natural condition previously mentioned, that the Gaussian curvature is positive and its

integral is infinite, and furthermore that this holds only in a neighbourhood of the incomplete point. Nevertheless, as stated, Theorem 1 is quite useful since metrics satisfying (1) arise naturally from higher-dimensional problems.

Two applications of Theorem 1 are given in the next two theorems. In Section 2 we use ODE methods to prove the existence of embedded closed geodesics in very particular circumstances. This allows us to show that condition (ii) of Theorem 1 holds on the incomplete two-spheres corresponding to the round three-sphere and the Fubini-Study metric on $\mathbb{CP}^2$. Thus, a sufficiently close metric still satisfies condition (ii) of Theorem 1 and has positive Gaussian curvature. One could pose the theorems in another way if discrete symmetries are present.

THEOREM 2. *A three-sphere equipped with a circle invariant metric sufficiently close to the round metric possesses an embedded minimal torus invariant under the circle action.*

Theorem 2 should be compared with the result of White [14] who proves the existence of a minimal torus for any metric on the three-sphere sufficiently close to the round metric. Unlike Theorem 2, the result of White does not specify whether or not the torus is invariant under a circle action. Furthermore, we have chosen the round three-sphere rather arbitrarily as an application of Theorem 1. The existence of embedded minimal tori for circle invariant metrics on the three-sphere with Ricci curvature not everywhere positive (so White's result does not apply) can be proven in particular cases by showing that they give rise to incomplete metrics on the disk satisfying the conditions of Theorem 1.

The following theorem can be proven for toric varieties other than $\mathbb{CP}^2$, for example $S^2 \times S^2$ and $\mathbb{CP}^2$ blown up twice. We cannot prove, although we believe, that *any* toric surface contains an embedded minimal 3-torus. (In particular, one expects it to be true for any toric Kahler metric on $\mathbb{CP}^2$.) It seems that the counterexample in Section 6 cannot be adjusted to apply to toric surfaces.

THEOREM 3. *A toric Kahler metric on $\mathbb{CP}^2$ sufficiently close to the Fubini-Study metric possesses an embedded minimal 3-torus invariant under the torus action.*

Previously, closed geodesics on surfaces with incomplete metrics have been studied using methods from ODEs [8, 9]. They arose in the context of minimal surfaces in higher dimensions, as in the applications above. The methods require the surfaces to be highly symmetric. In Section 2 we give such an argument for highly symmetric Riemannian

manifolds as a first step to proving a more general existence result when the metric is slightly perturbed so that it loses it symmetries. Section 3 contains the proof of Theorem 1. The applications are described in detail in the final few sections.

As mentioned previously, we expect the existence of an embedded closed geodesic to be true in greater generality. In Section 4 we describe the minimax technique and the intuition it brings to the problem. A *sweepout* of a manifold is a foliation (with singularities) of the manifold by a path of codimension 1 submanifolds that degenerate to lower dimensional submanifolds at both ends of the path. Well-known examples are sweepouts of a surface by circles, and sweepouts of the three-sphere by tori that degenerate to the two core circles at each end. Another example is a sweepout of $\mathbb{CP}^2$ by T^3s that degenerate to a two-torus at one end and the union of three lines at the other, i.e. take three lines $L_i \subset \mathbb{CP}^2$, $i = 1, 2, 3$ that do not intersect at a common point. The boundary of a disk neighbourhood of the three lines is homeomorphic to T^3, and the complement of the disk neighbourhood is homeomorphic to a neighbourhood of a homologically trivial $T^2 \subset \mathbb{CP}^2$.

A sweepout has a maximal volume leaf. The *minimax* of a family of sweepouts is the infimum of the volumes of the maximal volume leaves. If the family is large enough the minimax is minimal, and very often the minimax (or perhaps a piece of it) is a minimal submanifold.

On a two-sphere minus a point with an incomplete Riemannian metric there is a natural family of sweepouts by loops. The loops degenerate to the incomplete point at one end of the sweepout and to an interior point at the other end. Although the maximum length loop in each sweepout lies inside the two-sphere minus a point, unfortunately the minimax of a family of such sweepouts may not avoid the incomplete point. The maximum length loops may gradually move towards the incomplete point. In Section 4.1 we give two examples of sequences of sweepouts over a two-sphere minus a point with minimax not contained in the two-sphere minus a point. They correspond to a sequence of sweepouts of the round three-sphere by tori and $\mathbb{CP}^2$ by three-tori, with a change of topology in the minimax—rather than obtaining the Clifford torus or the minimal $T^3 \subset \mathbb{CP}^2$ the minimax is a minimal two-sphere, respectively minimal three-sphere.

Although we insist that geodesics on an incomplete surface avoid incomplete points, one can drop this condition since shortest paths still exist on the compact manifold with a singular metric. In [7] it is proven that when the incomplete surface is an orbifold there exists a closed geodesic. In the case of a two-sphere with one orbifold point, the closed geodesic contains the orbifold point.

In the following sections we prove the existence of embedded closed geodesics on the two-sphere minus a point equipped with an incomplete metric in increasingly more general circumstances. We begin with the circle symmetric case where the corresponding Hamiltonian system is integrable and closed geodesics are essentially understood. We move on to a class of metrics with discrete symmetries. Next we add the assumption of positive Gaussian curvature with infinite integral, and consider this case with and without finite symmetries. Finally we study minimax of sweepouts and higher dimensional families of loops.

2. ODE methods

It is often convenient and natural to represent an incomplete point on a surface as a circle boundary along which the metric is degenerate. Polar coordinates (r, θ) give such a representation around an incomplete point as in (1) or around a complete point. We shall take this viewpoint to represent the two-sphere as a disk with incomplete point given by its boundary.

2.1. Circle symmetric metrics

The simplest family of incomplete metrics on the two-sphere is

$$ds^2 = dr^2 + f(r)^2 d\theta^2 \tag{2}$$

where $r \in [0, 1]$ and $\theta \in [0, 2\pi]$, and f vanishes at $r = 0$ and $r = 1$. The case $f(r) = r\sqrt{1 - r^2}$ was studied in [9] (using different coordinates) giving rise to minimal tori in the round three-sphere.

We choose f to behave like r near 0, precisely $f'(0) = 1$, so that the metric is complete there. This condition is a consequence of the local Gauss-Bonnet formula

$$\int_\Omega K da + \int_{\partial \Omega} k ds = 2\pi \tag{3}$$

applied to the disk $r \leq \epsilon$, where $K = -f''/f$ is the Gaussian curvature, $k = f'/f$ is the geodesic curvature of the curve $r \equiv \epsilon$, $da = f dr d\theta$ and $ds = f d\theta$. After integrating out the θ terms (3) becomes

$$-\int_0^\epsilon f'' dr + f'(\epsilon) = 1$$

and the left hand side is $f'(0)$. We allow f to vanish in any way at $r = 1$, although we say the metric is incomplete at the 'point' $r = 1$ whether or not it can be completed. For example if it behaves like $1-r$,

respectively $\sqrt{1-r}$, then in the former case it can be completed, and in the latter it cannot.

Geodesics for (2) parametrised by arc-length satisfy

$$f(r)^2\dot\theta = c, \qquad \dot r^2 + f(r)^2\dot\theta^2 = 1 \qquad (4)$$

for some constant c. It immediately follows that $f(r) \geq c$ so geodesics either meet the incomplete point when $c = 0$, or they are confined to remain a bounded distance from the incomplete point if $c \neq 0$. In the latter case, by the symmetry of the metric each geodesic is periodic in θ, measured between two closest points to $r = 0$ or $r = 1$, and when the period is a rational multiple of 2π, the geodesic is closed.

Each critical point of f in the interior of $(0, 1)$ corresponds to an embedded closed geodesic, so $f_r(r_0) = 0$ implies $r \equiv r_0$ is a geodesic. This occurs at least once, at the maximum of $f(r)$, and precisely once when the Gaussian curvature is positive, since then $f(r)$ is convex and its maximum is the unique stationary point in $(0, 1)$.

In the example $f(r) = r\sqrt{1-r^2}$ where embedded closed geodesics correspond to embedded minimal tori in the three-sphere, the incomplete metric has positive Gaussian curvature with $\int K da = \infty$. The only embedded closed geodesic comes from the maximum of $f(r)$, it has index 3, and it corresponds to the Clifford torus. Nevertheless, in general when a metric can not be completed—we say a metric can be completed if $\int K da = 4\pi$—there may an embedded closed geodesic other than that coming from the maximum of $f(r)$. We state this precisely in the following proposition.

PROPOSITION 2.1. *Let γ be the embedded closed geodesic corresponding to the maximum of $f(r)$. The metric (2) has at least two embedded closed geodesics if*

$$\int K da < 4\pi \quad \text{and} \quad \text{index } \gamma > 2$$

or $\quad \int K da > 4\pi$, $\quad$ index $\gamma < 2$ and γ non $-$ degenerate.

Proof. Specify each geodesic by the minimum value c that f takes on the geodesic. As described above, each geodesic is periodic in θ, and the period Ω_c, defined for $c > 0$, is a rational multiple of 2π for closed geodesics. It is important to note that although the period of γ is 2π it is likely that $\lim_{c\to f_{\max}} \Omega_c \neq 2\pi$. If

$$\lim_{c\to 0} \Omega_c > 2\pi > \lim_{c\to f_{\max}} \Omega_c \qquad (5)$$

then by continuity of the period and the intermediate value theorem there must be a geodesic with r non-constant and period 2π. This gives an embedded closed geodesic other than γ.

For the moment assume the following two limits which may be infinite and are proven below:

$$\lim_{c \to 0} \Omega_c = \pi - \pi/f'(1), \quad \lim_{c \to c_{\mathrm{crit}}} \Omega_c = 2\pi/\sqrt{-f''(r_{\mathrm{crit}})f_{\max}} \quad (6)$$

where $f'(1) \in [-\infty, 0]$ and $f(r_{\mathrm{crit}}) = f_{\max}$ is the maximum of f. The length of γ is $L = 2\pi f_{\max}$, and the Gaussian curvature, which is given by $-f''/f$, is $-f''(r_{\mathrm{crit}})/f_{\max}$ on γ. Thus

$$\int K da = \int_0^{2\pi} \int_0^1 \frac{-f''(r)}{f} f \, dr \, d\theta = -2\pi f'(r)|_0^1 = 2\pi - 2\pi f'(1)$$

so

$$\lim_{c \to 0} \Omega_c \geq 2\pi \ \Leftrightarrow \ \int K da \leq 4\pi$$

and the inequalities can be replaced by strict inequalities.

