COLLOQUIUM MATHEMATICUM

VOL. XLIV 1981 FASC. 2

CUT LOCI OF CLOSED SURFACES WITHOUT CONJUGATE POINTS

BY

DAVID D. BLEECKER (HONOLULU)

1. Introduction. In the 1930's a study of cut loci of closed surfaces was made by Myers [4], [5]. Under additional assumption of non-positive curvature or, more generally, the non-existence of conjugate points, further results are possible, and these are summarized in Section 3 of the present work. The lemmas in Section 4 are of independent interest, and in many cases are stronger than what is needed to prove the theorems of Section 3.

2. Conventions and definitions. Unless otherwise stated, M will denote a compact, connected, C^∞ Riemannian 2-manifold without conjugate points (i.e. the exponential map at each point is non-singular everywhere). For $X \in M_p$, the tangent space of M at p, such that $\|X\| = 1$, define

$$c(X) = \inf\{a \in \mathbf{R}\colon\ g(t) = \exp_p tX,\ 0 \leqslant t \leqslant a, \text{ is not length-minimizing}\}.$$

The *cut locus* of M at p, denoted by L_p, is defined by

$$L_p = \{c(X)X\colon\ X \in M_p,\ \|X\| = 1\}.$$

Two points of L_p are *equivalent* if they are identified by $\exp_p$. $[X]$ will denote the equivalence class of $X \in L_p$, and $\# [X]$ the cardinality of $[X]$. A *vertex* of L_p will be a point $X \in L_p$ with $\# [X] \geqslant 3$. For convenience, $\# [X]$ will be called the *order* of X.

3. Main results.

THEOREM A. *If M is orientable and of genus G, then, for each $p \in M$, L_p may not have more than $12G-6$ vertices nor less than $4G$ vertices. For each $G \geqslant 1$ these bounds are achieved by some M of genus G at some $p \in M$.*

In fact, if m and n are positive integers with $mn = 4G-2$, we can find some M of genus G and $p \in M$ such that L_p has $(m+2)n$ vertices, all of order $m+2$.

THEOREM A′. *If M is non-orientable and G is the genus of the orientable 2-fold cover of M, say $\tilde{M}$, then, for $p \in M$, L_p has not more than $6G$ vertices* *nor less than* $2G+2$ *vertices. If* m *and* n *are positive integers with* $mn = 2G$, *we can find some* M *with* $\bar{M}$ *of genus* G *and some* $p \in M$ *such that* L_p *has* $(m+2)n$ *vertices, all of order* $m+2$.

THEOREM B. *Let* $f\colon M \to Z$ *be defined by the number* $f(p)$ *of vertices of* L_p. *Then, for each* $p \in M$, f *has a local minimum at* p. *In particular,* f *achieves its maximum on an open set. If* $X(M) \neq 0$, *then* f *is not constant.*

COROLLARY. *If* $X(M) \neq 0$, *then there are two points of* M *with* 4 *or more length-minimizing geodesics joining them. (If* $X(M) = 0$, *there are counterexamples.) For each compact surface* S *there is an* M *homeomorphic to* S *and there is an open set of* M, *say* U, *such that if* $p \in U$, *then there is no point of* M *joined to* p *by* 4 *or more length-minimizing geodesics.*

4. Proofs of lemmas and theorems.

LEMMA 1. *Let* (M, g) *be a complete Riemannian manifold of dimension* n *and let* $\pi\colon TM \to M$ *be the tangent bundle. Define* $P\colon TM \to M \times M$ *by* $P(X) = (\pi(X), \exp(X))$. *Then* P *is a Riemannian covering projection where the metric on* $M \times M$ *is* $g \times g$ *and that on* TM *is* $P^*(g \times g)$.

Proof. Since M has no conjugate points, $\exp_*$ is of rank n everywhere, and π_* under any conditions has rank n everywhere. Thus $\pi_* \times \exp_*$ has rank $2n$. That P_* is of rank $2n$ follows from the commutative diagram

$$\begin{array}{ccc} T^2M & \xrightarrow{\quad P_* \quad} & T(M \times M) \\ & {\scriptstyle \pi_* \times \exp_*} \searrow \qquad \swarrow {\scriptstyle \pi_{1*} \times \pi_{2*}} & \\ & TM \times TM & \end{array}$$

where π_1 and π_2 are the projections onto the first and the second factor of $M \times M$, respectively. Thus P is an isometric immersion. It remains to show that TM (with the prescribed metric) is complete ([3], p. 176). First note that the geodesics of $M \times M$ are precisely those curves of the form $\gamma_1 \times \gamma_2$, where γ_1 and γ_2 are geodesics in M, as is well known. Thus a curve of the form $\beta(s) = sX$, where $X \in TM$, is a geodesic in TM, since $P \circ \beta$ is a geodesic and P is a local isometry. Let $\gamma\colon [0, b) \to TM$ be any unit speed curve. We will prove that

