MOHAMMAD GHOMI'S SOLUTION TO THE ILLUMINATION PROBLEM

RALPH HOWARD
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTH CAROLINA
COLUMBIA, S.C. 29208, USA
HOWARD@MATH.SC.EDU

Contents

1. Introduction

Let M^2 be a compact oriented surface immersed in $\mathbf{R}^3$ with unit normal field $\mathbf{n}$. Recently Mohammad Ghomi has proven the following very pretty result.

Theorem. *If for all unit vectors e the sets $\{x \in M^2 : \langle \mathbf{n}(x), e\rangle > 0\}$ are simply connected, then M^2 is the boundary of a bounded convex set in $\mathbf{R}^3$.*

This problem has been around for a while in the from of the "illumination conjecture". Ghomi also gave an example to show that "simply connected" can not be weakened to "connected" as there is an imbedded torus M^2 in $\mathbf{R}^3$ so that all of the sets $\{x \in M^2 : \langle \mathbf{n}(x), e\rangle > 0\}$ are connected (see Figure 7).

Below is my exposition of Ghomi's results. Any errors are mine and not his.

Date: December 1998 with Sections 6 and 7 added in April 1999. Minor corrections added October 2003.

2. Elementary Results about the Boundaries of Shadows

Let M^2 be a compact oriented surface immersed in $\mathbf{R}^3$. The orientation of M^2 determines a unit normal field $\mathbf{n}$ along M^2 in the usual manner. We now fix our notation for the imbedding invariants of M^2 in $\mathbf{R}^3$. Let $\overline{\nabla}$ be the usual flat connection on $\mathbf{R}^3$. Then the second fundamental form $I\!I$ and the induced connection ∇ on M^2 are defined by

$$\overline{\nabla}_X Y = \nabla_X Y + I\!I(X,Y)\mathbf{n}$$

where X and Y are smooth vector fields defined on M^2 and $\nabla_X Y$ is the component of $\overline{\nabla}_X Y$ tangent to M^2. The Weingarten map A (or shape operator) is the linear map defined on vectors tangent to M^2 by

$$AX := -\overline{\nabla}_X \mathbf{n}$$

and is related to the second fundamental form by $\langle AX, Y\rangle = I\!I(X,Y)$.

Let S^2 be the unit sphere in $\mathbf{R}^3$ and for each $e \in S^2$ define a function $f_e \colon M^2 \to \mathbf{R}$ by

$$f_e(x) := \langle \mathbf{n}(x), e\rangle.$$

Then the ***shadow set*** defined by e is

$$S_e := \{x \in M^2 : f_e(x) > 0\}.$$

The boundary ∂S_e is the ***shadow boundary*** in the direction e. We now show that for most choices of $e \in S^2$ the shadow boundary is a smooth curve.

Proposition 2.1. *For almost all $e \in S^2$ (in the sense of Lebesgue measure on S^2) the set $\{x \in M^2 : f_e(x) = 0\}$ is a smooth curve in M^2. Thus for these e we have that $\partial S_e = \{x \in M^2 : f_e(x) = 0\}$ and so ∂S_e is a smooth curve for almost all $e \in S^2$.*

Proof. Let $U(M) := \{(x,u) : x \in M^2, u \in T(M)_x, \|u\| = 1\}$ be the unit sphere bundle of M^2. Define a map $\tau \colon U(M) \to S^2$ by $\tau(x,u) = u$ (where u is translated to the origin). By Sard's Theorem almost every $e \in S^2$ is a regular value of τ and for such e the preimage $\tau^{-1}[e]$ is a smooth curve in $U(M)$. If $(x,u) \in U(M)$ and $u_1 \in T(M)$ is a unit vector with $\langle u, u_1\rangle = 0$, then $c(t) = (x, \cos(t)u + \sin(t)u_1)$ parameterizes the fiber $U(M)_x$ and $c(0) = (x,u)$. Also

$$\tau_* c'(0) = \frac{d}{dt}\tau(x, \cos(t)u + \sin(t)u_1)\Big|_{t=0} = \frac{d}{dt}(\cos(t)u + \sin(t)u_1)\Big|_{t=0} = u_1$$

and so $\tau_* c'(0) \neq 0$. If e is a regular value of τ and $(x,u) \in \tau^{-1}[e]$ then the tangent space to $\tau^{-1}[e]$ is $\{X \in T(U(M))_{(x,u)} : \tau_* X = 0\}$. As $\tau_* c'(0) \neq 0$ this implies that the curve $\tau^{-1}[e]$ is never tangent to any of the fibers $U(M)_x$.

Let $\pi \colon U(M) \to M^2$ be the natural projection and e a regular value of τ. Then a chase through the definitions involved shows the map $\pi\big|_{\tau^{-1}[e]} \colon \tau^{-1}[e] \to S^2$ is injective. Because $\tau^{-1}[e]$ is never tangent to the fibers $\pi^{-1}[x] = U(M)_x$ the map $\pi\big|_{\tau^{-1}[e]}$ is an immersion. As $\tau^{-1}[e]$ is compact this implies that the projection $\pi\big[\tau^{-1}[e]\big]$ is a smooth embedded curve in M^2. To finish the proof note that $\{f_e = 0\}$ is equal to this projection:

$$\begin{aligned}\pi\big[\tau^{-1}[e]\big] &= \{x \in M^2 : e \in T(M)_x\} = \{x \in M^2 : \langle e, \mathbf{n}(x)\rangle = 0\} \\ &= \{x \in M^2 : f_e(x) = 0\}.\end{aligned}$$

□

Corollary 2.2. *Let M^2 be a compact oriented surface immersed in $\mathbf{R}^3$ so that all the shadow sets S_e are simply connected. Then M^2 is diffeomorphic to the sphere.*

