# MULTIPLE INTERSECTIONS ON NEGATIVELY CURVED SURFACES

CHARLES EPSTEIN, TROELS JORGENSEN & CURT McMULLEN

## 1. Introduction

Let $X$ be a compact surface with a Riemannian metric of negative curvature. An $n$-tuple point on $X$ is a point through which a single geodesic passes at least $n$ times in different directions. The geodesic is not required to be closed. Our purposes are to describe the loci of triple and quadruple points and to show that the *quadruple points are dense* on $X$ and the tangents at quadruple points can be made near to four given directions. The proof of this last fact is based on the theory of Anosov flows [1].

There is an elementary, intuitive way of studying geodesic triple points. We present it in §2 and §3 for the case of constant curvature. The quadruple points for variable curvature are studied in §4.

## 2. Hyperbolic surfaces

Consider a surface $X$ of constant curvature minus one. Its universal cover is the hyperbolic plane $\mathbf{H}^2$.

Three different lines in $\mathbf{H}^2$ are lifts of the same geodesic on $X$ if and only if there exist deck transformations $A$ and $B$, each mapping one of the pair of lines onto the third. Such lines will pass through a common point $\psi$, if and only if their projection on $X$ has a triple point at the image of $\psi$; this corresponds to $A\psi$, $B\psi$, and $\psi$ being colinear.

There is a single degenerate configuration which occurs if $\psi$ lies on the axis of $A$, $B$ or $A^{-1}B$. In this case $\psi$ projects to a single intersection and the line in $\mathbf{H}^2$ projects to a closed geodesic.

It is therefore natural, given two hyperbolic isometries $A$ and $B$, say noncommuting and without fixed points in $\mathbf{H}^2$, to study the locus $Z_{A,B}$ of points $\psi$ for which $A(\psi)$, $B(\psi)$, and $\psi$ are colinear. Upon introducing coordinates, $Z_{A,B}$ can be described as the solution set to a cubic

equation. In particular, $Z_{A,B}$ is the union of certain smooth curves. Due to symmetry, it is immediate that

$$Z_{A,B} = Z_{A^{-1},A^{-1}B} = Z_{B^{-1}A,B^{-1}}.$$

The aforementioned degeneracy occurs where $Z_{A,B}$ intersects an axis of $A$, $B$ or $A^{-1}B$. Whether these intersections exist depends on the elements $A$ and $B$; they will in any case consist of a finite set of points.

It is not surprising to find that $Z_{A,B}$ is made up of three disjoint components:

$$Z_{A,B} = C_{A,B} \cup C_{A^{-1},A^{-1}B} \cup C_{B^{-1}A,B^{-1}},$$

each of which is a smooth curve connecting the attractive fixed points of the associated transformations. To see this, let us first consider two "large" hyperbolic motions, $A$ and $B$, with distinct fixed points on the circle at infinity. For any point $\psi$, which is not too "close" to either of the repulsive fixed points, the images $A(\psi)$ and $B(\psi)$ will be "close" to the respective attractive fixed points. Thus, if such a point $\psi$ belongs to $Z_{A,B}$, it must be close to the line between the attractive fixed points of $A$ and $B$. Thus, when $A$, $B$, and $A^{-1}B$ are large, we can easily recognize the three components of $Z_{A,B}$. To verify the description of $Z_{A,B}$ in general, we continuously deform $A$ and $B$ and use a lemma, proved in §3, which shows that the three components cannot become tangent nor (self-) intersecting in $\mathbf{H}^2$.

The group of deck transformations for a hyperbolic surface $X$ is a Fuchsian group $\Gamma$. The pairs of attractive fixed points of elements in $\Gamma$ are dense among all pairs of limit points for $\Gamma$. Also, by taking powers, each pair of attractive fixed points belongs to a pair of arbitrarily large transformations. The above reasoning, therefore, shows that the curves $C_{A,B}$ of triple points for $\Gamma$ are dense in the subset of $\mathbf{H}^2$, which is the hyperbolic convex hull of the limit set of $\Gamma$. Thus, on $X$ geodesic triple points occur densely along smooth curves in the projection of this convex hull. In particular, if $X$ is compact, or more generally, if the limit set of $\Gamma$ is the entire circle at infinity, then the triple points are dense on all of $X$.