The Gaussian curvature is constant along γ so the Rauch comparison theorem tells us the exact distance between conjugate points along γ—it is $\pi/\sqrt{K|_\gamma}$. Then

$$L\sqrt{K|_\gamma} = 2\pi f_{\max} \times \sqrt{-f''(r_{\mathrm{crit}})/f_{\max}} = 2\pi\sqrt{-f''(r_{\mathrm{crit}})f_{\max}}.$$

From (6)

$$\lim_{c \to c_{\mathrm{crit}}} \Omega_c \leq 2\pi \ \Leftrightarrow \ L\sqrt{K}| \geq 2\pi$$

and the inequalities can be replaced by strict inequalities. Finally

$$\text{index } \gamma > 2 \ \Rightarrow \ L\sqrt{K}\big|_\gamma > 2\pi$$

$$\text{index } \gamma < 2 \ \Rightarrow \ L\sqrt{K}\big|_\gamma \leq 2\pi$$

since if index $\gamma > 2$ then the index with respect to variations that fix a point on the geodesic, call this index$_\Omega \gamma$, is at least two so γ must contain at least two disjoint arcs of length greater than $\pi/\sqrt{K}$. Similarly, if index $\gamma < 2$ then index$_\Omega \gamma \leq 1$ so the length of γ cannot exceed $2\pi/\sqrt{K}$. Here we have used the fact that on a surface

$$\text{index } \gamma \leq \text{index}_\Omega \ \gamma \leq \text{index } \gamma + 1.$$

In the second inequality, we can guarantee that the length of γ is strictly less than $2\pi/\sqrt{K}$ if γ is non-degenerate (it has no Jacobi fields.)

It remains to prove (6). From (4),

$$\frac{d\theta}{dr} = \frac{\dot{\theta}}{\dot{r}} = \frac{c}{f(r)\sqrt{f(r)^2 - c^2}}$$

so the period of the geodesic is

$$\Omega_c = 2 \int_{r_1}^{r_2} \frac{c}{f(r)\sqrt{f(r)^2 - c^2}} dr$$

where $f(r_1) = c$ and $f(r_2) = c$ is the next time along the geodesic f is c.

The limit $\lim_{c \to 0} \Omega_c$ depends only on the local behaviour of f near 0 and 1 since for any $\epsilon > 0$

$$\lim_{c \to 0} \int_{\epsilon}^{1-\epsilon} \frac{1}{f(r)\sqrt{f(r)^2 - c^2}} dr = \int_{\epsilon}^{1-\epsilon} \frac{dr}{f(r)^2} =< \infty$$

and hence it is annihilated by multiplication by c when $c \to 0$. Thus

$$\lim_{c \to 0} \Omega_c = 2 \lim_{c \to 0} \left(\int_{r_1}^{\epsilon} \frac{c}{f(r)\sqrt{f(r)^2 - c^2}} dr + \int_{1-\epsilon}^{r_2} \frac{c}{f(r)\sqrt{f(r)^2 - c^2}} dr \right).$$

Choose $\alpha > f'(0)$ and ϵ small enough so that for each $r_1 \in (0, \epsilon]$, $f(r)$ lies below the line $y = c + \alpha(r - r_1)$ for $r \in [r_1, \epsilon]$. This gives a lower bound for the limit

$$\begin{aligned}
2 \lim_{c \to 0} \int_{r_1}^{\epsilon} \frac{c}{f(r)\sqrt{f(r)^2 - c^2}} dr &\geq 2 \lim_{c \to 0} \int_{c}^{c_1} \frac{c}{\alpha y \sqrt{y^2 - c^2}} dy \\
&= 2 \lim_{c \to 0} \frac{1}{\alpha} \arctan \sqrt{(c_1/c)^2 - 1} \\
&= \pi / \alpha
\end{aligned}$$

where $c_1 = c + \alpha(\epsilon - r_1)$. Similarly, we can choose $\alpha < f'(0)$ so that $f(r)$ lies above a family of lines with slope α to get an upper bound for the limit. The same argument works near $r = 1$, using $f'(1)$. If $f'(1) = -\infty$ then we simply get an upper bound, with the lower bound of 0 automatic. Thus

$$\lim_{c \to 0} \Omega_c = \pi / f'(0) + \pi / |f'(1)| = \pi - \pi / f'(1).$$

The limit near the critical point, which we may translate to $r = 0$, is a local quantity. It can be simplified as follows.

$$\begin{aligned}
\lim_{c \to c_{\mathrm{crit}}} \Omega_c &= \lim_{c \to c_{\mathrm{crit}}} 2 \int_{r_1}^{r_2} \frac{c}{f(r)\sqrt{f(r)^2 - c^2}} dr \\
&= \lim_{c \to c_{\mathrm{crit}}} \frac{2}{\sqrt{2c_{\mathrm{crit}}}} \int_{r_1}^{r_2} \frac{dr}{\sqrt{f - c}}.
\end{aligned}$$

If we change the parametrisation to $r = r(t)$ then the limit transforms as

$$\lim_{c \to c_{\mathrm{crit}}} \int_{r_1}^{r_2} \frac{dr}{\sqrt{f - c}} = \lim_{c \to c_{\mathrm{crit}}} \int_{t_1}^{t_2} \frac{r'(t) dt}{\sqrt{f - c}} = \lim_{c \to c_{\mathrm{crit}}} \int_{t_1}^{t_2} \frac{r'(0) dt}{\sqrt{f - c}}$$

and

$$\frac{df^2}{dt^2}(0) = \frac{d^2 f}{dr^2}(0)r'(0)^2 + \frac{df}{dr}(0)r''(0) = \frac{df^2}{dr^2}(0)r'(0)^2.$$

Thus if we choose $r'(0) = 1$ then in the limit $f(r)$ is simply replaced by $f(r(t))$ and $f''(0)$ is well-defined. In the generic case, when the maximum is non-degenerate, by the Morse lemma we can choose $r(t)$ so that $f = c_{\mathrm{crit}} - \lambda^2 t^2$ for $\lambda^2 = -f''(0)/2$. The limit becomes

$$\lim_{c \to c_{\mathrm{crit}}} \int_{t_1}^{t_2} \frac{dt}{\sqrt{c_{\mathrm{crit}} - c - \lambda^2 t^2}} = \pi/\lambda.$$

so

$$\lim_{c \to c_{\mathrm{crit}}} \Omega_c = 2\pi / \sqrt{-f''(r_{\mathrm{crit}})c_{\mathrm{crit}}}.$$

When the maximum is a degenerate critical point an approximation argument gives

$$\lim_{c \to c_{\mathrm{crit}}} \Omega_c = +\infty$$

as expected, say when f is constant in a neighbourhood of its maximum value.

2.2. METRICS ON POLYGONS

A more complicated class of incomplete metrics on the two-sphere arise from toric geometry. See Section 5.2. Using ODE methods, we will analyse a specific example which possesses discrete symmetries.

On the square $|x| < 1$, $|y| < 1$ define the metric

$$ds^2 = (1 - y^2)dx^2 + (1 - x^2)dy^2. \tag{7}$$

This is an incomplete metric on the two-sphere, with the boundary the incomplete point.

The equations for the geodesic flow are

$$\ddot{x} = \frac{2y}{1 - y^2}\dot{x}\dot{y} - \frac{x}{1 - y^2}\dot{y}^2$$

$$\ddot{y} = \frac{-y}{1 - x^2}\dot{x}^2 + \frac{2x}{1 - x^2}\dot{x}\dot{y}$$

or implicitly

$$\frac{d^2 y}{dx^2} = \frac{x}{1 - y^2}\left(\frac{dy}{dx}\right)^3 - \frac{2y}{1 - y^2}\left(\frac{dy}{dx}\right)^2 + \frac{2x}{1 - x^2}\frac{dy}{dx} - \frac{y}{1 - x^2}. \tag{8}$$

A compact way to represent the geodesic equations is through the equivalent Hamiltonian system

$$H = \frac{1}{2}\left(\frac{p_1^2}{1-q_2^2} + \frac{p_2^2}{1-q_1^2}\right).\tag{9}$$

We will not use the Hamiltonian formulation, except to point out that it appears to be non-integrable so we expect the geodesic flow to be complicated.

The remainder of this section is devoted to the proof that there is an embedded closed geodesic on the two-sphere with incomplete metric given by (7).

DEFINITION 1. *Define the family of geodesics $\mathcal{H}$ to be those geodesics that meet the y-axis horizontally, or equivalently $\dot{y} = 0$ at $x = 0$.*

The next few lemmas prove that each geodesic in $\mathcal{H}$ meets the line $x = y$ and at least one of these meets $x = y$ orthogonally. By symmetry, this geodesic extends to a geodesic that is an embedded loop. Furthermore, we get estimates on the length and the index of the closed geodesic.

LEMMA 2.2. *A geodesic that meets the y-axis twice while remaining inside the region $|y| \geq |x|$ must be the y-axis.*

Proof. A geodesic that meets the y-axis twice while $|y| \geq |x|$ must be tangent somewhere to a line $y = cx$, and it must lie to the y-axis side of the line (unless the geodesic is the y-axis.) From (8), when $y = cx$ and $dy/dx = c$ we have

$$\begin{aligned}
\frac{d^2y}{dx^2} &= \frac{x}{1-c^2x^2}c^3 - \frac{2cx}{1-c^2x^2}c^2 + \frac{2x}{1-x^2}c - \frac{cx}{1-x^2}\\
&= \frac{cx}{(1-x^2)(1-c^2x^2)}\left(-(1-x^2)c^2 + (1-c^2x^2)\right)\\
&= \frac{y}{(1-x^2)(1-y^2)}(1-c^2)
\end{aligned}$$

Since $|y| > |x|$, $|c| > 1$ so $1 - c^2 < 0$ and d^2y/dx^2 has the opposite sign of y. But this means the geodesic and the y-axis are on opposite sides of the tangent line, which is a contradiction.

LEMMA 2.3. *Each geodesic in $\mathcal{H}$ moves from the y-axis to the lines $x = \pm y$ with y monotone in t.*

Proof. By symmetry, without loss of generality we may assume that $y > x > 0$. A geodesic that satisfies $\dot{y} \leq 0$ and $y > 0$ somewhere, must continue to have $\dot{y} < 0$ until $y < 0$—if $\dot{y}$ gets too close to zero

while $y > 0$ then $\ddot{y} < 0$ so $\dot{y}$ is sent back away from zero. This is an easy consequence of one of the equations for the geodesic, $\ddot{y} = (-y\dot{x}^2 + 2x\dot{x}\dot{y})/(1 - x^2)$. If $y > \epsilon > 0$ and $\dot{y}$ is small enough then $\ddot{y} < 0$ since $\dot{x}$ is bounded.

Thus, a geodesic in $\mathcal{H}$ must travel down and pass through $y = 0$ while $\dot{y} < 0$. From the previous lemma, it cannot pass through the y-axis while $y > 0$ so it must meet the line $y = x$.

LEMMA 2.4. *Any geodesic in $\mathcal{H}$ that meets the line $x = y$ at a slope greater than -1 is convex.*

Proof. Without loss of generality assume $y \geq x \geq 0$. Suppose a geodesic in $\mathcal{H}$ meets the line $x = y$ at a slope greater than -1. In the first quadrant $dy/dx \leq 0$ implies that $d^2y/dx^2 < 0$ since each term of (8) is negative. In particular a geodesic that meets the positive y-axis horizontally has dy/dx negative and decreasing for small positive x. Either this behaviour continues until $y < x$, proving the lemma, or $dy/dx \to -\infty$ and the geodesic becomes vertical. The latter case does not occur, because after it is vertical $\dot{x} < 0$ since $d^2x/dy^2 < 0$ by exchanging x and y in the argument earlier in this paragraph. Furthermore, it cannot become vertical again while $x > 0$ since $d^2x/dy^2 < 0$. In particular, it could not have met the line $x = y$ at a slope greater than -1.