$$(*) \qquad \frac{d\,\|\gamma(t)\|}{dt} \leqslant \sqrt{2},$$

where $\|\gamma(t)\|$ is the g-norm of the vector $\gamma(t)$. This will insure completeness since it implies that $\|\gamma(t)\| \leqslant \sqrt{2}b + \|\gamma(0)\|$, a fact easily verified using the mean value theorem. Hence for all $t \in [0, b)$ we have

$$\gamma(t) \in \{X \in TM\colon \|X\| \leqslant r_0,\ \pi(X) \in \overline{\operatorname{im} \pi \circ \gamma}\},$$

where $r_0 = \sqrt{2}b + \|\gamma(0)\|$. Thus any curve of finite length will lie in a compact set (i.e., the one in the preceding sentence). This clearly implies geodesic completeness. That (*) follows by a long and not straightforward calculation will be exhibited.

Let $a\colon [0, b) \times [0, 1] \to TM$ be given by $a(t, s) = s[\gamma(t)]$. We denote $a_*(\partial/\partial t)$ and $a_*(\partial/\partial s)$ by $\partial a/\partial t$ and $\partial a/\partial s$, respectively. Then

$$2\|\gamma(t)\|\frac{d\|\gamma(t)\|}{dt} = \frac{d\|\gamma(t)\|^2}{dt} = \frac{d}{dt}\int_0^1 \|\gamma(t)\|^2 ds = \frac{d}{dt}\int_0^1 \left\langle \frac{\partial a}{\partial s}, \frac{\partial a}{\partial s} \right\rangle ds,$$

since the curve $\beta_t(s) = a(t, s)$ has speed $\|\gamma(t)\|$ relative to the prescribed metric on TM and this metric restricted to a tangent space of M is just the pull-back of g by the exponential map on that tangent space. Now

$$\frac{d}{dt}\int_0^1 \left\langle \frac{\partial a}{\partial s}, \frac{\partial a}{\partial s} \right\rangle ds = \frac{d}{dt}\int_0^1 \left(\left\langle \pi_* \frac{\partial a}{\partial s}, \pi_* \frac{\partial a}{\partial s} \right\rangle + \left\langle \exp_* \frac{\partial a}{\partial s}, \exp_* \frac{\partial a}{\partial s} \right\rangle \right) ds$$

by the definition of the metric on TM. Let D/dt and D/ds denote covariant differentiation along the curves $s = \text{const}$ and $t = \text{const}$, respectively. Then the above is equal to

$$2\int_0^1 \left\langle \frac{D}{dt} \exp_* \frac{\partial a}{\partial s}, \exp_* \frac{\partial a}{\partial s} \right\rangle ds$$

(note that $\pi_*(\partial a/\partial s) = 0$) which is

$$2\int_0^1 \left\langle \frac{D}{ds} \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle ds,$$

since the Riemannian connection has 0-torsion. Now, since the metric is parallel, this is equal to

$$2\int_0^1 \frac{d}{ds}\left(\left\langle \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle - \left\langle \exp_* \frac{\partial a}{\partial t}, \frac{D}{ds} \exp_* \frac{\partial a}{\partial s} \right\rangle \right) ds.$$

We have

$$\frac{D}{ds} \exp_* \frac{\partial a}{\partial s} = 0$$

since $\exp \circ \beta_t$ is a geodesic, where β_t is defined as above. Thus we obtain

$$\begin{aligned} 2\int_0^1 \frac{d}{ds} \left\langle \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle ds & \\ = 2\left[\left\langle \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle (t, 1) - \left\langle \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle (t, 0) \right] & \\ = 2\left[\left\langle \exp_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle (t, 1) - \left\langle \pi_* \frac{\partial a}{\partial t}, \exp_* \frac{\partial a}{\partial s} \right\rangle (t, 0) \right], & \end{aligned}$$

since

$$\exp_* \frac{\partial a}{\partial t} = \pi_* \frac{\partial a}{\partial t} \quad \text{for } s = 0.$$

Now, noting that

$$\left\| \exp_* \frac{\partial a}{\partial s} \right\| = \|\gamma(t)\|$$

and using the Cauchy-Schwartz inequality, we infer that the expression above is less than or equal to

$$2\|\gamma(t)\| \left[\left\| \exp_* \frac{\partial a}{\partial t} \right\| + \left\| \pi_* \frac{\partial a}{\partial t} \right\| \right] (t, 1).$$