By considering points of intersection $\psi \in C_{A,B} \cap C_{C,A^{-1}}$ for suitably chosen transformations $A$, $B$, and $C$, one can show that quadruple points are dense. Choose $C$ so that topological considerations force $Z_{A,B}$ to intersect $Z_{C,A^{-1}}$; for example, if the axes of $A$ and $B$ are disjoint, as in Figure 2, one chooses the element $C$ so that the attractive fixed point of $C$ is separated from the fixed points of $A$ by the fixed points of $B$.

It is then apparent that the curve $C_{C,A^{-1}}$, which joins the repulsive fixed point of $A$ to the attractive fixed point of $C$, must cross $C_{A,B}$, called the point of intersection $\psi$. The points $\psi$, $A\psi$, $B\psi$, and $AC\psi$ will be colinear and, except when $\psi$ lies on an axis, will project to a quadruple point. By using the density in $S^1$ of the attractive fixed points of a discrete cofinite group and the fact that when two elements, $A$ and $B$, have very large translation lengths, $C_{A,B}$ is very close to the line connecting their attractive fixed points, one easily proves the density of quadruple points on the quotient surface.

In the constant curvature case we have shown that the locus of triple points is a union of real analytic varieties and the locus of quadruple points lies in the intersections of components of these varieties. In the smooth category one expects a similar description: triple points form continua while quadruple points are isolated though dense.

## 3. Computations

In order to make the above more precise we first remark that the set $Z_{A,B}$ is the zero locus of a smooth function defined in $\mathbf{H}^2$ by:

$$\Delta(X) = [\text{signed area of the triangle with vertices } (X, AX, BX)].$$

The important properties of $Z_{A,B}$ will follow from the fact that $d\Delta \neq 0$ on $Z_{A,B}$ and its behavior near $\partial_\infty \mathbf{H}^2$. Since the fixed points of $A$, $B$ and $A^{-1}B$ are pairwise disjoint, it follows easily $\overline{Z}_{A,B} \cap \partial_\infty \mathbf{H}^2$ consists precisely of the fixed points $A$, $B$ and $A^{-1}B$. A moment's consideration shows that $\Delta$ extends continuously to $\partial_\infty \mathbf{H}^2 / \{\text{fixed points of } A, B \text{ and } A^{-1}B\}$ where it equals either $\pi$ or $-\pi$.

**The Main Lemma.** *Let $A$ and $B$ be two distinct hyperbolic elements in $\mathrm{Aut}^+(\mathbf{H}^2)$. Then the following hold:*

(1) *$Z_{A,B}$ is a smoothly embedded submanifold of $\mathbf{H}^2$.*

(2) *The sign of $\Delta(X)$ changes as $X$ crosses $Z_{A,B}$.*

(3) *If $A^{-1}B$ is also hyperbolic and the fixed points of $A$, $B$ and $A^{-1}B$ are pairwise disjoint, then $Z_{A,B}$ is exponentially close to the pencils of geodesics emanating from the fixed points of $A$, $B$ and $A^{-1}B$.*