Proof. As in the last proof choose e to be a regular value of τ. Then ∂S_e is a smooth curve. As S_e is simply connected the closure $\overline{S}_e$ will be diffeomorphic to a closed disk. The boundary of S_{-e} is same as the boundary of S_e and so $\overline{S}_{-e}$ is also a closed disk. The surface M^2 is the disjoint union of S_e, S_{-e} and ∂S_e. Therefore M^2 is a pair of disks glued together along their boundaries and thus M^2 is a sphere. □

If is possible to be more precise about at which points the shadow boundaries ∂S_e are smooth curve. We first compute the differential of f_e. For a vector X tangent to M we have

$$df_e(X) = X\langle \mathbf{n}, e\rangle = \langle \overline{\nabla}_X \mathbf{n}, e\rangle + \langle \mathbf{n}, \overline{\nabla}_X e\rangle = -\langle A(X), e\rangle$$

as $\overline{\nabla}_X e = 0$. If X is tangent to M at a point $x \in \partial S_e$ then $f_e(x) = \langle e, \mathbf{n}(x)\rangle = 0$ and so $e \in T(M)_x$. Thus at points $x \in \partial S_e$ we can use that A is selfadjoint to write

$$df_e(X) = -\langle X, Ae\rangle.$$

This leads to:

Proposition 2.3. *If $x \in \partial S_e$ and $Ae \neq 0$ then ∂S_e is a smooth curve near x. In particular if the Gaussian curvature $K = \det(A)$ does not vanish at x then ∂S_e is a smooth curve near x for all $e \perp \mathbf{n}(x)$.*

Proof. If $Ae \neq 0$ then $df_e = -\langle \cdot, A(e)\rangle$ is a not zero at x and thus $\{f_e = 0\}$ is a smooth curve by the implicit function theorem. At points x where $K \neq 0$ we have $Ae \neq 0$ for all $e \neq 0$ and $e \perp \mathbf{n}(x)$ (which is the condition that $e \in T(M)_x$). □

3. An Application of Chern-Lashof Theory

We recall the basics of the Chern-Lashof theory as applied to surfaces in $\mathbf{R}^3$ (the original results work in all dimensions and codimensions). Let M^2 be a compact oriented surface immersed in $\mathbf{R}^3$. For any unit vector $u \in S^2$ define the height function $h_u\colon M^2 \to \mathbf{R}$ in the direction u as $h_u(x) := \langle x, u\rangle$. Recall that if $\mathbf{n}$ is the unit normal field along M^2 then the Gauss map of M^2 is the function $g\colon M^2 \to S^2$ given by $g(x) = \mathbf{n}(x)$ translated to the origin. Also recall that a function is a Morse function iff all of its critical points are non-degenerate. That is at any critical point the Hessian is non-singular.

Proposition 3.1. *If $u \in S^2$ and both u and $-u$ are regular values of the Gauss map then the height function h_u is a Morse function on M^2. For any such height function the Gauss curvature is non-zero at all critical points of h_u. Therefore by Sard's Theorem h_u is a Morse function for almost all $u \in S^2$ so that the Gauss curvature is non-zero at all critical points. As M^2 is compact the set of such u is an open set (as the set of critical values is the continuous image of the set of critical points which is compact).*

Proof. The derivative of the Gauss map is $g_{*x}X = -AX$ (where we are identifying the tangent spaces $T(M)_x$ and $T(S^2)_{g(x)}$ by parallel translation) and so x is a critical point of g if and only if $\det(A_x) = 0$. But the Gauss curvature is $K(x) = \det(A_x)$ and thus the set of critical points of g is just the set where the Gauss curvature vanishes. Now let u be a regular value of g. Then any x with $g(x) = u$ is a regular point of g and so $K(x) \neq 0$. The differential of h_u is $dh_u(X) = \langle X, u\rangle$ and so the critical points of the height function h_u are exactly the points x with $g(x) = \pm u$. Let x be a critical point of h_u, so that $u = \pm\mathbf{n}$, and let X, Y be smooth vector fields on M^2 so that $\nabla X = \nabla Y$ at x. Then we can compute the Hessian of h_u at x:

$$\begin{aligned} D^2h_u(X,Y) &= X dh_u(Y) = X\langle Y, u\rangle = \langle \overline{\nabla}_X Y, u\rangle \\ &= \langle \nabla_X Y + I\!I(X,Y)\mathbf{n}, u\rangle = \pm I\!I(X,Y) = \pm\langle AX, Y\rangle. \end{aligned}$$

But as $\det(A_x) = K(x) \neq 0$ this implies the Hessian is non-degenerate. □

If $u \in S^2$ is so that h_u is a Morse function on M^2 then let $C(h_u) := \{x \in M^2 : (dh_u)_x = 0\}$ be the set of critical points of h_u (which will be a finite set when h_u is a Morse function). Let $\#C(h_u)$ be the number of points in $C(h_u)$. Then the following is implicit in the paper [2] of Chern and Lashof.

Theorem 3.2 (Chern-Lashof [1, 2])**.** *If M^2 is a compact oriented surface immersed in R^3 then*

$$\frac{1}{2}\int_{S^2} \#C(h_u)\, du = \int_{M^2} |K(x)|\, dx. \tag{3.1}$$

(Here du is the area measure on S^2 and dx is the area measure on M^2.) Note that for almost all $u \in S^2$ the function h_u is a Morse function and thus $\#C(h_u)$ is finite. Since every smooth function on M^2 has at least two critical points (corresponding to the maximum and minimum) the inequality

$$\int_{M^2} |K(x)|\, dx \geq 4\pi$$

holds. Equality holds in this inequality if and only if M^2 is the boundary of a bounded convex set. □