PROPOSITION 2.5. *The square contains an embedded closed geodesic.*

Proof. We will prove that there is a geodesic that meets the lines $x = 0$ and $y = x$ orthogonally and thus by repeated reflections prove the existence of an embedded closed geodesic that meets the lines $x = 0$, $y = 0$ and $y = \pm x$ orthogonally .

Any geodesic in $\mathcal{H}$ close enough to the x-axis is almost horizontal so it meets the line $y = x$ at a slope approximately equal to 0. In the next paragraph we will show that any geodesic in $\mathcal{H}$ that meets $x = 0$ close enough to the boundary meets the line $y = x$ at a steep slope, meaning the slope is less than -1 there, or the geodesic has become vertical at some point and hence meets $y = x$ with a positive slope. By continuity of the angle and the intermediate value theorem, there is a geodesic in between that meets the line $y = x$ at an angle of $\pi/2$. The extension of this geodesic past the line $y = x$ is obtained by reflecting it through $y = x$. Reflect twice more to obtain an embedded closed geodesic.

Geodesics close to the boundary curve sharply away from the boundary. This is because the metric near the boundary (away from the vertices of the square) behaves like the model metric (1) where the geodesics are explicit. Now assume $y > x > 0$. By the proof of the previous lemma, once the slope of the geodesic is steeper than -1, it

must meet the line $y = x$ at a steep slope, i.e. slope less than -1 or positive, since it is either convex all the way to the line $y = x$, and hence steeper than -1 there, or it becomes vertical somewhere and hence must meet $y = x$ with positive slope.
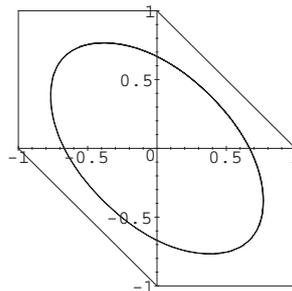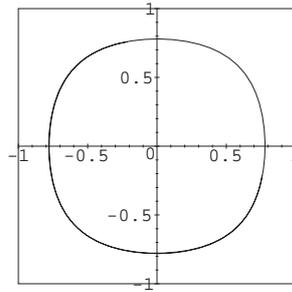

Figure 1. Embedded closed geodesic.

The same argument applies to other regular polygons to give embedded closed geodesics. For example, the triangle and hexagon defined by $abc = 0$ for $(a, b, c) = (1 + x, 1 + y, 1 - x - y)$, respectively $(a, b, c) = (1 - x^2, 1 - y^2, 1 - (x + y)^2$ equipped with the metric $ds^2 = (bcdx^2 + acdy^2 + abd(x + y)^2)/(a + b + c)$ are regular since their affine symmetries, which preserve the metric, make up the whole polygon symmetry group. In Section 5.2 we will see that the square, triangle and hexagon correspond to $S^2 \times S^2$, $C\!I\!P^2$ and $C\!I\!P^2$ blown up at three points, and the embedded closed geodesic in each corresponds to an embedded

minimal T^3. Figure 1 shows the closed geodesics obtained numerically using MAPLE. The geodesics lift to a minimal hypertorus in $S^2 \times S^2$, respectively $\mathbb{CP}^2$ blown up at three points.

3. Geodesic curvature flow.

Gage [5] considered the flow of a loop γ on a surface of curvature $K > 0$

$$\frac{\partial \gamma}{\partial t} = \frac{k}{K}\nu \tag{10}$$

in its normal direction ν, where k is the geodesic curvature of γ. The flow preserves embeddedness and the integral of K over the area enclosed by the loop behaves well under the flow. If the integral is 2π then it is preserved by the flow. Chou and Zhu [3] completed the proof of long time existence to prove that there exists an embedded closed geodesic on a 2-sphere with positive Gaussian curvature.

3.1. Solutions in the neighbourhood of a point

Around a complete point of the metric, any small loop flows to a point in finite time whereas around an incomplete point of the metric, it can take infinite time. This can be seen explicitly for $ds^2 = dr^2 + f(r)^2 d\theta^2$ when f is convex. The curve $r =$ constant has

$$k = -\frac{f'}{f}, \ K = -\frac{f''}{f}, \ \nu = (1,0)$$

so the flow is given by $dr/dt = f'/f''$ with solution

$$f'(r) = f'(r_0)e^t.$$

In a neighbourhood of $r = 0$, $0 < f'(r_0) < f'(r)$ since f is convex so when $f'(0)$ is finite, for example $f'(0) = 1$ in the complete case, the loop converges to a point in finite time.

For the remainder of this section we will assume that $f'(0)$ is infinite. In this case the curve $r =$ constant takes infinite time to converge to a point. For example, when the metric is the model metric (1), so $ds^2 = dr^2 + rd\theta^2$ the flow is given by

$$dr/dt = -2r \ \Rightarrow \ r = r_0 e^{-2t}$$

which shrinks exponentially.

Consider the metric defined on the positive quadrant with singular point given by the union of the axes

$$ds^2 = ydx^2 + xdy^2. \tag{11}$$

For this metric we do not find an exact solution of the flow however the levels sets of xy behave well under the flow. The curve $xy = \epsilon$ has

$$k = \frac{x^2 + y^2}{2xy(x+y)^{3/2}}, \ K = \frac{x+y}{4x^2y^2}, \ \nu = \frac{-1}{\sqrt{x+y}}(1,1)$$

so the infinitesimal change in $\epsilon = xy$ during the flow is

$$\frac{d\epsilon}{dt} = \nabla\epsilon \cdot (\dot{x}, \dot{y}) = -2xy\frac{x^2+y^2}{(x+y)^2}.$$

Denote the geodesic flow of $xy = \epsilon_0$ by $(x(t), y(t))$. We can use the fact that

$$1 \geq (x^2+y^2)/(x+y)^2 \geq 1/2 \tag{12}$$

to prove that

$$\epsilon_0 e^{-2t} \leq x(t)y(t) \leq \epsilon_0 e^{-t}.$$

The flow $\gamma(t) = \{xy = \epsilon_0 e^{-2t}\}$ coincides with the curve $xy = \epsilon_0$ at $t = 0$ and by the first inequality in (12), $\gamma(t)$ initially moves off more rapidly than the geodesic flow $(x(t), y(t))$. If $(x(t), y(t))$ was ever to catch $\gamma(t)$, then it would have to meet it at a tangent such that its geodesic curvature is bounded below by the geodesic curvature of $\gamma(t)$. But then $\gamma(t)$ is moving faster which contradicts the fact that it was ever caught by $(x(t), y(t))$. The same argument shows that the curve $\{xy = \epsilon_0 e^{-t}\}$ never catches $(x(t), y(t))$ when $t > 0$. Thus the curve takes infinite time to reach the boundary and it moves exponentially slowly there.

Any loop in a neighbourhood of the incomplete point (that does or does not surround the incomplete point) flows toward the incomplete point in infinite time. This is because the level sets $r = $ constant (or $xy = $ constant in the second example) form barriers for the loop, forcing it to remain between two level sets that move exponentially slowly toward the incomplete point.

A loop with $\int K da < 2\pi$ does not actually reach the incomplete point, instead it shrinks to a point (away from the incomplete point) in finite time. This is because we must have $\int_\gamma k ds > 0$. The flow satisfies

$$\frac{d \int_\gamma k ds}{dt} = \int_\gamma k ds$$

so $\int_\gamma k ds$ increases to 2π in finite time. At that time the flow stops since $\int K da = 0$ inside the loop. This occurs away from the incomplete point due to an exponentially shrinking barrier so we can apply the results of [3] that if the flow exists only for finite time then the loop flows to a point.

Before we can understand the behaviour of loops with $\int K\,da = 2\pi$ we must first consider scale invariant solutions. Consider solutions to (10) for the metric (1) that are scale invariant under the map

$$(r,\theta) \mapsto (\lambda(t)^2 r, \lambda(t)\theta) \tag{13}$$

where $\lambda(t) = e^{ct}$ determines the speed of a solution. Also rescale arc length $s \mapsto s/\lambda^2$ in order to preserve the third equation below. One consequence of this is that $k\,ds$ is preserved along a curve after rescaling. Lift to the cover with $\theta \in I\!R$. Scale invariant solutions are static solutions of the equation

$$\begin{aligned}
\frac{dr}{dt} &= 4r^2\ddot{r} - cr\theta\dot{r}\dot{\theta} + 2(c-1)r^2\dot{\theta}^2, \\
\frac{d\theta}{dt} &= 4r^2\ddot{\theta} + 2(2-c)r\dot{r}\dot{\theta} + c\theta\dot{r}^2, \\
\dot{r}^2 &+ r\dot{\theta}^2 = 1.
\end{aligned}$$

The variable t is the flow parameter and the dotted derivatives are taken with respect to the arc length s. If we choose $c = 0$ then this produces a geodesic. In fact, we choose $c = 1$, in order to scale at the same rate as the flow of the curves $r =$ constant, or in other words, so that $r =$ constant is a scale invariant solution. The ODE

$$0 = 4r^2\ddot{r} - r\theta\dot{r}\dot{\theta}, \quad 0 = 4r^2\ddot{\theta} + 2r\dot{r}\dot{\theta} + \theta\dot{r}^2, \quad \dot{r}^2 + r\dot{\theta}^2 = 1 \tag{14}$$

has a unique solution for each set of initial data. We call such solutions *scale invariant* solutions (with scaling parameter $c = 1$.)
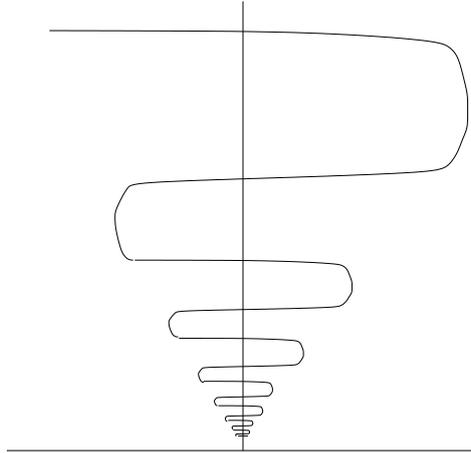

Figure 2. Scale invariant solution $\xi_{a,b}$ acts as a barrier.

Associate to any pair $(a, b) \in \mathbb{R}^+ \times \mathbb{R}$ the unique solution $\xi_{a,b}$ of the ODE (14) with initial conditions at $s = 0$

$$r = a, \ \theta = b, \ \dot{r} = 1, \ \dot{\theta} = 0. \tag{15}$$

Figure 2 shows the solution around the vertical axis $\theta = 0$. The position $s = 0$ occurs at, say the uppermost vertical in the picture to the right of the axis. The solution curve and the line $\theta = 0$ bound closed regions to either side of $\theta = 0$ as shown in the following proposition.