*Proof.* As remarked above, we deduce these properties from the fact that $Z_{A,B}$ is noncritical for $\Delta$. Instead of studying $\Delta$, we will introduce a simpler function and work in the Lorentz model of $\mathbf{H}^2$. Briefly, this is

defined by putting an inner product of signature $(2, 1)$ on $\mathbf{R}^3$:

$$\langle X, Y \rangle = X_1 Y_1 + X_2 Y_2 - X_3 Y_3.$$

Then $\mathbf{H}^2$ is isometric to the submanifold:

$$H^+ = \{\langle X, X \rangle = -1; X_3 > 0\},$$

with the induced inner product. Geodesics are the intersections of planes through $(0, 0, 0)$ with $H^+$ and the group $\mathrm{Aut}^+(\mathbf{H}^2)$ is identified with the subgroup, $\mathrm{SO}^+(2, 1)$ of the orthogonal group of $\langle \ , \ \rangle$ which preserves the condition: $X_3 > 0$. The action of $\mathrm{SO}^+(2, 1)$ is the standard linear action of $\mathrm{Gl}(3)$ on $\mathbf{R}^3$. Because geodesics in $\mathbf{H}^2$ are given by intersections of planes through $(0, 0, 0)$ with $H^+$, one easily sees that:

$$Z_{A,B} = \{X \in H^+: \det(X, AX, BX) = 0\};$$

we define $F(X) = \det(X, AX, BX)$. The signed area is a positive multiple of $F(X)$. Thus $Z_{A,B}$ is the intersection of the zero set of a homogeneous cubic with $H^+$. We will let $Z = \{X \in \mathbf{R}^3: F(X) = 0\}$. To prove statements (1) and (2), it suffices to show that $dF(X)(Y) \neq 0$ for $X \in Z_{A,B}$ and some $Y \in T_X H^+$. If $X \in Z$, then there are real numbers $a$ and $b$ such that

$$TX = (I + aA + bB)X = 0.$$

From this it follows easily that for $X \in Z$,

$$\begin{aligned} dF(X)(Y) &= \det(TY, AX, BX) \\ &= Y \cdot {}^t T(AX \times BX). \end{aligned}$$

Hence $dF(X) = 0$ if and only if ${}^t T(AX \times BX) = 0$. Suppose that $X \in Z_{A,B}$. Then we can find a transformation $C \in \mathrm{SO}^+(2, 1)$, such that

$$CX = (0, 0, 1),$$

$$CAC^{-1} = \begin{pmatrix} \cos\theta\,\mathrm{ch}\,\lambda & -\sin\theta\,\mathrm{ch}\,\lambda & \mathrm{sh}\,\lambda \\ \sin\theta & \cos\theta & 0 \\ \cos\theta\,\mathrm{sh}\,\lambda & -\sin\theta\,\mathrm{sh}\,\lambda & \mathrm{ch}\,\lambda \end{pmatrix},$$

$$CBC^{-1} = \begin{pmatrix} \cos\phi\,\mathrm{ch}(-\mu) & -\sin\phi\,\mathrm{ch}(-\mu) & \mathrm{sh}(-\mu) \\ \sin\theta & \cos\phi & 0 \\ \cos\phi\,\mathrm{sh}(-\mu) & -\sin\phi\,\mathrm{sh}(-\mu) & \mathrm{ch}(-\mu) \end{pmatrix}.$$

In the sequel we will simply write $X$ for $CX$, $A$ for $CAC^{-1}$, and $B$ for $CBC^{-1}$ as this should cause no confusion.

In light of the fact that $A \neq B$, we can assume without loss of generality that the ordering of points along the line is $AX$, $X$, $BX$; were this not

the case we could simply replace the pair $(A, B)$ with either $(A^{-1}, A^{-1}B)$ or $(B^{-1}, B^{-1}A)$ and proceed as below. This implies that we can assume that $\lambda$ and $\mu$ are both positive. With these normalizations we have:

$$T = \left( I - \frac{\operatorname{sh}\mu A}{\operatorname{sh}(\lambda + \mu)} - \frac{\operatorname{sh}\lambda B}{\operatorname{sh}(\lambda + \mu)} \right)$$

and

(4)
$$AX \times BX = (0, -\operatorname{sh}(\lambda + \mu), 0).$$

It follows from (4) above that ${}^tT(AX \times BX) = 0$, if and only if both $T_{21}$ and $T_{22}$ are zero. An elementary calculation gives:

$$T_{21} = -(\operatorname{sh}\mu \sin\theta + \operatorname{sh}\lambda \sin\phi) \Big/ \operatorname{sh}(\lambda + \mu),$$

$$T_{22} = 1 - (\operatorname{sh}\mu \cos\theta + \operatorname{sh}\lambda \cos\phi) / \operatorname{sh}(\lambda + \mu).$$