Let $f\colon M^2 \to \mathbf{R}$ be a Morse function. If x is a critical point of f the index of f at x is defined to be in index of the Hessian $D^2 f_x$ at x. (The index of a symmetric matrix is the number of negative eigenvalues. The index of a symmetric bilinear form is the index of any matrix that represents it.) Let $\iota_k(f)$ be the number of critical points of f of index k. Then $\iota_0(f)$ is the number of local minima of f, $\iota_2(f)$ is the number of local maxima of f and $\iota_1(M)$ is the number of saddle points of f. Let $\beta_k(M^2) = \dim H_k(M, \mathbf{R})$ be the k-th Betti number of M^2. Then $\beta_0(M^2) = \beta_2(M^2) = 1$ and $\beta_1(M^2) = 2g$ where g is the genus of M^2. Then basic results in Morse theory [3] imply

$$\iota_0(f) - \iota_1(f) + \iota_2(f) = 2 - 2g$$
$$\iota_0(f),\ \iota_2(f) \geq 1$$
$$\iota_1(f) \geq \beta_1(M^2) = 2g.$$

As an aside we note these inequalities yield that for a surface of genus g that for any Morse function $f\colon M^2 \to \mathbf{R}$ that the number of critical points of f satisfies

$$\#C(f) = \iota_0(f) + \iota_1(f) + \iota_2(f) \geq 2 + 2g.$$

Using this with the Chern-Lashof theorem yields

$$\int_{M^2} |K(x)|\, dx \geq 4\pi(1 + g)$$

for compact oriented surfaces of genus g immersed in $\mathbf{R}^3$.

If M^2 is a sphere so that $g = 0$ then for any Morse function f on M^2 we have $\iota_0(f) - \iota_1(f) + \iota_2(f) = 2$. Thus if f has only one local maximum and one local minimum then $\iota_1(f) = 0$. The total number of critical points is then $\#C(f) = \iota_0(f) + \iota_1(f) + \iota_2(f) = 2$. This observation can be combined with the Chern-Lashof theorem to give the following (which can be proven by more elementary methods, but I like Chern-Lashof theory).

Proposition 3.3. *Let M^2 be a sphere immersed in $\mathbf{R}^3$ in such a way that for almost all $u \in S^2$ the height function h_u has only one local maximum. Then M^2 is the boundary of a bounded convex set in $\mathbf{R}^3$.*

Proof. As a local minimum of h_u is a local maximum of h_{-u} for almost all $u \in S^2$ the function will have at most one local maximum and at most one local minimum and thus exactly one of each. As above this implies $\#C(h_u) = 2$. But using $\#C(h_u) = 2$ for almost all u in (3.1) gives $\int_{M^2} |K(x)|\, dx = 4\pi$. Then Theorem 3.2 implies that M^2 is the boundary of a convex set. This completes the proof. □

4. Nicely Placed Triples in Oriented Surfaces

We now study the elementary combinatorics of triples of points on the boundaries of simply connected sets in compact surfaces.

Definition 4.1. Let M^2 be a compact oriented surface. Then (p_1, p_2, p_3, U) is a ***nicely placed triple*** in M^2 iff

1. U is a simply connected open set in M^2,
2. p_1, p_2, and p_3 are distinct points on the boundary, ∂U, of U, and
3. for each $i = 1, 2, 3$ there is an open neighborhood V_i of p_i in M^2 so that $V_i \cap \partial U$ is a smooth curve. □

Note that while the boundary of U in a nicely placed triple (p_1, p_2, p_3, U) is smooth near each of the points p_i, there is no regularity assumption on other parts of the boundary. (See Figure 1).

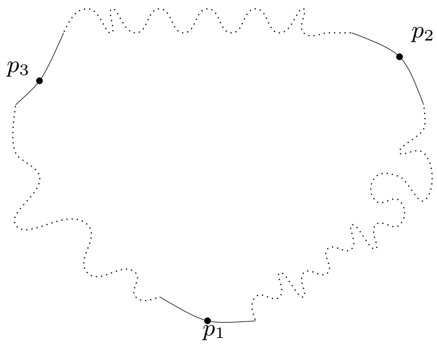

FIGURE 1. A nicely placed triple (p_1, p_2, p_3, U). The boundary, ∂U, of U is smooth near p_1, p_2, and p_3, but no regularity is required on the rest of ∂U.

If (p_1, p_2, p_3, U) is a nicely placed triple then for any pair of the points p_1, p_2, p_3, to be definite say the pair p_1 and p_2, a curve $\gamma\colon [0,1] \to M^2$ is a ***standard curve*** iff γ is a smooth embedded curve with $\gamma(0) = p_1$, $\gamma(1) = p_2$ and $\gamma(t) \in U$ for all $t \in (0,1)$. As U is simply connected any two standard curves $\gamma, \tilde{\gamma}\colon [0,1] \to M^2$ from p_1 to p_2 will be homotopic to each other through a family of standard curves. Also the curve γ will divide U into two simply connected regions, one on each side of γ.

Start with a standard curve $\gamma_{12}\colon [0,1] \to M^2$ from p_1 to p_2. Now p_3 will lie in the closure of one of the connected components of $U \setminus \gamma_{12}\big[(0,1)\big]$. As this component is simply connected there is a standard curve $\gamma_{2,3}\colon [0,1] \to M^2$ connecting p_2 to p_3 and which is disjoint from γ_{12} except at p_2. Then $U \setminus (\gamma_{12}\big[(0,1)\big] \cup \gamma_{23}\big[(0,1)\big])$ has three simply connected componets and only one of them has both p_1 and p_3 in its closure. Then there is a standard curve $\gamma_{31}\colon [0,1] \to M^2$ from p_3 to p_1 and so that γ_{31} is disjoint from both of γ_{12} and γ_{23} excepty at the points p_1 and p_3. Call such a triple of curves γ_{12}, γ_{23} and γ_{31} a ***standard triangle*** for (p_1, p_2, p_3, U). (Figure 2 shows a standard triangle for (p_1, p_2, p_3, U).) Again using that U is simply connected if $\tilde{\gamma}_{12}$, $\tilde{\gamma}_{23}$ and $\tilde{\gamma}_{31}$ is another standard triangle for (p_1, p_2, p_3, U) then $\tilde{\gamma}_{12}$, $\tilde{\gamma}_{23}$, $\tilde{\gamma}_{31}$ is homotopic to γ_{12}, γ_{23}, γ_{31} through standard triangles. That is standard triangles are unique up to homotopy.