Here we also note that $\pi_* (\partial a / \partial t)$ is independent of s. Thus we finally get

$$\begin{aligned} 2\|\gamma(t)\| \frac{d\|\gamma(t)\|}{dt} &\leqslant 2\|\gamma(t)\| \left[\left\| \exp_* \frac{\partial a}{\partial t} \right\| + \left\| \pi_* \frac{\partial a}{\partial t} \right\| \right] \\ &= 2\|\gamma(t)\| [\|\exp_* \gamma'(t)\| + \|\pi_* \gamma'(t)\|] \\ &\leqslant 2\|\gamma(t)\| \sqrt{2} [\|\exp_* \gamma'(t)\|^2 + \|\pi_* \gamma'(t)\|^2]^{1/2} \\ &= 2\sqrt{2} \|\gamma(t)\|. \end{aligned}$$

Assuming that $\|\gamma(t)\| \neq 0$, we obtain (*) by dividing both sides by $2\|\gamma(t)\|$.

LEMMA 2. *Let (M, g) be a compact Riemannian manifold of dimension n and let π_0: $S(M) \to M$ be the unit sphere bundle. Define*

$$h: S(M) \to L \equiv \bigcup_{p \in M} L_p$$

by $h(X) = c(X)X$, where $c(X) \in \mathbf{R}$ is such that $c(X)X \in L_{\pi(X)}$. (Note that $c(X)$ exists, since M is compact, and hence of finite diameter.) Then h is a homeomorphism.

Proof. First we show that L is closed. Let A_n be a sequence of points in L converging to A. Then $\pi(A_n) \to \pi(A)$ and $\exp(A_n) \to \exp(A)$, by sequential continuity. Let d: $M \times M \to \mathbf{R}$ be the metric on M arising from g. Then

$$d(\pi(A), \exp(A)) = \lim_{n \to \infty} d(\pi(A_n), \exp(A_n)) = \lim_{n \to \infty} \|A_n\| = \|A\|$$

by sequential continuity, the fact that $A_n \in L$, and sequential continuity again, respectively. Thus the geodesic determined by A is length-minimizing. We must now show that no extension of this geodesic is length-minimizing. Consider the sequence

$$A_n' = \left(1 + \frac{1}{n}\right) A_n.$$

Let $B_n \in TM$ correspond to a length-minimizing geodesic joining $\pi(A'_n)$ to $\exp(A'_n)$. Some subsequence of B_n, say B_{n_i}, converges to a point B, since the B_n lie at a distance not greater than the diameter of M plus 1 from the set of zero vectors in TM (i.e., they lie in a compact set). If $B \neq A$, then no extension of A will yield a length-minimizing geodesic, by a corner-cutting argument, and thus, in this case, $A \in L$. If $B = A$, then for any neighborhood U of A there exists an $N \in \mathbf{Z}$ such that, for $i > N$, B_{n_i} and A'_{n_i} both lie in U. But $P(B_{n_i}) = P(A'_{n_i})$ and $B_{n_i} \neq A'_{n_i}$, the latter assertion following from the fact that the B_{n_i} correspond to length-minimizing geodesics while the A'_{n_i} do not since they are extensions of cut-locus vectors. Hence P is not locally one-to-one, a contradiction.

We now proceed with the proof of the continuity of h and h^{-1}. Let X_i be a sequence of points in $S(M)$ converging to X. Then any subsequence of $h(X_i)$ has a limit point $Y \in L$, since $h(X_i)$ lies in a compact set and L is closed. We may pick a subsequence of this arbitrary subsequence converging to Y. Since the obvious projection of TM modulo the zero vectors onto $S(M)$ is continuous, we get $Y/\|Y\| = X$ by sequential continuity of the projection with respect to the sub-subsequence. Thus $Y = h(X)$, and hence every subsequence of $h(X_i)$ has one and only one limit point, namely $h(X)$. It follows that $h(X_i)$ converges to $h(X)$ and h is continuous. Now h is obviously one-to-one and onto. If C is a closed set in $S(M)$, then C is compact. Therefore, $h(C)$ is compact, and hence closed. Thus h^{-1} is also continuous, and h is a homeomorphism.

LEMMA 3. *Let p and q be distinct points of S, a complete, simply connected, C^∞ Riemannian 2-manifold without conjugate points. Then the set of points equidistant from p and q, say $C(p, q)$, is a closed, connected, noncompact, C^∞-submanifold of S of dimension* 1. *Moreover, $C(p, q)$ has the following properties*:

(1) *If (r, θ) are polar coordinates about p, then $C(p, q)$ is regularly parametrized for its entire length by some function $F\colon (\theta_1, \theta_2) \to \mathbf{R}^2$, where $F(\theta) = \big(R(\theta), \theta\big)$ and $\theta_2 - \theta_1 \leqslant \pi$. Here we identify points of $S - p$ with their polar coordinates.*

(2) *If $R(\theta)$ is as in* (1), *then there is a unique θ_0 such that $R'(\theta_0) = 0$ and $R(\theta) > R(\theta_0)$ for $\theta \neq \theta_0$.*

(3) *At $x \in C(p, q)$, $C(p, q)$ bisects the angle formed by the geodesics from p to x and from q to x.*

We call the curve $C(p, q)$ the *equicurve* between p and q. If S has constant curvature, the equicurves in S are easily shown to be geodesics.