In order for $T_{21} = 0$, it is necessary that $(\sin\theta, \sin\phi) = t(\operatorname{sh}\lambda, -\operatorname{sh}\mu)$ for a $t$ with $|t| \leq \min(\frac{1}{\operatorname{sh}\lambda}, \frac{1}{\operatorname{sh}\mu})$. Therefore

$$(\cos\theta, \cos\phi) = \left( \pm\sqrt{1 - t^2\operatorname{sh}^2\lambda}, \pm\sqrt{1 - t^2\operatorname{sh}^2\mu} \right).$$

Using this in the formula for $T_{22}$, we see that if $T_{22} = 0$ for any $t$ in the allowable range, then it will vanish for some $t$ with both $\cos\theta$ and $\cos\phi$ positive, with these choices: $T_{22} = 1 - g(t)$, where

$$g(t) = \left( \operatorname{sh}\mu\sqrt{1 - t^2\operatorname{sh}^2\lambda} + \operatorname{sh}\lambda\sqrt{1 - t^2\operatorname{sh}^2\mu} \right) \Big/ \operatorname{sh}(\lambda + \mu).$$

This function assumes its maximum value in the allowable range at $t = 0$, where

$$g(0) = (\operatorname{sh}\mu + \operatorname{sh}\lambda) / \operatorname{sh}(\mu + \lambda) < 1.$$

Thus if $T_{21} = 0$, then $T_{22} \neq 0$. This proves that for $X \in Z_{A,B}$, there is $Y \in T_X R^3$ such that $dF(X)(Y) \neq 0$; since $dF(X)(X) = 0$ and $X$ is transverse to $T_X H^+$, it follows that there is a $Y \in T_X H^+$ such that $dF(X)(Y) \neq 0$. From this, (1) and (2) follow immediately.

To complete the proof we need to examine the behavior of $Z_{A,B}$ near to $\partial_\infty H^2$. The model in which the asymptotic boundary is most smoothly attached is the Klein model. This is obtained by "stereographically" projecting $H^+$ into the plane $\{X_3 = 1\}$: a line through $(0, 0, 0)$ which lies inside $\langle X, X \rangle = 0$ will intersect one point on $H^+$ and one point on $\{X_3 = 1\}$; this defines a projection of $H^+$ onto

$$K = \{(X_1, X_2, X_3): X_1^2 + X_2^2 < 1; X_3 = 1\},$$
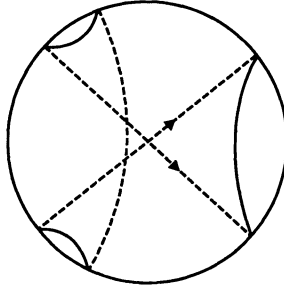
which is called the Klein model of hyperbolic space.
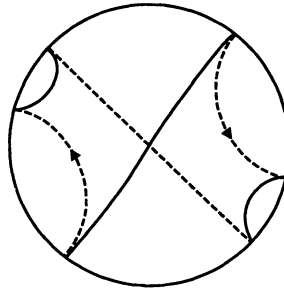
FIGURE 1



FIGURE 2

Hyperbolic geodesics in the Klein model are segments of straight lines in the plane $\{X_3 = 1\}$. From this observation and the fact that the distance between hyperbolic geodesics with a common endpoint on $\partial_\infty \mathbf{H}^2$ tends exponentially to zero, it follows that a smooth curve in $\overline{K}$, which intersects $\partial K$ transversely at a point $p$, is exponentially close in the hyperbolic metric, to the pencil of geodesics emanating from $p$.