PROPOSITION 3.1. *The scale invariant solution $\xi_{a,b}$ reaches $\theta = 0$ for $s > 0$ and $s < 0$ at interior angles α and β, and over the interior of the region bounded by the solution curve and the line $\theta = 0$, $\int K da = \alpha + \beta$.*

Proof. The Gauss-Bonnet formula over R, the interior of the region bounded by the solution curve and the line $\theta = 0$, is

$$\int_R K da + \int_{\partial R} k ds + \pi - \alpha + \pi - \beta = 2\pi$$

so the second claim $\int_R K da = \alpha + \beta$ is equivalent to the statement that $\int_{\partial R} k ds = 0$. We begin with a lemma stronger than the second claim.

LEMMA 3.2. *Let R be a region bounded by arcs consisting of solutions of (14) together with an arc along $\theta(0) = 0$. Then $k ds$ is exact along ∂R so*

$$\int_{\partial R} k ds = 0.$$

Proof. The unit normal to $\gamma(s) = (r(s), \theta(s))$ is $\nu = r^{-1/2}(r\dot{\theta}, -\dot{r})$ so for solutions of (14)

$$
\begin{aligned}
k ds &= \nabla_{d/ds}(\dot{r}, \dot{\theta}) \cdot \nu ds \\
&= r^{-1/2}(r\ddot{r}\dot{\theta} - r\dot{\theta}^3/2 - r\ddot{\theta}\dot{r} - \dot{r}^2\dot{\theta})ds \\
&\quad \text{use (14) to substitute for } \ddot{r} \text{ and } \ddot{\theta} \\
&= r^{-1/2}(\theta\dot{r}\dot{\theta}^2/4 - r\dot{\theta}^3/2 - \dot{r}^2\dot{\theta}/2 + \theta\dot{r}^3/(4r))ds \\
&= r^{-1/2}(\dot{r}\theta/(4r) - \dot{\theta}/2)(\dot{r}^2 + r\dot{\theta}^2)ds \\
&= r^{-1/2}(\dot{r}\theta/(4r) - \dot{\theta}/2)ds \quad \text{again by (14)} \\
&= -\frac{1}{2}d(r^{-1/2}\theta)
\end{aligned}
$$

The same conclusion holds along the line $\theta = 0$, since $k = 0 = -\frac{1}{2}d(r^{-1/2}\theta)$ when $\theta \equiv 0$, and the lemma is proven.

The proof of the proposition breaks into the two cases $s < 0$ and $s > 0$ for the solution $\xi_{a,b}$ satisfying the initial conditions (15).

LEMMA 3.3. *The solution $\xi_{a,b}$ meets $\theta = 0$ for $s < 0$.*

Proof. First note that for $s < 0$, $\dot{r}(s) > 0$ and while $\theta(s) > 0$, $\dot{\theta}(s) > 0$, since the lines $r = $ constant are solutions of (14), respectively the lines $\theta = $ constant > 0 give $\ddot{\theta} < 0$. Thus, while $\theta > 0$, $\ddot{\theta} < 0$ since $\ddot{\theta} = -\dot{r}\dot{\theta}/(2r) - \dot{\theta}\dot{r}^2/(4r^2)$. It follows that for $s < 0$, $\dot{\theta}(s)$ is decreasing so the solution meets $\theta = 0$ (possibly at $r = 0$) if we can show it stays away from $r = 0$.

The solution $\xi_{a,b}$ does not meet $r = 0$ while $\theta > 0$ since suppose the solution $\xi_{a,b}$ meets the line $r = \epsilon$ at $\theta_1 > 0$. Then the region R bounded by the arcs $\theta = 0$, $r = r(0)$, $r = \epsilon$ and the solution curve satisfies the conditions of Lemma 3.2 (since $r = $ constant is a solution of (14).) Thus if α is the interior angle between the solution curve and the line $r = \epsilon$ then $\int_R K da = \alpha - \pi/2 < \pi/2$. But since $K da = dr d\theta/4r^{3/2}$, $\int_R K da$ is arbitrarily large for small enough ϵ if θ_1 stays away from 0. This gives the desired contradiction.

For the case $s > 0$ we we turn the problem around and prove that any scale invariant solution with $\theta(0) = 0$, $\dot{\theta}(0) > 0$ and $\dot{r}(0) < 0$ eventually becomes vertical ($\dot{\theta} = 0$), and that a vertical at any point (a, b) is reached above by a solution beginning at $\theta = 0$.

LEMMA 3.4. *A solution of (14) with $\theta(0) = 0$, $\dot{\theta}(0) > 0$ and $\dot{r}(0) < 0$ eventually becomes vertical, i.e. $\dot{\theta} = 0$.*

Proof. As in the proof of the previous lemma note that for $s > 0$, $\dot{r}(s) < 0$ and if the solution does not become vertical then $\dot{\theta}(s) > 0$.

If the solution does not become vertical then it must either meet $r = 0$ or become asymptotic to $r = r_1$. In both cases the solution meets the line $\theta = 0$ at an angle α, say, and a line $r = \epsilon$ at an angle β where ϵ is arbitrarily small in the first case and $\beta < \pi/2$, and in the second case ϵ is arbitrarily close to r_1 and β is arbitrarily close to 0. The region R bounded by the arcs $\theta = 0$, $r = \epsilon$ and the solution curve satisfies the conditions of Lemma 3.2 so

$$\int_R K da = \alpha + \beta - \pi/2.$$

In the first case, as in the previous lemma, $\int_R K da$ is arbitrarily large for small enough ϵ which contradicts $\int_R K da = \alpha + \beta - \pi/2 < \alpha$. In the second case, since $\alpha < \pi/2$, when β is small enough $\alpha + \beta - \pi/2 < 0$ contradicting $\int_R K da > 0$.

LEMMA 3.5. *Lemma 3.4 gives a well-defined map*

$$\mathcal{F} : I\!R^+ \times (0, \pi/2) \to \{(a, b) | a > 0, b > 0\}$$

where the domain gives the intial conditions and the range the point at which the solution becomes vertical. The image of $\mathcal{F}$ is both open and closed and hence $\mathcal{F}$ is surjective.

Proof. By continuity of initial conditons of ODEs, in a small enough neighbourhood of a point (a, b) reached vertically above by a solution beginning at $\theta = 0$, any vertical solution remains close to the solution through (a, b) and in particular meets $\theta = 0$ transversally and at an angle not equal to $\pi/2$. Thus the image of $\mathcal{F}$ is open.

The map $\mathcal{F}$ is continuous by continuity of initial conditons of ODEs. Thus the image of any compact set in $\mathbb{R}^+ \times (0, \pi/2)$ is compact and hence closed. In order to prove that the image of $\mathcal{F}$ is closed we must understand the behaviour of $\mathcal{F}$ as the initial angle tends to 0 or to $\pi/2$. As the initial angle tends to 0, the solution tends to the solution $\theta = 0$ and the vertical tends to the line $\theta = 0$ which lies outside the set $\{(a, b) | a > 0, b > 0\}$. As the inital angle, α say, tends to $\pi/2$, take the region R bounded by the arcs $\theta = 0$, the solution curve, and $r = a$, the value of r when the solution becomes vertical. By Lemma 3.2, $\int_R K da = \alpha \approx \pi/2$. But the solution tends towards a horizontal line $r = $ constant, and in order for the integral of K to remain large ($\approx \pi/2$) we must have $a \to \infty$, which lies outside the set $\{(a, b) | a > 0, b > 0\}$ so we have proven that the image of $\mathcal{F}$ is closed in $\{(a, b) | a > 0, b > 0\}$.

Proposition 3.1 follows from combining the previous four lemmas.

The shape of a solution is given in Figure 2. If $b >> a$, then the solution becomes almost horizontal as θ approaches 0, i.e. the two angles α and β of the solution curve with $\theta = 0$ are approximately $\pi/2$. This is because $\dot{r}$ must remain positive, α and β are both less than $\pi/2$ but then $\int K da < \pi$ over the region bounded by the solution curve and $\theta = 0$. In order to maintain such a small integral when $b >> a$, the solution curve may only span a small range along the r direction. Note also that for such a solution $\int K da \approx \pi$.

PROPOSITION 3.6. *(i) A loop with $\int K da = 2\pi$ reaches the incomplete point from a well-defined direction.*

(ii) A loop with $\int K da > 2\pi$ may smear out at the incomplete point meeting it at an interval of angles.

Proof. Any loop γ_t with γ_0 bounding a disk over which $\int K da \geq 2\pi$ reaches the incomplete point (in infinite time) since it maintains $\int K da \geq 2\pi$ throughout the flow so cannot disappear. It is easy to see that (i) implies (ii) if we take a loop with $\int K da > 2\pi$ symmetric around a vertical axis $\theta = \theta_0$. Inside such a loop is a loop with $\int K da = 2\pi$ symmetric around $\theta = \theta_1$ for an interval of θ_1 values, and each such

loop forms a barrier for the outer loop, forcing it to meet the incomplete point at $\theta = \theta_1$.

To prove (i), suppose the converse. If $\int K da = 2\pi$ and γ_t smears out at the incomplete point, take the maximum angle $\theta = \theta_{\max}$ in the limit of γ_t. (Recall that all of this takes place in the universal cover, so $\theta \in \mathbb{R}$.) This exists, since during the flow, γ_t cannot cross a line $\theta \equiv \theta_1$ so take the minimum of the (closed) set of lines θ_1 that lie to the right of γ_t at some finite time T. Similarly, take the minimum angle $\theta_{\min}$.

Since $\int K da = 2\pi$ as $t \to \infty$, γ_t must become long and thin in order to reach $\theta_{\min}$ and $\theta_{\max}$ and to maintain its small integral. Another way to state this is to rescale the flow by (13) to get a solution $\tilde{\gamma}_t$ of (14). The static solutions $r =$ constant serve as barriers above and below the rescaled solution. The lines $\theta = \theta_{\min}$ and $\theta = \theta_{\max}$ now move outwards exponentially. The curve $\tilde{\gamma}_t$ becomes very flat along some $r =$ constant in order to maintain $\int K da = 2\pi$. It moves outwards exponentially fast. Now, place a static solution as in Figure 2 with one of its verticals at (a, b) just to the right of $\tilde{\gamma}_t$. The height of $\tilde{\gamma}_t$ is much less than that of a static solution since by translating the θ coordinate we can make the integral $\int K da$ of $\tilde{\gamma}_t$ to the right of the line $\theta = 0$ arbitrarily small. The static solution always maintains $\int K da \approx \pi$ to the right of the line $\theta = 0$ so it is much thicker (spans a greater range along the r direction) than $\tilde{\gamma}_t$. Since it is static it will serve as a barrier to $\tilde{\gamma}_t$, preventing it from moving outwards exponentially as assumed.