Proof. First we note that $C(p, q)$ is closed because it is the set of points of S where two continuous functions agree, namely the distance-from-p function, denoted by $d_p\colon M \to \mathbf{R}$, and d_q. Connectedness and noncompactness along with differentiability will be proved after showing a local version of property (1).

Let $x \in C(p, q)$ and let γ_p and γ_q be the unique geodesics joining p to x and q to x, respectively, and having unit speed and common minimizing-length r_0. From the Gauss lemma it is evident that the gradients of d_p and d_q at x are $\gamma_p'(r_0)$ and $\gamma_q'(r_0)$, respectively. Now, $\gamma_p'(r_0)$ and $\gamma_q'(r_0)$ are distinct, for otherwise, by uniqueness of geodesics having a common tangent vector, we would get $p = q$. Now, that $C(p, q)$ is locally differentiable follows from the fact that $d(d_p - d_q) = d(d_p) - d(d_q) \neq 0$ by the duality of differentials and gradients via the Riemannian metric. This proves that $C(p, q)$ is a C^∞-submanifold of dimension 1, since $d_p - d_q$ is a submersion in a neighborhood of $C(p, q)$. Now suppose we consider the component of $C(p, q)$ containing x and parametrize it via the polar coordinate θ with p as the pole. We note that $C(p, q)$ is never tangent to a geodesic ray from p, since this implies $\langle \nabla(d_p - d_q), \nabla d_p \rangle = 0$, so we have $\langle \nabla d_q, \nabla d_p \rangle = 1$, whence $\gamma_p'(r_0) = \gamma_q'(r_0)$, and $p = q$ by uniqueness of geodesics. Thus, each component c_i of $C(p, q)$ is regularly parametrized by θ on some interval (θ_1^i, θ_2^i). Assume $\theta_1^i - \theta_2^i > \pi$. Then we get a contradiction from the fact that there would be a geodesic through p connecting two points of C_i and a broken geodesic of the same length through q connecting them, which violates the fact that geodesics are length-minimizing under our hypotheses on S. Thus we have proved property (1) for each component of $C(p, q)$.

We proceed to prove (2) in general. Suppose there are two points of $C(p, q)$ which are tangent to the curves $r = \text{const}$. This implies the existence of two distinct unbroken geodesics from p to q since the gradients of d_p and d_q are anti-parallel at those two points, a contradiction. This establishes uniqueness. As for existence, we note that since $R(\theta)$ is monotone if θ_0 does not exist, a right-hand or left-hand limit of $R(\theta)$ for any component (of $C(p, q)$) must exist and not equal ∞ or 0, whence the component could be continued because of the fact that $C(p, q)$ is closed and transverse to the lines $\theta = \text{const}$. This also establishes that there is only one component of $C(p, q)$.

To prove property (3) assume that $x \in C(p, q)$ and that β is a curve through x and lying in $C(p, q)$, taking $\beta(0) = x$ and $\beta'(0) \neq 0$. Now

$$\langle \beta'(0), \gamma_p'(r_0) \rangle - \langle \beta'(0), \gamma_q'(r_0) \rangle = \langle \beta'(0), \gamma_p'(r_0) - \gamma_q'(r_0) \rangle$$

$$= \langle \beta'(0), \operatorname{grad}(d_p - d_q) \rangle = d(d_p - d_q)(\beta'(0)) = \frac{d}{dt}(d_p - d_q)(\beta(t))_{t=0} = 0.$$

LEMMA 4. *L_p has a finite number of vertices. Between consecutive vertices, L_p is a segment of an equicurve between 0 and some $Y \in \exp_p^{-1}(p)$ in M_p with metric $\exp_p^* g$.*