Since $F(X)$ is homogeneous, the image of $Z_{A, B}$ under the projection to $K$ is simply $Z \cap K$. To complete the proof of the lemma, all we need to show is that $dF(X)(W) \neq 0$ for $X \in Z \cap \partial K$ and $W$ a tangent vector to $\partial K$ at $X$. It is convenient to use a different normalization from that used above, we will assume that $X$ and $A$ are normalized so that:

$$X^{\pm} = (\pm 1, 0, 0),$$
$$A = \begin{pmatrix} \operatorname{ch}\lambda & 0 & \operatorname{sh}\lambda \\ 0 & 1 & 0 \\ \operatorname{sh}\lambda & 0 & \operatorname{ch}\lambda \end{pmatrix}, \quad \lambda \neq 0.$$

We see that

$$dF(X^{\pm})(Y) = \det(T_{\pm}Y, AX^{\pm}, BX^{\pm}),$$

where $T_{\pm} = (I - e^{\mp\lambda}A)Y$. With the above normalizations, $W = (0, x, 0)$, and

$$W \cdot {}^{t}T_{\pm}[AX^{\pm} \times BX^{\pm}] = \begin{cases} -x(e^{\lambda} - 1)[(B_{11} + B_{13}) - (B_{31} + B_{33})] = Z^{+}, \\ x(e^{-\lambda} - 1)[(B_{13} - B_{11}) - (B_{33} - B_{31})] = Z^{-}. \end{cases}$$

Note that if $Z^{\pm} = 0$, then $BX^{\pm} = \mu X^{\pm}$ for some $\mu$. This follows because $\langle BX^{\pm}, BX^{\pm} \rangle = \langle X^{\pm}, X^{\pm} \rangle = 0$; if $Z^{\pm} = 0$, then $\langle BX^{\pm}, BX^{\pm} \rangle = (B_{21} \pm B_{23})^2 = 0$. By assumption, the fixed points of $A$ and $B$ are disjoint; hence, $dF(X^{\pm})(W) \neq 0$. This completes the proof of the main lemma.   q.e.d.

Figures 1 and 2 show $Z_{A,B}$ in the Poincaré model; the axes of $A$, $B$ and $A^{-1}B$ are drawn as dashed lines, and $Z_{A,B}$ is drawn as a solid line. In Figure 1 the axes intersect; in Figure 2 the axes are disjoint.

## 4. Variable Curvature

In order to prove the density theorem, we first make a few definitions and recall some basic facts.

An *n-cross* consists of a point on $X$ and $n$ unit tangent vectors at that point. When an *n*-tuple point occurs (a point through which a geodesic passes $n$ times), the point and the tangents to the geodesic determine an *n*-cross.

As in the hyperbolic cases, the universal cover $\hat{X}$ has a *sphere at infinity* $S^1$, which can be constructed from the visual sphere at any point.

The geodesic flow is mixing relative to the invariant Liouville measure on the unit tangent bundle [1].

A *pencil of geodesics* $G(s)$ is determined by a smooth immersed path $T(s)$ in the unit tangent bundle, transverse to the geodesic flow; $G(s)$ denotes the geodesic through $T(s)$. The geodesics in a pencil have the following "sweeping" property: The positive (or if not, then the negative) ends of the geodesics spread apart exponentially fast, so that:

(1) as $s$ varies, the speed of motion of $G(s)$ normal to its direction as a geodesic on $X$ tends to infinity as we travel toward the positive (negative) end of $G(s)$; and

(2) for any open set $U$ in the unit tangent bundle to $X$, there are arbitrarily small values of $s$ for which the far positive (negative) end of $G(s)$ enters $U$. These conditions fail for the positive ends of $G(s)$ only

if these ends are focused at a single point on the sphere at infinity $S^1$; in that case, they hold for the negative ends of $G(s)$.