3.2. Global flow

For the remainder of this section we will consider the two-sphere minus a point equipped with an incomplete metric with the properties:

- it has positive Gaussian curvature
- it is asymptotic to (1) or (11) near the incomplete point.

$$(16)$$

Over a smooth Riemannian 2-sphere, an embedded closed geodesic separates it into two pieces with integral of K over each equal to 2π. In the incomplete case, we begin with an embedded loop with integral of K over its interior 2π and infinite on its exterior and expect it to flow to an embedded closed geodesic.

PROPOSITION 3.7. *On $S^2 - \{\text{point}\}$ equipped with a metric satisfying (16), a smooth loop bounding a region with $\int K da = 2\pi$ converges under the flow (10) to one of the following:*
  (i) an embedded closed geodesic;
  (ii) the incomplete point, approaching from a well-defined direction;

(iii) the double of a geodesic arc beginning and ending at the incomplete point.

Proof. If the curve remains a bounded distance from the incomplete point then one can apply the results of [3], that the geodesic flow produces an embedded closed geodesic, since the proof uses only upper and lower bounds for K. Hence (i) results.

Suppose a loop does not converge to an embedded closed geodesic. Then at least part of the loop must flow arbitrarily close to the incomplete point. If there are disconnected intervals of directions at the incomplete point such that the loop moves arbitrarily close to the incomplete point from these directions, but misses an interval of directions in between, then the number of such intervals of directions must be exactly two, the intervals must be points, and the loop must be converging to (the double of) a geodesic beginning and ending at the incomplete point along the two directions. This is because outside a neighbourhood of the incomplete point, arcs of the loop must converge to geodesics. If the limit direction set consists of more than two points, then it must contain intervals in order to accommodate different geodesics leaving the set. In that case, the integral $\int K da$ over the interior of the loop would be infinite. Alternatively, studying scale invariant solutions locally, one can show that the integral of the Gaussian curvature on the interior of a solution to the flow concentrates near the incomplete point along a well-defined direction with $\int K da \approx \pi$. Thus, a loop with $\int K da = 2\pi$ can afford only two such directions. Hence (iii) results.

If neither (i) nor (iii) results then the loop must be eventually contained entirely in an arbitarily small disk neighbourhood of the incomplete point. The neighbourhood of the incomplete point is asymptotically given by (1), (or (11) near vertices of polygons arising from toric surfaces.) Since the loop moves inside arbitrarily small neighbourhoods of the incomplete point, the model metric approximates the given metric arbitrarily closely. By continuity of solutions of the flow, we can conclude from the local study that the loop takes infinite time to reach the incomplete point, that it remains for all time inside any small neighbourhood of the incomplete point, and that it approaches it from a well-defined direction.

To complete the proof of Theorem 1 we use the topology of the disk and its boundary. First we need the following result about small disks.

LEMMA 3.8. *Over $S^2 - \{$point$\}$ equipped with a metric satisfying (16) one can continuously assign a family of expanding disks around any given point $x \in S^2 - \{$point$\}$ so that the integral of K over the disk is arbitrarily large. The assignment is continuous in x and $\int K da$.*

Proof. Any well-defined family of disks is suitable for our purposes. We can arbitrarily choose a background metric and take fixed radius disks. In terms of the local picture $ds^2 = dr^2 + rd\theta^2$ we choose a background metric $ds^2 = dr^2 + d\theta^2$. We may assume the point $x = (c, 0)$ by translating θ. Consider the disk $(r - c)^2 + \theta^2 = c^2$. Then

$$\int_D K da = \int_0^{2c} \int_{-\sqrt{r(2c-r)}}^{\sqrt{r(2c-r)}} \frac{r^{1/2} d\theta dr}{4r^2} = \int_0^{2c} \frac{\sqrt{2c-r}}{2r} dr = \infty.$$

When the metric is only asymptotically like (1), then almost any continuous assignment will do, as long as it locally looks like the construction above in a neighbourhood of the incomplete point. In the case of polygons arising from toric geometry, it is convenient to choose the disks to be (smoothed) polygons centred at each interior point similar to the given polygon.

COROLLARY 3.9. *One can continuously assign to each point on a two-sphere minus a point equipped with a metric satisfying (16) a disk over which $\int K da = 2\pi$.*

Proof. This is immediate from Lemma 3.8 since the integral of K over a very small disk will be small and a large enough disk will give large integral of K.

Before stating the next result, we explain what is meant when "the shortest geodesic is long" which appears in the statement of Theorem 1.

The *length* of a homology class of loops in S^2 is defined as follows. Let η be a k-dimensional family of loops in S^2 representing a homology class $[\eta] \in H_k(\text{loops} \subset S^2)$ and take $l(\eta) =$ length of the longest loop in η. Define $l([\eta])$ to be the infimum of $l(\eta)$ over all representatives η. If there is an index two (possibly index three when the condition $\int K da = 2\pi$ is dropped) embedded closed geodesic on the incomplete two-sphere, then its length is expected to be the length of the second homology class of loops. Note that the assignment of a family of disks with $\int K da = 2\pi$ given by Corollary 3.9 gives an upper bound for the length of the second homology class of unparametrised embedded loops in S^2. Simply take the maximum length in the family. This can be calculated in many cases, in particular for the metrics arising from the round three-sphere, $\mathbb{CP}^2$ and $S^2 \times S^2$. In these cases, one uses a family of Euclidean circles, respectively (smoothed) triangles and squares. The maximum length loop in the family comes from simple calculus.

The length of the shortest geodesic arc running from the incomplete point back to itself can also often be calculated, in particular for the same three examples as in the previous paragraph. In these three cases, the geodesic arc can be explicitly calculated.

When twice the length of the shortest geodesic arc running from the incomplete point back to itself is greater than the length of the second homology class of embedded loops in the two-sphere we expect the geodesic curvature flow to produce an embedded closed geodesic. In fact, we need something a little stronger to hold:

DEFINITION 2. *The shortest geodesic arc running from the incomplete point back to itself is* **long** *if twice its length is greater than the maximum length loop in the family constructed in Corollary 3.9.*

THEOREM 4. *Suppose a metric on a two-sphere minus a point satisfies (16) and the shortest geodesic arc running from the incomplete point back to itself is long. Then there exists a loop that converges to an embedded closed geodesic under the geodesic curvature flow.*

Proof. For each point on the two sphere minus the incomplete point, continuously choose the disk centred at the point over which $\int K da = 2\pi$ given by Corollary 3.9. The boundary of the disk is the initial loop in the geodesic flow.

We will argue by contradiction that one of the loops must flow to an embedded closed geodesic. First, replace the incomplete point with its circle of directions, so the two-sphere with an incomplete point becomes a disk with circle boundary. Each loop must converge to (i), (ii) or (iii) of Proposition 3.7. No loop can converge to (iii), double a geodesic arc, since the length of each initial loop is less than twice the length of the shortest geodesic arc, and the flow is length decreasing. Now suppose that no loop flows to (i), an embedded closed geodesic. Thus each loop flows to (ii), the incomplete point along a well-defined direction, or equivalently to a boundary point on the circle. This gives a map from the disk to its boundary, defined by sending the centre point of an initial disk to the destination of its loop on the boundary.

The map is continuous by continuity of the partial differential equation governing the flow. Of course, continuity only applies to finite time intervals, but by the proof of Proposition 3.6, a loop evolves quite predictably from a very large time until infinity, confined to a cone containing the well-defined limit direction. The continuous map can be extended to the boundary of the disk by setting it to be the identity there. This is because an initial disk centred close to a boundary point evolves close to the boundary point by the proof of Proposition 3.6.

A continuous map from the closed disk to its boundary that is the identity on the boundary cannot exist by algebraic topology considerations ($S^1 \hookrightarrow D^2 \to S^1$ induces the identity on the fundamental group that factors through the zero map) thus giving the desired contradiction.

The flow simplifies considerably if there is a cyclic symmetry of order at least three. In that case, if one begins with a loop that possesses the symmetry neither (ii) nor (iii) of Proposition 3.7 can occur since if a point of a loop approaches the incomplete point from some direction then it must approach from at least three directions, contradicting (ii) and (iii). Thus, (i) remains, so the loop must converge to an embedded closed geodesic.

4. Minimax and sweepouts.

A sweepout of a manifold is a foliation of the manifold by a path of codimension 1 submanifolds that degenerate to lower dimensional submanifolds at each end. An example of a sweepout is the constant radius loops on the the two-sphere with a circle invariant metric described in Section 2.1. The loops degenerate to a point at each end of the sweepout. We showed that the maximum length loop is a geodesic. For more general sweepouts of a two-sphere by loops degenerating to points at each end, it is no longer true that the maximal length loop is a geodesic, however the *minimax* in the family of sweepouts is minimal. The minimax is obtained from a family of sweepouts by taking a sequence of sweepouts in the family with maximal length loop decreasing and converging to the infimum of all maximal lengths in the family. A priori the minimax may not be a smooth loop, and hence a geodesic, however in [12] it is shown that over a complete Riemannian surface the minimax is a geodesic.

4.1. Change of topology in a minimax sequence.

Here we give two explicit minimax sequences to show how one can fail to detect an embedded closed geodesic on an incomplete surface using sweepouts. This corresponds in higher dimensions to a change of topology of the limit from the topology of the leaves of each sweepout. The two examples are based on the same idea.

Consider the two-sphere with incomplete metric $ds^2 = dr^2 + r^2(1 - r^2)d\theta^2$. Any point (r, θ) with $0 < \theta < \pi$ determines a loop in the disk given by joining (r, θ) to $(r, -\theta)$ by a vertical line and by a constant radius circle to the right of the line as in Figure 3. The length of the loop is

$$L(r, \theta) = 2r \sin\theta \sqrt{1 - r^2 \cos^2\theta} + 2r\theta \sqrt{1 - r^2}.$$

(This is clear for $L(r, \pi) = 2\pi r \sqrt{1 - r^2}$, $L(r, \pi/2) = 2r + \pi r \sqrt{1 - r^2}$, and when $\theta \approx 0$, the two summands of L are approximately equal.) A sweepout by loops that degenerate to the boundary at one end and to
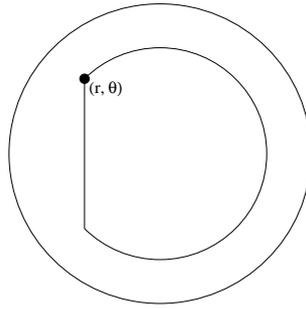
Figure 3. The loop is determined by its left upper corner.

an interior point at the other end is simply given by a path $r(\theta)$ such that $r(\pi) = 1$—the boundary is an end point; $r'(\theta) > 0$—the leaves are disjoint; and $r(0) \geq 0$—the endpoint is an interior point.

We will now construct a sweepout with maximum length loop represented by $(c, \pi/2)$ for c close enough to 1. As $c \to 1$ the maximum length loop decreases in length and converges to the minimax which is simply the line $\theta = \pm \pi/2$.

Along $\theta = \pi/2$, $L(r, \pi/2)$ is a maximum at $r \approx .83$ and decreases for $r > .83$. Along $r = 1$, $L(1, \theta)$ is increasing as θ travels from 0 to $\pi/2$ and decreasing from $\pi/2$ to π. From this it is easy to see that for $c > .83$ one can take a path from the point $(c, \pi/2)$ to $(1, \pi)$ such that L decreases along the path and similarly a path from $(c, \pi/2)$ to $(r_0, 0)$ along which L decreases. In other words there is a sweepout with maximum length loop represented by $(c, \pi/2)$ and as $c \to 1$ the maximum length loop decreases.