Proof. We first prove that if $X \in L_p$, then X lies on some equicurve of the form $C(0, Y)$, where $Y \in \exp^{-1}(p)-\{0\}$. Let γ_X be the unit speed geodesic in M associated to X via the exponential map at p. We recall here that, in the absence of conjugate points along geodesics in M, this exponential map is a Riemannian covering projection if we use the pulled-back metric on M_p. Since γ_X minimizes the distance between p and $\gamma_X(\|X\|)$, X must not be strictly closer than $\|X\|$ to any member of $\exp^{-1}(p)-\{0\}$. Indeed, otherwise we may find a geodesic from this member to X of length shorter than γ_X, and thus this geodesic projects to a geodesic shorter than γ_X joining p and the end of γ_X. On the other hand, if X is strictly closer to 0 than to $\exp_p^{-1}(p)-\{0\}$, then the same will be true of $(1+\varepsilon)X$ for sufficiently small $\varepsilon > 0$. Hence the corresponding ε-extension of γ_X would still be length-minimizing. In either case we contradict $X \in L_p$. Therefore

$$d\left(X, \exp_p^{-1}(p)-\{0\}\right) = \|X\|,$$

and since $\exp_p^{-1}(p)-\{0\}$ is closed, there is some $Y \in \exp_p^{-1}(p)-\{0\}$ realizing the distance. Hence $X \in C(0, Y)$.

Now we establish that there are a finite number of vertices. We know that there are at most a finite number of points of $\exp_p^{-1}(p)$ which are in the closed disk, say D, of radius twice the diameter of M and centered at 0, since otherwise $\exp_p$ would fail to be one-to-one in any neighborhood of a limit point of $\exp_p^{-1}(p)$. Now, since L_p is contained in the closed concentric subdisk of D of half the radius of D, it is clear that only a finite number of equicurves can contain points of L_p (i.e., only those between 0 and points of $\exp_p^{-1}(p)$ in D). Each vertex of L_p is the intersection of 2 or more of these equicurves. In fact, the number of these intersecting equicurves is by 1 less than the order of the vertex. We just lift the geodesic segments corresponding to the other members of $[V]$ in reverse (starting at V) to locate the members of $\exp_p^{-1}(p)$ generating the equicurves passing through V. Suppose that L_p has an infinite number of vertices. Then, since the number of equicurves with points on L_p is finite, there must be 2 distinct equicurves among them which intersect in an infinite set within the compact set D. It is not difficult to show that at any limit point of this infinite intersection set the two curves are tangent. We deduce easily from property (3) of Lemma 3 that both equicurves are between 0 and the same member of $\exp_p^{-1}(p)-\{0\}$ and must therefore coincide. Thus we arrive at a contradiction, and so L_p can only have a finite number of vertices.

Now it is trivial to observe that since L_p is a continuous image of the unit circle (by Lemma 1), it cannot pass from one equicurve to another without passing through a vertex. Hence, between consecutive vertices, L is a segment of a single equicurve.

LEMMA 5. *$\exp_p(L_p)$ is a 1-dimensional CW complex. Let $v = \exp_p(V)$, where $V \in L_p$. There are exactly $\#[V]$ length-minimizing geodesics joining p to v. For some small convex disk about v, these geodesics divide the disk into $\#[V]$ sectors. In each sector, there issues from v a 1-cell of $\exp_p(L_p)$, and this 1-cell bisects the angle of the sector.*

Proof. In making $\exp_p(L_p)$ a CW complex we let the 0-cells be $\exp_p\{V: V$ is a vertex of $L_p\}$. The 1-cells will be the images of the equicurve segments of L_p between consecutive vertices of L_p. Of course, we must verify that either the images of 2 of these segments are disjoint or coincide. Suppose

$$\exp_p(S_1) \cap \exp_p(S_2) \neq \emptyset,$$

where S_1 and S_2 are open equicurve segments between consecutive vertices of L_p. Then there are points $X_1 \in S_1$ and $X_2 \in S_2$ such that $\exp_p(X_1) = \exp_p(X_2)$. Since $\exp_p$ is a universal covering projection, there is a covering transformation $T: T_pM \to T_pM$ such that $T(X_1) = X_2$. Let Y_1 and $Y_2 \in \exp_p^{-1}(p)$ be the unique points such that $X_i \in C(0, Y_i)$. Now

$$d(0, X_2) = d(0, X_1) = d(T(0), T(X_1)) = d(T(0), X_2),$$

and thus $X_2 \in C(0, T(0))$. Consequently, $T(0) = Y_2$ and $T(Y_1) = 0$, or T may be the identity. In either case,

$$T(C(0, Y_1)) = C(T(0), T(Y_1)) = C(Y_2, 0).$$

Hence $T(S_1) = S_2$ since $T(S_1)$ can contain no vertex, $T(X_1) = X_2$, and the bounding vertices of S_1 cannot be mapped to interior points of S_2 because all interior points are of order 2. Thus S_1 and S_2 are identified by $\exp_p$. If $S_1 = S_2$, then $T(S_1) = S_1$ and $T(\bar{S}_1) = \bar{S}_1$, whence T has a fixed point in $\bar{S}_1$, T must be the identity and $X_1 = X_2$. Thus $\exp_p | S_1$ is one-to-one. Hence $\exp_p(L_p)$ is a 1-dimensional CW complex as indicated.