Any standard triangle γ_{12}, γ_{23}, γ_{31} for (p_1, p_2, p_3, U) bounds a simply connected region, $\operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})$, of U which can be defined as follows. Let U_{12} be the component of $U \setminus \gamma_{12}\big[(0,1)\big]$ that contains p_3 in its closure, U_{23} the component of $U \setminus \gamma_{23}\big[(0,1)\big]$ that contains p_1 in its closure, and U_{31} the component of $U \setminus \gamma_{31}\big[(0,1)\big]$ that contains p_2 in its closure. Then $\operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31}) = U_{12} \cap U_{23} \cap U_{31}$. (If we wish to indicate the dependence on the set U we will write $\operatorname{Int}_U(\gamma_{12}, \gamma_{23}, \gamma_{31})$.) Then $\operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})$ has

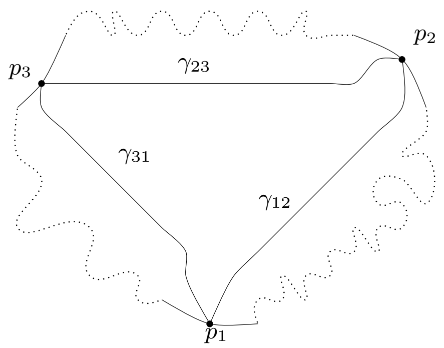

FIGURE 2. This shows a standard triangle for the nicely place triple (p_1, p_2, p_2, U). As U is simply connected any pair of standard triangles for (p_1, p_2, p_2, U) will be homotopic.

a piecewise smooth boundary and as M^2 is oriented there is an induced boundary orientation on $\partial \operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31}) = \gamma_{12} \cup \gamma_{23} \cup \gamma_{31}$. (Figure 3 shows one of the two possible boundary orintations on $\partial \operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})$.) This boundary orientation determines a premutation, $\sigma = \sigma(p_1, p_2, p_3, U)$ of the set $\{p_1, p_2, p_3\}$ in a natural manner. That is if when moving along $\partial \operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})$ in the direction of the orientation when leaving p_i the next of the points encountered is p_j then $\sigma(p_i) = p_j$. (In Figure 3 with the pictured boundary orientation we have (using the standard cycle notation for permutations) $\sigma(p_1, p_2, p_3, U) = (p_1p_2p_3)$. If the boundary orientation were reversed then $\sigma(p_1, p_2, p_3, U) = (p_1p_3p_2)$.) As standard triangles are unique up to homotopy it follows that $\sigma(p_1, p_2, p_3, U)$ is independent of the choice of the standard triangle used to define it. For future use we record the following elementary result whose proof is left to the reader.

Lemma 4.2. *If (p_1, p_2, p_3, U) is a standard triple and ∂U is a smooth curve, then $\sigma(p_1, p_2, p_3, U)$ is the permutation of $\{p_1, p_2, p_3\}$ defined by moving along ∂U with the induced boundary orientation.* □

Let (p_1, p_2, p_3, U) be a standard triple then another standard triple $(p_1, p_2, p_3, \widetilde{U})$ is ***adjacent*** to (p_1, p_2, p_3, U) iff there is some standard triangle $\gamma_{12}, \gamma_{23}, \gamma_{31}$ for (p_1, p_2, p_3, U) so that γ_{12}, γ_{23}, and γ_{31} are also standard curves in $\widetilde{U}$ and $\operatorname{Int}_U(\gamma_{12}, \gamma_{23}, \gamma_{31}) \subset \widetilde{U}$. It then follows that $\gamma_{12}, \gamma_{23}, \gamma_{31}$ is a standard triangle in $\widetilde{U}$ and $\operatorname{Int}_U(\gamma_{12}, \gamma_{23}, \gamma_{31}) = \operatorname{Int}_{\widetilde{U}}(\gamma_{12}, \gamma_{23}, \gamma_{31})$. Therefore the following is clear.

Lemma 4.3. *Let (p_1, p_2, p_3, U) and $(p_1, p_2, p_3, \widetilde{U})$ be adjacent standard triples in M^2. Then $\sigma(p_1, p_2, p_3, U) = \sigma(p_1, p_2, p_3, \widetilde{U})$. (See Figure 4.)* □

Note that for two standard triples to be adjacent all that matters is that they have a common standard triangle. They do not have to be close in any other sense, such as in the Hausdorff metric on closed sets.

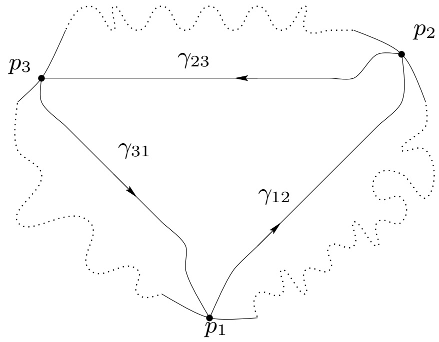

FIGURE 3. A standard triangle for (p_1, p_2, p_3, U) bounds a region of U. The orientation of M^2 induces a boundary orientation on the standard triangle, which in turn induces a permutation $\sigma(p_1, p_2, p_3, U)$ on the set $\{p_1, p_2, p_3\}$. As any two standard triangles are homotopic this permutation is independent of the choice of the standard triangle. (For the orientation shown the permutation is $(p_1 p_2 p_3)$.)