The above example corresponds to minimal tori in the three-sphere. We know that there is an embedded closed geodesic, corresponding to the Clifford torus, and this is missed by taking the minimax of sweepouts. The sweepouts by loops correspond to sweepouts by tori on the three-sphere. The maximal volume tori converge to a two-sphere union a one-dimensional arc meeting the two-sphere at two points. We remove the arc, leaving a minimal two-sphere.

The next example uses the metric over the triangle

$$ds^2 = y(1 - x - y)dx^2 + x(1 - x - y)dy^2 + xyd(1 - x - y)^2 \quad (17)$$

defined in Section 2.2 where we proved the existence of an embedded closed geodesic. In the following proposition we show that this geodesic is missed by taking the minimax of sweepouts. This corresponds to sweepouts of tori in $\mathbb{CP}^2$ with minimax an embedded minimal three-sphere. Inside the maximal volume T^3 of each sweepout, there

is a $[0, 1] \times T^2 \subset T^3$ that collapses to a pair of disks $[0, 1/2] \times S^1 \cup$ $[1/2, 1] \times S^1/\{1/2\} \times S^1$. So the minimax is a three-sphere union two two-dimensional disks that intersect each other at a point and the three-sphere at a circle. We remove the lower-dimensional part of the limit.

PROPOSITION 4.1. *There exists a sequence of sweepouts of loops of the triangle with metric (17) with minimax an arc that starts and ends at the incomplete point.*

Proof. Consider sweepouts of the triangle $x \geq 0$, $y \geq 0$, $x + y \leq 1$ by similar triangles symmetric under $x \leftrightarrow y$. (There are many such sweepouts, each given by a continuous monotone function encoded by the top left vertex of each triangle.)

Given any $a > 0$ small enough, we will construct a sweepout of the triangle by a family of similar triangles such that the maximum length triangle is close to the triangle T_a with edges $x = a$, $y = a$, $x + y = 3/4$. Furthermore, there is a sequence $a_i \to 0$, such that the maximum length triangles in the associated sweepouts decrease in length. The minimax of such a sequence is the triangle $x = 0$, $y = 0$, $x + y = 3/4$. We throw away $x = 0$ and $y = 0$ to be left with the geodesic $x + y = 3/4$ that meets the boundary.

The length of the path $x + y = c$ running from $x = 0$ to $y = 0$ is $l(c) = c\sqrt{2c(1 - c)}$. By considering l^2 it is easy to see that the maximum occurs at $c = 3/4$ and the function is monotone outside the maximum. The length of the triangle T_a, with edges $x = a$, $y = a$, $x + y = 3/4$, is $L(a) = (3/4 - 2a)(2\sqrt{2a(1 - a)} + \sqrt{3/8})$. Since $L'(a) \to +\infty$ as $a \to 0$, for small enough a, $L(a) > L(0) = l(3/4) > l(c)$ for $c \neq 3/4$.

Given $\epsilon > 0$, there is a $\delta > 0$ such that $l(c) < l(3/4) - \delta$ for $c > 3/4 + \epsilon$. Choose a small enough so that the path $x = a$ has length less than $\delta/2$. Foliate the outside of the triangle T_a with similar triangles (invariant under $x \leftrightarrow y$.) Then, any triangle with side $x + y = c$ and $c > 3/4 + \epsilon$ has length bounded above by $l(c) + 2\delta/2 < l(3/4) < L(a)$.

Inside T_a we can foliate by triangles each of length less than $L(a)$ as follows. The paths $x + y = c$ for $c < 3/4$ running from $x = a$ to $y = a$ have lengths less than the path $x + y = 3/4$ running from $x = a$ to $y = a$ since they are the same proportion of the full paths from $x = 0$ to $y = 0$. In the extreme, if we were to take the triangles with edges $x + y = c$, $x = a$ and $y = a$, for c running between $3/4$ and a, they would all have length strictly less than the triangle T_a. This set of triangles doesn't foliate the interior of T_a, but we can adjust them slightly to foliate, as follows. Choose the vertical $x = x(c)$ so that the increase in the lengths of the path $x = x(c)$ and the path $x = a$ from $y = a$ to $y = 3/4 - x(c)$ (respectively $y = 3/4 - a$) is exactly equal to

the decrease in the lengths of $x + y = c$ and $x + y = 3/4$ from $x = a$ to $y = a$. (If $x(c)$ gets big enough so that the latter lengths do not decrease, then choose any foliation of verticals.) The lengths of all of these triangles are less than the length of T_a since we have measured a bit more than necessary.

Thus, the maximum length triangle in the sweepout contains the side $x + y = c$ for $c < 3/4 + \epsilon$. The maximum length is bounded above by $L(a) + \delta$ where $\delta \to 0$ as $\epsilon \to 0$. Thus, we can choose a sequence $a_i \to 0$ so that the maximum length decreases. Moreover, it converges to $L(0)$ and the maximum length triangle converges to T_0, the triangle with edges $x = 0$, $y = 0$, $x + y = 3/4$.

4.2. MINIMAX IN FINITE DIMENSIONS.

Consider the problem of locating the critical points of a smooth function $f : M \to I\!R$ defined on a compact manifold M. The local minima of f can be found by flowing from a generic point along the path of steepest descent—the gradient flow, or simply any path of descent. A more sophisticated method is needed to locate critical points of higher index.

In place of a generic point in M, take a submanifold $\Sigma \subset M$, that represents a non-trivial homology class. If $[\Sigma] \in H_k(M)$ is non-trivial, then there is a point of Σ that flows to an index k critical point of M, under the gradient flow. This follows from Morse theory ideas, where the critical points of f represent cohomology classes on M (or more precisely, cochains), and the non-trivial evaluation of a cohomology class on $[\Sigma]$ detects the intersection of Σ with the stable manifold of a critical point.

From $\Sigma \subset M$, one might use the gradient flow to locate the index k critical point, although this method is limited when we adapt it to infinite-dimensional analogues where the gradient flow is not so well-behaved.

Alternatively, a minimax argument can be used. Take the point in the image of Σ that is a maximum under f. Now vary Σ in its homology class in such a way that its maximum decreases. The *minimax* of Σ is the minimum such maximum value over all homologous maps $\Sigma \to M$. One can prove that it is the index k critical point. This is because the maximum point on Σ becomes arbitarily close to the intersection of Σ with the stable manifold of the critical point, and this intersection can be brought arbitrarily close to the critical point.

An example is given in Figure 4 for the height function on a torus. Consider maps $\gamma : [0,1] \to M$ that satisfy $\gamma(0) = q = \gamma(1)$ where
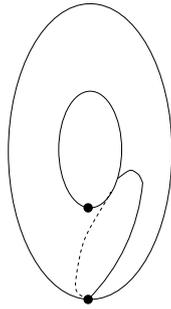
Figure 4. The index 1 critical point is a minimax.

$q \in M$ is the minimum, and the image of γ represents a non-trivial class in $H_1(M)$. The example shows that the image of γ contains a point that flows to the index 1 critical point under the gradient flow. The maximum point on γ is forced to lie above the index 1 critical point, and a minimax sequence of such maximum points will converge to the index 1 critical point.

This shows that sweepouts can give index one closed geodesics at best and that we need to take the minimax of higher dimensional families of loops to get some closed geodesics. Lusternik-Schnirelmann theory successfully does this on the infinite-dimenional manifold of loops on a manifold equipped with the length functional. More generally, Almgren [2] used geometric measure theory techniques to prove the existence of minimax limits and Pitts [12] proved regularity results for the varifolds produced from Almgren's work.

One way of producing the minimax of a family of paths on manifold is to apply a curve shortening flow to the entire family. As long as such a process is continuous, the topology of the underlying manifold can be used to deduce the existence of closed geodesics. Hass and Scott [10] constructed a curve shortening flow by covering a surface with small disks and then straightening a given path in each disk, in turn. They proved that this process produces a sequence of paths that converge to a geodesic. Furthermore, any given continuous family of paths can be shortened continuously—this requires some fixing of discontinuities. When the surface has incomplete points, this process can still be used to shorten a given curve. However, the curve shortening may not be unique and this produces discontinuities in the curve shortening of a family that cannot be fixed. Figure 5 shows two shortest paths of the same length joining a pair of points on the boundary of a small disk around the incomplete point with metric $ds^2 = dr^2 + rd\theta^2$. Arbitrarily
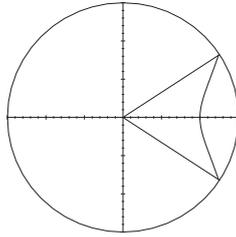
Figure 5. Equal length shortest paths joining boundary points.

small disks possess such pairs since in a disk of radius ϵ, geodesics (with θ non-constant) from the boundary that come close enough to the incomplete point have length greater than 2ϵ so the intermediate value theorem produces a geodesic from the boundary with length exactly 2ϵ and it can be shown it is minimal. This results in a choice for the curve shortening flow, which produces a discontinuity in the flow of a family. A similar type of discontinuity occurs in [10], and there they fill in the "gap" between the two shortest paths with a family of paths. In our case, the family of paths joining the two boundary points consists of paths longer than the two bounding shortest paths, and this destroys the *shortening* of a family. The key difference with [10] is that their discontinuity arises between a shortest path and a (local) longest path.

5. Minimal submanifolds.

A natural source of incomplete metrics over a surface comes from the study of minimal submanifolds in higher dimensions. If $Y \to M$ is a codimension one minimal immersion invariant under a subgroup G of isometries of M such that $\dim M/G = 2$, then it is the pull-back of a geodesic on the quotient with an adjusted metric. If g is the quotient metric on M/G, and $V(x)$ is the volume of the orbit of $x \in M$ under G (defined with multiplicity), then the adjusted metric is $V^2 g$ [9]. Since the volume of each orbit is taken into account, the length of a geodesic in the quotient is equal to the volume of its pull-back upstairs.

5.1. Three-spheres

Consider S^3 with the circle action that fixes a circle. It is convenient to respresent this with respect to a cylindrical coordinate system (r, θ, ϕ) where (θ, ϕ) parametrise tori which degenerate to circles at $r = 0$ amd $r = \pi/2$. The circle action is given by rotation by ϕ and the fixed circle is $r = \pi/2$. The quotient space of the action is a disk with boundary corresponding to the fixed point set upstairs. The lengths of the orbits go to zero as the fixed point set is approached so this produces a metric on the disk which is degenerate on the boundary, or equivalently an incomplete metric on the two-sphere minus a point. For concreteness, we will consider the special case of circle invariant metrics on S^3 of the form

$$ds^2 = dr^2 + u'(r)^2 d\phi^2 + \frac{f(u(r), \theta)^2}{u'(r)^2} d\theta^2$$

where $f(u)^2 \sim u^2(1 - u^2)$ near $u = 0$ and $u = 1$ and $u(r) \sim \sin(r)$ near $r = 0$ and $r = \pi/2$ (strictly we mean $u'(0) = 1$, etc.) This means the metric is asymptotically the same as the round metric at the two invariant circles $r = 0$ and $r = \pi/2$. The more general metric would still behave like this asymptotically, although it would not be diagonal, and the coefficient of $d\phi^2$ would depend on θ too. The quotient metric is

$$d\bar{s}^2 = dr^2 + \frac{f(u(r), \theta)^2}{u'(r)^2} d\theta^2$$

and the adjusted metric is

$$d\hat{s}^2 = u'(r)^2 dr^2 + f(u(r), \theta)^2 d\theta^2 = du^2 + f(u, \theta)^2 d\theta^2.$$

This is complete at $u = 0$ and incomplete at $u = 1$ with $\int K da = \infty$ in a neighbourhood.