Each length-minimizing geodesic from p to v can be lifted to M_p at 0, and thus determines an element of L_p in $[V]$ and conversely. Thus there are $\#[V]$ length-minimizing geodesics from p to v. It is clear that these geodesics divide any convex disk about v into $\#[V]$ sectors. Let S be a sector bounded by consecutive incoming geodesics g_1 and g_2. Lift g_1 starting at 0 and ending at $V' \in [V]$ and lift g_2 in reverse starting at V' and necessarily ending at some point $X \in \exp^{-1}(p) - \{0\}$. By Lemma 3 (3), $C(0, X)$ bisects the angle between the lifts of g_1 and g_2 at V', and the part of $C(0, X)$ issuing from V' into the sheet of S at V' is a part of L_p since all other equicurves through V' must lie outside the sector bounded by the lift of g_1 and $C(0, X)$ because g_1 and g_2 are consecutive and Lemma 3 (3) applies again. Projecting the situation isometrically by $\exp_p$ establishes the final sentence of Lemma 5.

LEMMA 6. *The vertices of L_p are precisely those points where L_p has the interior angle less than π. In particular, the segments of L_p issuing from a vertex of L_p belong to different equicurves.*

Proof. At a non-vertex, L_p is smooth since in Lemma 4 we have shown that, between consecutive vertices, L_p is a segment of an equicurve and we know equicurves are smooth. Now, if V is a vertex of L_p, then its order is greater than 2, and hence 3 or more distinct geodesics corresponding to elements of the class of V connect p to $\exp(V)$. By Lemma 5, the images of the sides of L_p with members of the class of V as endpoints bisect the sectors determined by the above-mentioned geodesics, and so it is clear that the angle between consecutive side images is less than π. Pulling back angles to M_p preserves this inequality and we are done.

LEMMA 7. *Let D be a convex disk about $p \in M$ such that $D \times D$ is evenly covered by P. Let D_0 be the diagonal of $D \times D$, and let D^1 and D^2 be sheets of $D \times D$. Denote $D^i \cap P^{-1}(D_0)$ by D_0^i for $i = 1, 2$. Define $R^i\colon \pi^{-1}(D) \to \mathbf{R}$ by $R^i(X)$ being the distance from X to the point $D_0^i \cap M_{\pi(X)}$ in $M_{\pi(X)}$ with metric $\exp_{\pi(X)}^* g$. Then*

$$E \overset{\text{def}}{=} \{X \in \pi^{-1}(D) : R^1(X) = R^2(X)\}$$

is a C^∞ 3-dimensional submanifold of $\pi^{-1}(D)$. If $i\colon M_q \to TM$ is the inclusion, then i is transverse regular on E for each $q \in D$.

Note that M_q meets E in an equicurve. We call E the *equihypersurface* between D_0^1 and D_0^2.

Proof. The functions R^i, $i = 1, 2$, are mutually differentiable everywhere except on $D_0^1 \cup D_0^2$. The proof that E is a smooth submanifold of dimension 3 follows once we show that the differential of the function $R^1 - R^2$ is non-zero everywhere on E. Let $X \in E \cap M_q$ and $Y_i = D_0^i \cap M_q$, $i = 1, 2$. Let γ_1 and γ_2 be the geodesics (in M_q) with unit speed such that $\gamma_i(0) = Y_i$ and $\gamma_i(r_0) = X$, $i = 1, 2$. Now, we have

$$\begin{aligned} i^* d(R^1 - R^2)\big(\operatorname{grad}(R^1 - R^2)\circ i\big) &= d\big((R^1 - R^2)\circ i\big)\big(\operatorname{grad}(R^1 - R^2)\circ i\big) \\ &= \|\operatorname{grad}(R^1 - R^2)\circ i\|^2 \end{aligned}$$

but

$$\big(\operatorname{grad}(R^1 - R^2)\circ i\big)_X = (\operatorname{grad} R^1 \circ i)_X - (\operatorname{grad} R^2 \circ i)_X = \gamma_1'(r_0) - \gamma_2'(r_0) \neq 0$$

since $Y_1 \neq Y_2$ (recall they are in different sheets). Hence we have proved that not only is $d(R^1 - R^2)$ non-trivial on E, but also that $i^* d(R^1 - R^2) \neq 0$. The second relation implies that E is transverse to the submanifolds M_q, $q \in D$.