5. Proof of the Theorem

Let M^2 be a compact oriented surface smoothly immersed in $\mathbf{R}^3$ so that all the shadow sets S_e are simply connected. Then by Corollary 2.2 M^2 is diffeomorphic to a sphere. Now assume, toward a contradiction, that M^2 is not the boundary of a convex set. Then by Propositions 3.1 and 3.3 there is an open set $V \subset S^2$ so that for every $u \in V$ the height function h_u is a Morse function that has at least two local maximums. For each $u \in V$ let $u^\perp \cap S^2$ be the great circle of S^2 orthogonal to u. Then as u varies over V the circles $u^\perp \cap S^2$ will cover a set F of positive measure in S^2. By Proposition 2.1 there is an $e \in F$ so that ∂S_e is a smooth curve in M^2. Now chose $u \in V$ so that $e \in u^\perp \cap S^2$. Then we have a pair u, e so that the height function

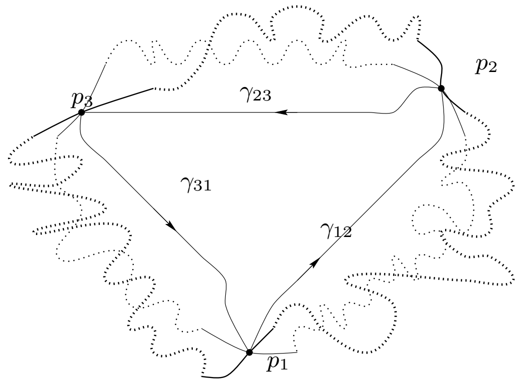

FIGURE 4. Pictured are two adjacent standard triples (p_1, p_2, p_3, U) and $(p_1, p_2, p_3, \widetilde{U})$ with a common standard triangle. As $\sigma(p_1, p_2, p_3, U)$ and $\sigma(p_1, p_2, p_3, \widetilde{U})$ are defined in terms of the standard triangle the equality $\sigma(p_1, p_2, p_3, U) = \sigma(p_1, p_2, p_3, \widetilde{U})$ holds.

h_u is a height function with two (or more) local maximums and an $e \perp u$ so that ∂S_e is a smooth curve.

By rotation we can assume that u is in the direction of the positive z axis and e in the direction of the positive x axis. The height function h_u is then just the restriction of the z coordinate to M^2. Let $e(\theta) := (\cos\theta, \sin\theta, 0)$. Then $e(0) = e$. Let p_1 and p_2 be two local maxima of h_u and let p_3 be a local minima. As $u \perp e(\theta)$ for all θ we see that $p_1, p_2, p_3 \in S_{e(\theta)}$ for all θ.

By Proposition 3.1 the Gauss curvature of M^2 is non-zero at all of the critical points p_1, p_2 and p_3 and thus the Gauss map g of M^2 is a local diffeomorphism near each of these points. Now as p_1 is a local maximum of $h_u = z\big|_{M^2}$ the tangent plane to M^2 at p_1 is defined by $z = c_1$ for some constant c_1. If $\delta > 0$ is sufficiently small there is a connected component of $M^2 \cap \{z = c_1 - \delta\}$ which is a simple closed curve on M^2 that bounds a disk containing p_1. As the Gauss curvature is not zero at p_1 the Gauss map g is a local diffeomorphism near p_1. By making δ small enough we can assume that $g\big|_{D_1} : D_1 \to S^2$ is a diffeomorphism onto its image $g[D_1]$. For each θ we have that the shadow boundary $\partial S_{e(\theta)}$ is the preimage of the great circle $e(\theta)^\perp S^2$ under the Gauss map. As $g\big|_{D_1}$ is a diffeomorphism this implies that $D_1 \cap \partial S_{e(\theta)}$ is a smooth curve through the point p_1. (This is pictured in Figure 5.) Letting $c_i = h_u(p_i)$, we can construct similar disks D_2 about p_1 (with boundary a component of $M^2 \cap \{z = c_2 - \delta\}$) and D_3 about p_3 (with boundary a component of $M^2 \cap \{z = c_3 + \delta\}$) so that $g\big|_{D_i} : D_i \to S^2$ is a diffeomorphism for $i = 1, 2, 3$.

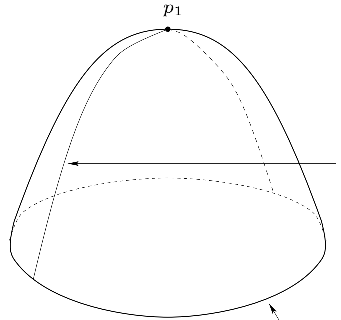

FIGURE 5. Near the local maxima p_1 a plane parallel to the tangent $T(M)_{p_1}$ and a little below p_1 will cut off a disk D_1 in M^2 so that the restriction of the Gauss map to D_1 is a diffeomorphism with $g[D_1] \subset S^2$. Any of the shadow boundaries $\partial S_{e(\theta)}$ will meet U_1 in a smooth curve.

As $D_i \cap \partial S_{e(\theta)}$ is smooth for all θ and we are assuming that each $S_{e(\theta)}$ is simple connected we see that $(p_1, p_2, p_3, S_{e(\theta)})$ is a nicely placed triple in the sense of Section 4.

Lemma 5.1. *For any θ_0 there is a $\epsilon > 0$ so that if $|\theta - \theta_0| < \epsilon$ then $(p_1, p_2, p_3, S_{e(\theta)})$ and $(p_1, p_2, p_3, S_{e(\theta_0)})$ are adjacent nicely placed triples.*

Proof. We choose a standard triangle $\gamma_{12}, \gamma_{23}, \gamma_{31}$ for $S_{e(\theta_0)}$ so that at all the corners p_1, p_2, p_3 the sides of the triangle form non-zero angles with ∂S_{θ_0}. In other terminology this is the same as requiring that all of the sides γ_{ij} are transverse to ∂S_{θ_0} (i.e. not tangent to ∂S_{θ_0}). (This is pictured in Figure 6.) By continuity if $|\theta_0 - \theta|$ is sufficiently small then $\partial S_{e(\theta)}$ will also be transverse to the sides γ_{ij}. Let $\mathcal{T} := \overline{\operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})} \cap S_{\theta_0}$. (Here $\overline{A}$ is the closure of A in M^2.) From the definition of a standard triangle we have

$$\mathcal{T} = \overline{\operatorname{Int}(\gamma_{12}, \gamma_{23}, \gamma_{31})} \setminus \{p_1, p_2, p_3\}.$$

This discussion implies that there is an $\epsilon_1 > 0$ so that

$$\mathcal{T} \cap (D_1 \cup D_2 \cup D_3) \subset S_\theta. \tag{5.1}$$

holds whenever $|\theta - \theta_0| < \epsilon_1$.