More generally, the $SO(n-1)$ action on the first factor of $\mathbb{R}^{n-1} \times \mathbb{R}$ induces an action on S^n. The quotient is a disk with boundary corresponding to the fixed point set upstairs, so again we get an incomplete metric on the two-sphere minus a point. Near the incomplete point, the metric looks like

$$ds^2 = dr^2 + r^p d\theta^2, \ p = 2 - 2/n.$$

When

$$ds^2 = du^2 + u^2(1 - u^2) d\theta^2 \tag{18}$$

corresponding to the round three-sphere, the length of the second homology class of loops with $\int K da = 2\pi$ over the disks they bound, is the same as the length of the embedded closed geodesic which is π.

The length of the shortest geodesic arc is 2. Since $\pi < 4$, the length of the second homology class of loops is less than twice the length of the shortest geodesic arc. The same is true for a metric sufficiently close to (18), and furthermore a sufficiently close metric has positive Gaussian curvature. Thus we have proven Theorem 2 in the case that the circle action has a fixed point set. When the circle action is free, then the existence of an embedded minimal torus is equivalent to the existence of an embedded closed geodesic on the quotient two-sphere where the quotient metric is adjusted by the lengths of the orbits. In this case, there are at least three embedded closed geodesics, see for example [10], and hence three embedded minimal tori. Thus Theorem 2 is proven. Note that, as mentioned in the introduction, a three-sphere that gives rise to a positive Gaussian curvature incomplete two-sphere that satisfies the conditions of Theorem 1 does not have to be close to the round three-sphere.

5.2. TORIC VARIETIES

A Kahler four-manifold is *toric* if it admits an action of the torus T^2 that preserves the Kahler structure. We will denote it by its underlying symplectic manifold (M^4, ω) and prove results for a large class of compatible toric Kahler metrics. It is fibred by Lagrangian tori over a convex polygon base $P \subset \mathbb{R}^2$ with some degenerate fibres.

$$
\begin{array}{c}
T^2 \hookrightarrow \quad (M^4, \omega) \\
\downarrow \\
P \subset \mathbb{R}^2
\end{array} .
$$

The quotient equipped with the adjusted metric of Hsiang and Lawson is an incomplete metric on the two-sphere minus a point. Before we study this specific case of four-dimensional toric manifolds we will describe some background in any dimension.

Let (M^{2n}, ω) be a closed symplectic manifold equipped with an effective action of a torus T^n that preserves ω. A basic example of this is complex projective space, $\mathbb{CP}^n$, with its natural symplectic form $\partial\bar{\partial} \ln |z|^2$ where $z = (z_0, .., z_n)$ is a projective coordinate. The symplectic form is invariant under the action of $U(n+1)$, and in particular under $T^{n+1} \subset U(n+1)$ which acts as T^n (since one circle acts trivially.) Another basic example is the subset of all Hermitian matrices with a given set of eigenvalues. The torus acts by conjugation and the symplectic form uses the trace and commutator of matrices. (The first example is a case of the latter: the space of $(n+1) \times (n+1)$ Hermitian matrices with one eigenvalue 1 and n eigenvalues 0, is isomorphic, as a symplectic manifold with torus action, to $\mathbb{CP}^n$.)

Such an action is said to be *Hamiltonian*, and every $\xi \in \mathbf{t} = \mathbf{R^n}$ gives rise to the function $H_\xi : M \to I\!R$ that satisfies $dH_\xi = \omega(\tilde{\xi}, \cdot)$ for $\tilde{\xi}$ the vector field generated by ξ and the torus action. Define the *moment map* to be $\phi : M \to \mathbf{t}^* = \mathbf{R^n}$ that satisfies

$$(\phi(p), \xi) = H_\xi(p).$$

In the case of $C\!I\!P^n$,

$$\phi(z_0, .., z_n) = \left(\frac{|z_1|^2}{|z|^2}, .., \frac{|z_n|^2}{|z|^2} \right)$$

which has image $\{(x_1, .., x_n) | x_i \geq 0, \ \sum x_i \leq 1\}$. In general, the image of ϕ is a convex polytope P in $I\!R^n$. This generalises the theorem of Schur that the diagonal elements of an $n \times n$ Hermitian matrix lie in the convex hull of the points in $I\!R^n$ obtained from the eigenvalues of the matrix, listed in every possible order.

Those polytopes that arise as the image of a moment map of a torus action are characterised in the following definition.

DEFINITION 3. *A convex polytope $P \subset I\!R^n$ is Delzant if:*
   (1) there are n edges meeting at each vertex v;
   (2) the edges meeting at the vertex v are rational, i.e. $v + tv_i, 0 \leq t \leq \infty, v_i \in \mathbb{Z}^n$;
   (3) the $v_1, \ldots, v_n$ can be chosen to be a generating set of $\mathbb{Z}^n$.

The Delzant polytope P can be defined by inequalities

$$l_r(x) = \langle x, \mu_r \rangle - \lambda_r \geq 0 \tag{19}$$

where μ_r is a primitive element of $\mathbb{Z}^n$ giving the inward pointing normal of the rth face of P. For each Delzant polytope P, Guillemin constructed a symplectic manifold M_P using the beautiful idea of varying the numbers λ_r. The variations generate a torus action, and one obtains M_P as a symplectic quotient of this action. See [6] for details. There is a bijective correspondence between Delzant polytopes and toric manifolds $P \mapsto M_P$.

THEOREM 5.1 ([4]). *Let (M^{2n}, ω) be a closed symplectic manifold equipped with an effective action of a torus T^n that preserves ω and moment map $\phi : M \to \mathbf{t}^* = \mathbf{R^n}$. Then the image of ϕ is a Delzant polytope, and M and the T^n action are isomorphic to M_P.*

Toric manifolds are Kahler manifolds and so far we have only described a symplectic structure. When we also specify a (compatible)

complex structure, this together with the symplectic structure gives rise to the Kahler metric.

The metric is usually expressed with respect to one of two *canonical coordinate systems*. Denote by $M^\circ \subset M$ those points on which the torus action is free. The torus action complexifies and allows us to describe M° in complex coordinates by

$$M^\circ \cong C^n/2\pi i Z^n = R^n \times iT^n = \{u + iv : u \in R^n, v \in R^n/Z^n\}.$$

Alternatively, one can use action-angle coordinates:

$$M^\circ \cong P^\circ \times T^n = \{(x,y) : x \in P^\circ \subset R^n, y \in R^n/Z^n\} \qquad (20)$$

where P° is the interior of the Delzant polytope.

The quotient manifold can be identified with either R^n or the interior of the Delzant polytope P°, equipped with the quotient metric. The metric on the toric manifold is expressed in terms of functions over the quotient explicitly using the quotient metric. For our purposes, the action-angle coordinate system is more suitable, since one can extend to the compact quotient. When we adjust the quotient metric by the volumes of the fibres, the new metric is degenerate at the faces of the Delzant polytope.

We will follow Abreu's treatment [1] of Guillemin's construction [6] of metrics over toric manifolds. With respect to the coordinates (20), the symplectic form is the standard $\omega = \sum_j dx_j \wedge dy_j$, given in matrix form by

$$\begin{pmatrix} 0 & \vdots & I \\ \cdots & \cdots & \cdots \\ -I & \vdots & 0 \end{pmatrix}.$$

The complex Kahler structure J is given by

$$\begin{pmatrix} 0 & \vdots & -G^{-1} \\ \cdots & \cdots & \cdots \\ G & \vdots & 0 \end{pmatrix}$$

where $G = \mathrm{Hess}(h)$ is the Hessian of a potential $h(x)$,

$$G = \left[\frac{\partial^2 h}{\partial x_j \partial x_k} \right], \ 1 \leq j, k \leq n.$$

We can now describe the potential $h(x)$ defined over the polytope P, using the defining expressions for P (19). Define

$$h_P(x) = \frac{1}{2} \sum_{r=1}^{d} l_r(x) \log l_r(x).$$

This is well-defined on the interior of P since $l_r(x) > 0$ there. More generally, choose any function f smooth on the whole of P, and set $h = h_P + f$. It follows that

$$\det(G) = \left(\delta(x) \prod_{r=1}^{d} l_r(x) \right)^{-1} \tag{21}$$

for $\delta(x)$ a strictly positive smooth function on P.

The quotient metric on P is given by the $n \times n$ matrix G. It is a well-defined finite metric at the boundary of P although it looks like it might blow up there. For example, on $C\!I\!P^2$ the Fubini-Study metric is described by the canonical potential on the Delzant polygon $x_1 \geq 0$, $x_2 \geq 0$ and $x_1 + x_2 \leq 1$,

$$h_P = x_1 \ln x_1 + x_2 \ln x_2 + (1 - x_1 - x_2) \ln(1 - x_1 - x_2).$$

The quotient metric is the round metric on an octant of the two-sphere (identified with a triangle) which is certainly finite at the boundary of the octant.

We can associate minimal surfaces in the quotient to minimal surfaces in the toric manifold upstairs once we adjust the quotient metric by the volumes of the torus fibres. From (21), the volume $V(x)$ of the torus fibre over $x \in P$ is given by $V(x)^2 = \det G^{-1}(x)$.

Replace the quotient metric G by

$$g(x) = V(x)^{2/(n-1)} G(x) = \det G^{1/(1-n)}(x) G(x).$$

By (21), this satisfies $\det g(x) = \left(\delta(x) \prod_{r=1}^{d} l_r(x) \right)^{1/(n-1)}$ which vanishes on the boundary of P. When $n = 2$, it vanishes linearly and explicitly when $l_r = a_r x + b_r y + c_r$ we get:

$$g(x) = \sum \frac{1}{2 l_r} \begin{pmatrix} a_r^2 & a_r b_r \\ a_r b_r & b_r^2 \end{pmatrix}$$

which has determinant $\delta(x)/(\prod l_r)$ where $\delta(x) \neq 0$ on the closed polygon. Equivalently, the quotient metric is

$$d\bar{s}^2 = \frac{1}{2} \sum \frac{dl_r^2}{l_r}$$

so the adjusted metric is

$$ds^2 = \frac{1}{2\delta} \sum \prod_{j \neq r} l_j dl_r^2.$$

At each edge $l_r = 0$ the metric is asymptotic to (1).