LEMMA 8. *Let $X \in L_q$ with $q \in D$, and suppose X is of order 2. Then there is a neighborhood W of X in TM such that $W \cap L = W \cap E$ for some equihypersurface E in TM.*

Proof. Since X is of order 2, it is a member of precisely one equicurve of the form $C(0_q, Y_q)$ with $Y_q \in \exp_q^{-1}(q)$. Now, this equicurve belongs to the equihypersurface E between the sheets above $D \times D$ containing 0_q and Y_q, respectively. It is apparent that an equihypersurface is uniquely determined by an equicurve lying within it in the case where we only consider those equihypersurfaces which have non-empty intersections with the set $\bigcup_{q \in D} L_q$ and where D_0^1 is the image of the 0-section of $\pi: \pi^{-1}(D) \to D$. Now it is evident that there are at most a finite number of such equihypersurfaces as above even if we allow q to range over all of D, since the second sheets of those equihypersurfaces are all disjoint, have the same volume and occupy a region of finite volume. Now, since this collection of equihypersurfaces is finite, the union of the ones not containing X is closed in $\pi^{-1}(D)$. We take W to be a neighborhood of X which is contained in the complement of this union and which is such that $W \cap E$ and $W \cap L$ are homeomorphic to R^3. (Note that L is homeomorphic to $S(M)$ by Lemma 2.) Now $W \cap L \subset W \cap E$ since every point of $W \cap L$ belongs to some equicurve which must belong to E by construction. Also, since L is closed, $L \cap W \cap E = W \cap L$ is closed in $W \cap E$. However, $W \cap L \approx R^3$, and hence is open in $W \cap E$ by invariance of domain. Thus $W \cap L = W \cap E$.

Proof of Theorems A and A′. We give $\exp_p(L_p)$ the CW structure defined in Lemma 5. Let $v_1, \ldots, v_k$ be the 0-cells of this complex. Let o_i be the number of length-minimizing geodesics from p to v_i (i.e., the number of points of L_p identified with v_i by the exponential map at p). By Lemma 5, the number of 1-dimensional simplices in the star of v_i is o_i since these 1-simplices bisect the sectors determined by the incoming geodesics. Remembering that each 1-simplex is bounded by 2 vertices and that $\exp_p(L_p)$ is homotopically equivalent to $M-p$ (via retracting radially along the geodesics starting at p), we have the following formula:

$$\chi(M)-1 = \chi(M-\{p\}) = k-\frac{1}{2}\sum_{i=1}^{k} o_i = \sum_{i=1}^{k}\left(1-\frac{1}{2}\,o_i\right).$$

Now, since $o_i \geqslant 3$, we have $k \leqslant -2(\chi(M)-1) = -2+4G(M)$, and since $1-2G(M) \neq 0$, we get $k \geqslant 1$. From the formula and the fact that $\chi(M) = 2-2G(M)$, we infer that

$$\sum_{i=1}^{k} o_i = 4G(M)+2k-2,$$

and using our bounds on k we get

$$4G(M) \leqslant \sum_{i=1}^{k} o_i \leqslant 12G(M)-6.$$

Thus we have proved the first sentence since $\sum_{i=1}^{k} o_i$ is the number of vertices of L_p. To prove the corresponding result for Theorem A′ we need only to substitute $\frac{1}{2}\chi(\bar{M})$ for $\chi(M)$ and proceed in an analogous manner.

Now we show that if $mn = 4G-2$, we can put a metric on a surface S of genus G (S compact) and find a point p on S such that L_p has $(m+2)n$ vertices, all of order $m+2$. Once this is shown, if we let $m = 1$ and $n = 4G-2$, we have proved that the upper bound on the number of vertices can be achieved, and $m = 4G-2$ and $n = 1$ will realize the lower bound. First we show that S can be represented topologically by a polygon with $(m+2)n$ sides with the sides identified in pairs. We start with the standard representation of S as a $4G$-gon with sides identified via the scheme $aba^{-1}b^{-1}cdc^{-1}d^{-1}$, etc., and consider a neighborhood of the point with which the $4G$ vertices are identified. We illustrate the procedure for the double torus (Fig. 1). Here we take $m = 2$ and $n = 3$.

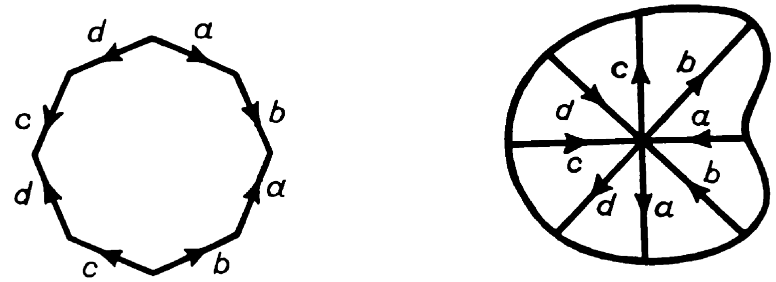

Fig. 1

We now "pull apart" the vertex, forming a string of n vertices with $m+2$ rays issuing from each and in the process we create $n-1$ new sides. Note that this is possible only if $mn+2 = 4G$. Fig. 2 clarifies this procedure.