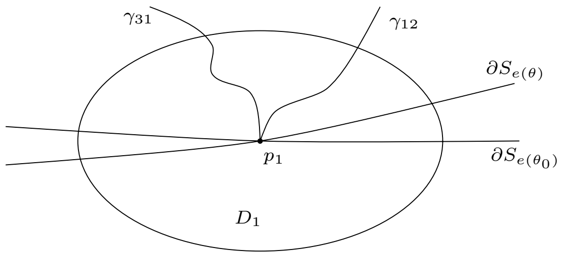

FIGURE 6. If $\gamma_{12}, \gamma_{23}, \gamma_{31}$ is a standard triangle for $(p_1, p_2, p_3, S_{e(\theta_0)})$ so that at the corners p_i the sides γjk make a positive angle with $\partial S_{e(\theta_0)}$ then the same will be true for $\partial S_{e(\theta)}$ with $|\theta - \theta_0|$ sufficiently small. Therefore $\mathcal{T} \cap D_1 \subset S_\theta$ for these θ.

The set $E := \mathcal{T} \setminus (D_1 \cup D_2 \cup D_3)$ is compact and $f_{e(\theta_0)}$ is positive on E. Thus by compactness and continuity there is a $\epsilon_2 > 0$ so that $f_{e(\theta)}$ is positive on E for $|\theta - \theta_0| < \epsilon_2$. But from the definition of S_θ in terms of $f_{e(\theta)}$ this implies

$$\mathcal{T} \setminus (D_1 \cup D_2 \cup D_3) \subset S_{e(\theta)} \tag{5.2}$$

for $|\theta - \theta_0| < \epsilon_2$. Let $\epsilon = \min\{\epsilon_1, \epsilon_2\}$. Then (5.1) and (5.2) imply that $\mathcal{T} \subset S_{e(\theta)}$ when $|\theta - \theta_0| < \epsilon$. But this is just the condition that is required to show that $S_{e(\theta)}$ is adjacent to $S_{e(\theta_0)}$. □

Recall from Section 4 that for each θ the nicely placed triple $(p_1, p_2, p_3, S_{e(\theta)})$ determines a permutation $\sigma(p_1, p_2, p_3, S_{e(\theta)})$ of the set $\{p_1, p_2, p_3\}$. By possibly renaming p_1 and p_2 (the two local maxima) we can assume that when $\theta = 0$ that $\sigma(p_1, p_2, p_3, S_{e(0)}) = (p_1 p_2 p_3)$. Lemmas 4.3 and 5.1 imply that $\theta \mapsto \sigma(p_1, p_2, p_3, S_{e(0)})$ is locally constant. As the real numbers are connected this implies that $\sigma(p_1, p_2, p_3, S_{e(\theta)}) = (p_1 p_2 p_3)$ for all θ. Recall that for $\theta = 0$ we have that $S_{e(0)}$ has a smooth boundary $\partial S_{e(0)}$. By Lemma 4.2 the permutation $\sigma(p_1, p_2, p_3, S_{e(0)})$ is just the cyclic permutation induced on $\{p_1, p_2, p_3\}$ by the boundary orientation of $\partial S_{e(0)}$. But $e(\pi) = -e(0)$ and $S_{e(\pi)} = M^2 \setminus \overline{S}_{e(0)}$. Therefore as sets $\partial S_{e(0)} = \partial S_{e(\pi)}$, but as the inward normal determined by $S_{e(0)}$ is the outward normal determined by $S_{e(\pi)}$ the boundary orientation $\partial S_{e(0)}$ determined by $S_{e(\pi)}$ is opposite of the one determined by $S_{e(0)}$. Therefore if $\sigma(p_1, p_2, p_3, S_{e(0)}) = (p_1 p_2 p_3)$ then $\sigma(p_1, p_2, p_3, S_{e(\pi)}) = (p_1 p_3 p_2)$ which contradicts that $\sigma(p_1, p_2, p_3, S_{e(\theta)}) = (p_1 p_2 p_3)$ for all θ. This completes the proof.

6. Tubes about Curves and an Example of a Genus One Surface with Connected Shadows

Let $c\colon S^1 \to \mathbf{R}^3$ be a smooth immersed curve. Let M^2 be the tube of radius r about c. Then for r sufficiently small M^2 is a smooth immersed torus in $\mathbf{R}^3$ (and if c is embedded M^2 will be embedded). Explicitly

$$M^2 := \{c(t) + ru : t \in S^1, u \in S^2,\ u \perp c'(t)\}.$$

The ***fiber*** of M^2 over t is $C_t := \{c(t) + ru : u \in S^2,\ u \perp c'(t)\}$ which is a circle. The Gauss map of M^2 is given by

$$\mathbf{n}(c(t) + ru) = u.$$

Therefore for any $e \in S^2$ and $t \in S^1$ the fiber C_t either is half inside (open

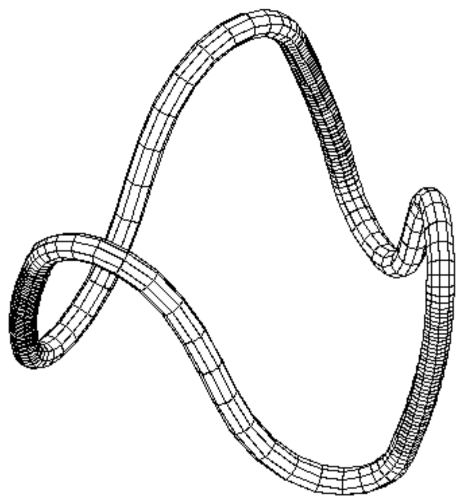

FIGURE 7. A tube about a curve that has no two parallel tangent lines. Such a tube has all its shadow sets connected.