Property (2) is an easy consequence of mixing. Consider a lift of $G(s)$ to $\widehat{X}$, and let $I$ denote the interval of $S^1$ traced by the positive endpoint of $G(s)$ as $s$ ranges in $(-\varepsilon, \varepsilon)$. The lifted pencil corresponds to a smooth path $T(s)$ in the unit tangent bundle to $\widehat{X}$. Let $W$ denote an open neighborhood of $T(0)$ chosen small enough that on $\widehat{X}$, the geodesics through $W$ tend towards a subinterval of $I$ and are therefore eventually contained in the region swept out by $G(s)$. By mixing, under the geodesic flow $W$ meets lifts of $U$ infinitely often. But geodesics through $W$ become nearly parallel to those of $G(s)$ as one tends towards $I$ (looking backward, the visual distance between $W$ and $T(-\varepsilon, \varepsilon)$ tends to 0). Therefore, the far positive end of some $G(s)$ also enters $U$.

**Proposition 1.** *The double points of almost any geodesic are dense in the space of 2-crosses.*

*Proof.* By mixing, almost any geodesic is dense in the unit tangent bundle.

**Proposition 2.** *The triple points are dense in the space of 3-crosses.*

*Proof.* Take a pencil of geodesics $G(s)$ such that $G(0)$ has a double point at a given transverse 2-cross, and the positive ends are sweeping. By transversality, the double point moves by a differentiable function $b(s)$ for $s$ small. For large positive time, the geodesic $G(s)$ sweeps by the original double point with high normal velocity (much higher than $db/ds$) and nearly parallel to an arbitrary third unit vector. Hence there is a triple point near a given 3-cross (Figure 3).

**Theorem 3.** *The quadruple points are dense in the space of 4-crosses.*

*Proof.* Consider in the preceding construction the function $f(s) = \mathrm{dist}(b(s), G(s))$, where the distance is measured for the particular branch of $G(s)$ sweeping by nearly parallel to the third vector. Since $G(s)$ is moving rapidly, the derivative $df/ds$ is nonzero at the small value of $s$ which achieves the triple point. Let $T_{\mathrm{triple}}$ denote the unit tangent vector on the path $T(s)$ corresponding to this value of $s$.

Let $U$ be a two-dimensional submanifold of the unit tangent bundle transverse to the geodesic flow and containing $T_{\mathrm{triple}}$. The functions $b(s)$ and $f(s)$ can be defined locally on a neighborhood of $T_{\mathrm{triple}}$ in $U$ (since the double point is assumed transverse). By the above remarks, $f$ gives an immersion of $U$ into the real line, covering 0; by the implicit function theorem there is a smooth path $T(t)$ in $U$ such that $T(0) = T_{\mathrm{triple}}$ and $f(T(t)) = 0$, i.e., such that the corresponding geodesic $G(t)$ has a triple point at $b(t)$.
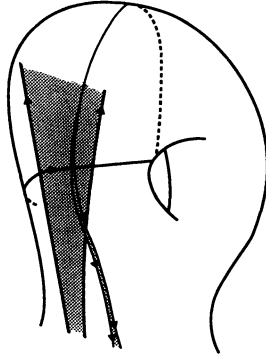
FIGURE 3

Now reapply the proof of Proposition 2. We may assume the positive ends of the geodesics $G(t)$ come sweeping by the original triple point for $G(0)$, nearly parallel to a fourth unit tangent vector and moving much faster than $b(t)$. (If the positive ends happen to be focused, we use the negative ends.) In any case the resulting collision produces a quadruple point approximating a given 4-cross.

# References

[1] D. V. Anosov, *Geodesic flows on closed Riemann manifolds with negative curvature*, Trudy Mat. Inst. Steklov **90** (1967). (Russian)

UNIVERSITY OF PENNSYLVANIA
COLUMBIA UNIVERSITY
PRINCETON UNIVERSITY