Large families of metrics can be obtained for a given polygon this way. Simply add "edges" to the polygon that lie outside the polygon. For example, over $\mathbb{CP}^2$ with polygon defined by $x \geq 0$, $y \geq 0$ and $1 - x - y \geq 0$ we can include the inequality $l(x,y) = 2 - x - y \geq 0$ and still get a positive definite metric on the interior of the triangle. We may also take multiples $\epsilon l(x,y)$ for an extra edge, which is desirable when we choose ϵ small so that the new metric is close to the original metric. (In fact one can take $h_P(x) + f(x)$ for any smooth function $f(x)$ defined on a neighbourhood of P, and when $f(x)$ is small enough the new metric remains positive definite.)

For the Fubini-Study metric on $\mathbb{CP}^2$ we proved the existence of an embedded closed geodesic. An easier calculation shows that the shortest arc is given by the line $x = 1/4$ which has length $3\sqrt{3}/8\sqrt{2} \approx .46$. The maximum length triangle with $\int K da = 2\pi$ on its interior is less than .9 which is less than $2 \times 3\sqrt{3}/8\sqrt{2}$. Thus, the shortest geodesic is long and we can apply Theorem 4 to get Theorem 3. The same reasoning applies to $S^2 \times S^2$ with the product of round metrics of the same area.

The Fubini-Study metric on $\mathbb{CP}^2$ is invariant under a $\mathbb{Z}_3$ action. We can drop the condition that a metric on $\mathbb{CP}^2$ be sufficiently close to the Fubini-Study metric and instead require that it is invariant under the $\mathbb{Z}_3$ action and that the adjusted metric on the triangle has positive Gaussian curvature to again deduce the existence of an embedded closed geodesic.

The geodesic curvature flow proof corresponds to mean curvature flow in $\mathbb{CP}^2$. The *mean curvature* of a hypersurface $Y^3 \subset M^4$ is a function on Y^3 given by the trace of the second fundamental form of Y^3. As a (symmetric) bilinear form, the second fundamental form is defined by $l(X,Y) = \langle \nabla_X \nu, Y \rangle$ for $X, Y \in T_p Y^3$, ν the unit normal vector field to Y^3 and ∇ the Levi-Civita connection on M^4.

The mean curvature flow of a surface Y is an evolution of Y in its normal direction with magnitude given by the mean curvature:

$$\frac{dY_t}{dt} = \text{mean curvature} \cdot \nu.$$

In actual fact, we need to adjust the flow further, by multiplying the right-hand side by a function defined on M^4 (and hence independent of the way in which Y embeds.) For a hypersurface invariant under the torus action we have

$$\text{mean curvature} = \frac{1}{\text{area of torus}} k$$

where k is the geodesic curvature of the corresponding curve on the polygon. The reciprocal of the Gaussian curvature of the polygon is an

invariant function ψ on the polygon that seems not to have a natural interpretation. Thus the flow in the toric surface is given by

$$\frac{dY_t}{dt} = \psi \cdot (\text{area of torus})^2 \cdot \text{mean curvature} \cdot \nu$$

where an extra factor of the area of the torus arises from comparing the normal vectors in two-dimensions and four-dimensions. The factor $\psi \cdot (\text{area of torus})^2$ is a function on the toric surface that vanishes on the divisor of $C\!I\!P^2$ on which T^2 does not act freely.

6.   Counterexample

It is not always true that there exists an embedded closed geodesic on an incomplete two-sphere. A counterexample uses the metric:

$$ds^2 = \sin^2(y)(dx^2 + dy^2), \ (x,y) \in [-R, R] \times (0, \pi). \tag{22}$$

Choose R very large and cap off the metric at the two ends in such a way that any geodesic in the cap must leave the cap. (It can be chosen to have positive Gaussian curvature.) For example, choose

$$ds^2 = \cos^2(|x| - R)\sin^2 y(dx^2 + \cos^2(|x| - R)dy^2), \ |x| \in [R, R + \pi/2). \tag{23}$$

Alternatively, the metric can be capped off smoothly.

THEOREM 5.   *There is no embedded closed geodesic for the metric (22), (23).*

   Proof. We call a geodesic vertical if it is given by $x =$ constant for $x \in [-R, R]$. The theorem follows from the following facts:

   (i) any non-vertical geodesic cannot be vertical anywhere in $x \in [-R, R]$;

   (ii) any two non-vertical geodesics must intersect;

   (iii) a closed geodesic must enter $x \in [-R, R]$ twice and hence it must be self-intersecting.

   Uniqueness of geodesics gives (i) immediately. For (ii) we use the fact that the geodesic flow is integrable and translation invariant in $x \in [-R, R]$, as dealt with in Section 2.1, to deduce that inside $x \in [-R, R]$ each geodesic oscillates around $y = \pi/2$ with period

$$\begin{aligned} \Omega_c &= 4\int_{\sin x = c}^{\pi/2} \frac{c}{\sqrt{\sin^2 x - c^2}}dx \\ &= 4c\int_c^1 \frac{du}{\sqrt{(u^2 - c^2)(1 - u^2)}} \end{aligned}$$

$$< \; 4\sqrt{\frac{c}{2(1+c)}} \int_c^1 \frac{du}{\sqrt{(u-c)(1-u)}}$$

$$= \; 4\pi\sqrt{\frac{c}{2(1+c)}}.$$

where c denotes the geodesic that reaches a maximum value of $y = c^2$ in $x \in [-R, R]$. The period is bounded and we choose R to be greater than (half) this period so that each non-vertical geodesic must meet the line $y = \pi/2$. Since any geodesic meets $y = \pi/2$ with a given period, any two non-vertical geodesics meet in $x \in [-R, R]$.

Now a geodesic in the cap (that does not meet the boundary) must leave the cap since the equation for geodesics in the cap is

$$\ddot{x} \; = \; \dot{x}^2 \tan(x - R) - 2\dot{x}\dot{y}\cot y - \dot{y}^2 \sin 2(x - R)$$

$$\ddot{y} \; = \; \dot{x}^2 \sec^2(x - R)\cot y + 4\dot{x}\dot{y}\tan(x - R) - \dot{y}^2 \cot y$$

Thus $\dot{x} = 0 \;\Rightarrow\; \ddot{x} < 0$ so no loops can occur inside the cap. By (ii), the two arcs of any closed geodesic must intersect in $x \in [-R, R]$ so (iii) follows and the theorem is proven.

Remark. The proof of Theorem 5 uses the explicit knowledge of geodesics given by the integrability of the geodesic flow. There is an alternative proof that uses less structure which we only sketch. Suppose R is very large and there are two geodesics travelling from $x = -R$ to $x = R$ without intersecting. If we choose a point strictly between the geodesics at $x = R$ and a point strictly between the geodesics at $x = -R$ then there is a shortest path joining them which must have index 0. The two geodesics form barriers so the shortest path lies between them. We can assume the two geodesics do not come close to the boundary $y = 0$ or π since then they would be quite steep and intersect, so in particular the shortest path stays away from the boundary. Thus its length is comparable to the length along $y = \pi/2$, so for large enough R the length of the shortest path is large. The Gaussian curvature on the disk is bounded below so the Rauch comparison theorem forces the shortest path to have index at least one which is a contradiction. We conclude that the two geodesics intersect.

Geodesics on the disk with metric (22), (23) arise from minimal surfaces on S^3 equipped with the circle invariant metric obtained by stretching the round metric in a neighbourhood of a minimal two-sphere, or more precisely $S^2 \times \mathbb{R}$ with the round metric times the flat metric capped off with round three balls. The counterexample shows that there is no minimally embedded $T^2 \subset S^3$ invariant under the circle action. It would be interesting to know if there is any minimally

38

embedded T^2 in this three-sphere. White [14] proves that for any C^4 metric on S^3 with positive Ricci curvature there exists at least one embedded minimal torus. Our counterexample (which can be smoothed) has non-negative Ricci curvature.

The following corollary shows that condition (ii) in Theorem 1 does not hold for the metric in the counterexample.

COROLLARY 6.1. *There is a length decreasing sequence of loops that converges to a double geodesic arc for the metric (22), (23).*

Proof. Take a loop that bounds a region with $\int K da = 2\pi$ (where $K =$Gaussian curvature) that is contained inside $x \in [-R, R]$ and that is symmetric under reflection in $y = \pi/2$. Under the flow (10) the loop cannot move past vertical geodesics and $\int K da = 2\pi$ is preserved. It cannot converge to a closed geodesic since no vertical tangencies are allowed so it travels to both boundaries symmetrically, converging to a double vertical geodesic. The flow is length decreasing so the corollary follows.

References

1. Abreu, M. *Kahler geometry of toric manifolds in symplectic coordinates*, Symplectic and contact topology: interactions and perspectives (Toronto/Montreal, 2001), AMS, Providence, RI, Fields Inst. Commun., **35**, (2003), 1-24.

2. Almgren, F. J., Jr. *The structure of limit varifolds associated with minimizing sequences of mappings*, Symposia Mathematica, Vol. XIV, Academic Press, London, (1974), 413–428.

3. Chou, Kai-Seng, Zhu, Xi-Ping. *The curve shortening problem.* Chapman & Hall/CRC, Boca Raton, FL, (2001).

4. Delzant, T. *Hamiltoniens périodiques et image convex de l'application moment*, Bull. Soc. Math. France, **116** (1988), 315-339.

5. Gage, M. E. *Deforming curves on convex surfaces to simple closed geodesics*, Indiana Univ. Math. J. **39** (1990), 1037–1059.

6. Guillemin, V. *Kahler structures on toric varieties*, J. Diff. Geom. **40** (1994), 285-309.

7. Guruprasad, K. and Haefliger, A. *Closed geodesics on orbifolds*, arXiv:math.DG/0306238

8. Hsiang, W-T., Hsiang, W-Y. and Tomter, P. *On the existence of minimal hyperspheres in compact symmetric spaces*, Ann. Sci. Ec. Norm. Sup., **21** (1988), 287-305.

9. Hsiang, W-Y. and Lawson, H.B. *Minimal submanifolds of low cohomogeneity*, J. Diff. Geom. **5** (1971), 1-38.

10. Hass, J. and Scott, P. *Shortening curves on surfaces*, Topology **33** (1994), 25-43.

11. Milnor, J. *Morse Theory*, Annals of Mathematics Studies **51**. Princeton, N.J., Princeton University Press (1963).

12. Pitts, Jon T. *Regularity and singularity of one dimensional stationary integral varifolds on manifolds arising from variational methods in the large.* Symposia Mathematica, Vol. XIV, Academic Press, London, (1974), 465–472.

13. Pitts, J.T. and Rubinstein, J.H. *Applications of minimax to minimal surfaces and the topology of 3-manifolds*, Miniconference on geometry and partial differential equations, 2 (Canberra, 1986), Proc. Centre Math. Anal. ANU, **12** (1987), 137-170.

14. White, Brian *The space of minimal submanifolds for varying Riemannian metrics.* Indiana Univ. Math. J. **40** (1991), 161–200.