half circle) and half outside (closed half circle) the shadow set S_e or C_t is entirely contained in the boundary of S_e. The latter only happens at points where $c'(t) \perp e$. Thus for each t the set $C_t \cap S_e$ is either empty (when $c'(t) \perp e$) or $C_t \cap S_e$ is an open half circle. If $[t_0, t_1]$ is an interval in S^1 so that for no point $t \in [t_0, t_1]$ does $c'(t) \perp e$ hold, then it is easily seen that that part of the shadow S_e over $[t_0, t_1]$ is connected. (That is $S_e \cap \bigcup_{t\in[t_0,t_1]} C_t$ is connected.) This implies that the number of connected components of S_e is at most the number of connected components of $S^1 \setminus \{t \in S^1 : c'(t) \perp e\}$. As at least two points of S^1 must be removed before it is disconnected, if some shadow S_e is not connected there must be two values of $t \in S^1$ with $c'(t) \perp e$.

Thus if the curve c has the property that its unit tangent curve $\mathbf{t}(t) := c'(t)/\|c'(t)\|$ is a simple closed curve on the sphere that does not contain any of its antipodal points (or what is the same thing c has no two parallel tangent lines), then for each $e \in S^2$ there is at most one $t \in S^1$ so that $c'(t) \perp e$. Then the tube M^2 will have all its shadow sets S_e connected. Figure 7 shows such a tube.

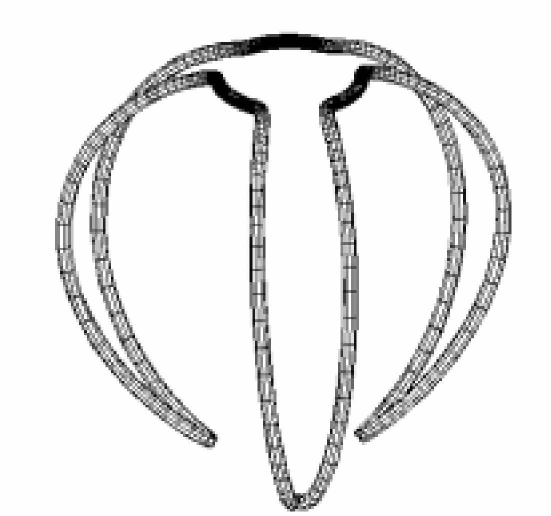

FIGURE 8. A small tube about the tangent image. This curve has the origin in the interior of its convex hull and also is disjoint from its antipodal image (refection through the origin).

As it is not obvious that a curve with these properties exists we now give a construction for one. Start with any simple closed curve $\mathbf{t}\colon S^1 \to S^2$ that does not contain any of its antipodal points and so that the origin is in the interior of the convex hull of $\mathbf{t}[S^1]$. Figure 8 shows such a curve. This curve was constructed along the following lines. Start with a circle on the sphere which is parallel to a great circle. For instance, take a circle of constant latitude in the northern hemisphere, such as the Arctic circle, then construct three tongs (legs) which go down from the Arctic circle along three longitudes which are 120 degrees apart. Send these tongs below the equator, but stop just short of the Antarctic circle (the antipodal image of the Arctic circle).

The curve in Figure 8 has a parameterization

$$\mathbf{t}(t) := \begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix} = \begin{bmatrix} \sin r(t) \cos t \\ \sin r(t) \sin t \\ \cos r(t) \end{bmatrix}$$

where

$$r(t) := \big(.1 + .78 \cos^{20}(3t/2)\big)\pi.$$

(This defines the functions $x(t)$, $y(t)$, and $z(t)$.) Then $\mathbf{t}$ is periodic of period 2π and thus can be viewed as defined on $S^1 = \mathbf{R}/2\pi\mathbf{Z}$. But if a curve on the sphere has the origin in the interior of its convex hull then then it is the unit tangent curve of a closed space curve (see Propostion 7.1 for the exact statement and its proof). If c is a curve that has $\mathbf{t}$ as its unit tangent curve then c will have no parallel tangents and therefore a small tube about c will give an example of an embedded genus one surface in $\mathbf{R}^3$ with all shadow sets connected.

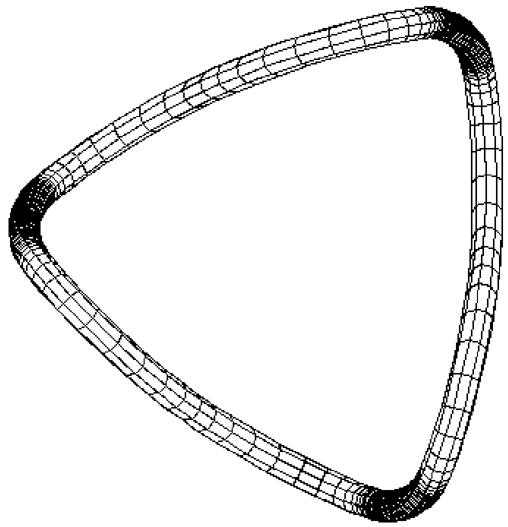

FIGURE 9. A top view of the tube of Figure 7 showing that its central curve lies on a cylinder over a convex plane curve.

The proof of Proposition 7.1 is contructive and we can use it to explicitly find the curve c. First let

$$v(t) := .1 + a \cos^{20}(3t/2)$$

where a is a positive number to be chosen shortly. Then for symmetry reasons (or direct calculation)

$$\int_0^{2\pi} v(t)x(t)\,dt = \int_0^{2\pi} v(t)y(t)\,dt = 0.$$

Now choose a so that

$$\int_0^{2\pi} v(t)z(t)\,dt = 0.$$

Then $a \approx .8824923325\ldots$ to enough accuracy to produce decent graphics. The integrals above and the periodicity of $\mathbf{t}$ imply that if

$$c(t) := \int_0^t v(\tau)\mathbf{t}(\tau)\,d\tau$$

then $c(t+2\pi) = c(t)$. Therefore $c\colon S^1 \to \mathbf{R}^3$ is a closed curve and $c'(t) = v(t)\mathbf{t}(t)$ and so $\mathbf{t}$ is the unit tangent curve of $\mathbf{t}$. Then, as $\mathbf{t}$ does not meet its antipodal curve, c will not have any two parallel tangent lines and gives us our required curve. Figure 7 is the tube of radius $r = .01$ about this curve.

We leave it as an exercise to show that the projection of c onto the x-y plane is a convex curve. This can be seen in Figure 9 which is the tube of Figure 7 viewed from high above the x-y plane with the direction of sight parallel to the z axis.

7. Appendix: A Sufficient Condition for a Spherical Curve to be the Tangent Mapping of a Space Curve

The following sufficient condition for a closed curve on a sphere to be the tangent map to a space curve is, in the proper circles, a well known folk theorem (I learned of the result and its proof from Mike Gage), however the only explicit reference I have seen is the original paper of Fenchel on the total curvature of space curves (*Uber Krummungund Windung geschlossener Raumkurven*, Math. Ann. **101** (1929) p. 238–252) where it is stated without proof and attributed to Löwner[1].

Proposition 7.1. *Let* $\mathbf{t}\colon S^1 \to S^2$ *be a smooth map so that the origin is in the interior of the convex hull of the image of* $\mathbf{t}$*. Then there is an immersed space curve* $c\colon S^1 \to \mathbf{R}^3$ *that has* $\mathbf{t}$ *as unit tangent map (that is* $\mathbf{t}(t) = c'(t)/\|c'(t)\|$*).*

Remark 7.2. It is all right if $\mathbf{t}$ is not injective, or if it is not an immersion.

Proof. We view S^1 as $\mathbf{R}/\mathbf{Z}$ so that $\mathbf{t}$ can be thought of as a map $\mathbf{t}\colon \mathbf{R} \to S^2$ with $\mathbf{t}(t+1) = \mathbf{t}(t)$. If we can find a positive function $v\colon \mathbf{R} \to (0,\infty)$ with $v(t+1) = v(t)$ and so that the "center of gravity" of the product $v\mathbf{t}$ is the origin, that is

$$\int_0^1 v(\tau)\mathbf{t}(\tau)\,d\tau = 0. \tag{7.1}$$

[1]Fenchel attributes the result to Löwner and gives as reference Pólya and Szegő *Aufgaben und Lehrsätze aus der Analysis. Band II: Funktionentheorie, Nullstellen, Polynome Determinanten, Zahlentheorie*, Springer, Berlin, (1925) S. 165 und 391 Aufgabe, 13. But in fact Pólya and Szegő only show that the origin is in the convex hull of the image of a tangent mapping and do not state or prove the converse. They credit this result to Löwner but without explicit reference.

Then periodicity implies that for all t

$$\int_t^{t+1} v(\tau)\mathbf{t}(\tau)\,d\tau = 0. \tag{7.2}$$

Set

$$c(t) := \int_0^t v(\tau)\mathbf{t}(\tau)\,d\tau.$$

The integral condition (7.2) implies $c(t+1) = c(t)$ so c is a closed curve $c\colon S^1 \to \mathbf{R}^3$. Also $c'(t) = v(t)\mathbf{t}(t)$ and as v is positive this implies that c is an immersion. Finally $c'(t) = v(t)\mathbf{t}(t)$ clearly implies that $\mathbf{t}$ is the unit tangent map to c.

We now show the existence of the function v so that (7.1) holds. As the origin is in the interior of the convex hull of c there are distinct $t_0, t_1, t_2, t_3 \in [0,1)$ and $\alpha_0, \alpha_1, \alpha_2, \alpha_3 \in (0,1)$ with each α_i positive, $\sum_{i=0}^3 \alpha_i = 1$ and

$$\sum_{i=0}^3 \alpha_i \mathbf{t}(t_i) = 0.$$

By continuity there is a $\delta > 0$ so that if $P_0, P_1, P_2, P_3 \in \mathbf{R}^3$ then

$$\|\mathbf{t}(t_i) - P_i\| < \delta \text{ for } i = 1, \ldots, 4 \implies 0 \in \text{convex hull of } P_0, \ldots, P_3. \tag{7.3}$$

Now for $i = 0, \ldots, 4$ there is a smooth positive C^∞ function v_i on $\mathbf{R}$ with $v_i(t+1) = v_i(t)$, $\int_0^1 v_i(t)\,dt = 1$ and that approximates the point mass at t_i well enough that

$$\left\| \mathbf{t}(t_i) - \int_0^1 v_i(t)\mathbf{t}(t)\,dt \right\| < \delta.$$

Therefore the implication (7.3) yields $\beta_0, \ldots, \beta_3 \geq 0$ with $\sum_{i=0}^3 \beta_i = 1$ and

$$0 = \sum_{i=0}^3 \beta_i \int_0^1 v_i(t)\mathbf{t}(t)\,dt = \int_0^1 \sum_{i=0}^3 \beta_i v_i(t)\mathbf{t}(t)\,dt = \int_0^1 v(t)\mathbf{t}(t)\,dt$$

where $v(t) = \sum_{i=0}^3 \beta_i v_i(t)$. This gives us the v so that the desired relation (7.1) holds and completes the proof. ☐

Acknowledgment: I would like to thank Sara Kasiri for suggesting some improvements to these notes.

References

1. S. S. Chern and R. L. Lashof, *On the total curvature of immersed manifolds*, Amer. J. Math. **79** (1957), 306–318.
2. ———, *On the total curvature of immersed manifolds. II*, Michigan Math. J. **5** (1958), 5–12.
3. J. Milnor, *Morse theory*, Princeton University Press, Princeton, N.J., 1963, Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51.