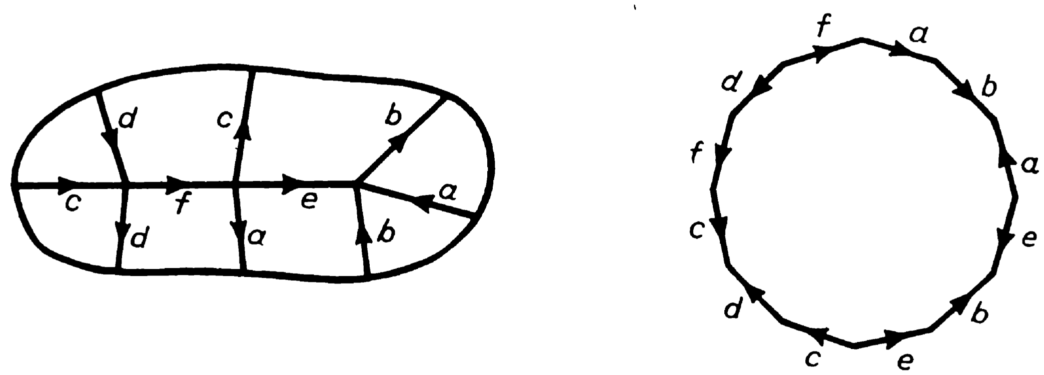

Fig. 2

We label the new sides e and f and cut along all the sides and obtain the desired polygonal representation upon unfolding (Fig. 2). Now we embed the new polygonal representation in the Euclidean plane E^2 (or the hyperbolic disk H^2) in such a way that it is symmetrically placed about the origin (or center) and has geodesic sides of equal length depending on whether $\chi(M) = 0$ (or $\chi(M) < 0$). Now, in the case where $\chi(M) < 0$ we may dilate or contract the embedding to make the vertices have a common interior angle of $2\pi/(m+2)$, once we note that in the Euclidean plane the vertices will have a common exterior angle of $2\pi/n(m+2)$, whereas the desired angle is $\pi - 2\pi/(m+2)$ which is always greater if $mn \neq 2$, as it is in the case where $G \neq 1$. In the case where $mn = 2$ (i.e., $G = 1$), the exterior angles are equal and the embedding must be into E^2. Now it is tedious but not difficult to prove that, since the angles at the vertices have been chosen correctly and the sides are all geodesic segments of equivalent length, upon identifying the vertices and sides in the prescribed manner we obtain the surface S together with a metric of constant curvature. We need only to observe that the boundary of the polygon in E^2 (or H^2) is the cut locus at the origin (or center), say O, if we identify E^2 (or H^2) with S_0 (with the metric pulled-back via $\exp_0$). Note that the proof for the case of non-orientable surfaces proceeds in an analogous way if we use the $a^2b^2c^2 \ldots$ representations to begin the process.

Proof of Theorem B. Let $f(p) = n$ and let $P_1, \ldots, P_n$ be points on the sides of L_p which are of order 2 and separate the n vertices. We let $W_1, \ldots, W_n$ be neighborhoods of these points as in Lemma 8, and put

$$U = \pi(W_1') \cap \ldots \cap \pi(W_n'),$$

where $W_i' = W \cap E_i$ and E_i is the equihypersurface containing P_i. Now U is open since the W_i' are 3-dimensional and transverse to the tangent spaces. We note that if $P_1, \ldots, P_n$ are in order of increasing θ (say $\theta = 0$ for P_1 and we have a fixed orientation on U), then P_i and P_{i+1} lie on different equihypersurfaces (we use addition mod n in adding indices), for otherwise the vertex between them would be smooth, violating Lemma 6. Since each $W_i \cap L = W_i \cap E_i$ is a smooth 3-dimensional submanifold of TM projecting onto an open superset of U, we can find sections $\Gamma_1, \ldots, \Gamma_n$, defined on U and taking values in $W_i \cap L$, such that $\Gamma_1(p) = P_1, \ldots, \Gamma_n(p) = P_n$. Furthermore, we note that these sections preserve the angular ordering of $P_1, \ldots, P_n$ (relative to the vector field Γ_1 on U defining $\theta = 0$) on some neighborhood of p contained in U, say U_0. Thus it is clear that $f(q) \geqslant f(p)$ for $q \in U_0$, since $\Gamma_1(q), \ldots, \Gamma_n(q)$ lie on distance sides of L_q.

Now let the function f be constant (with value n) and normalize the vertices of L_p for each p. We define a properly discontinuous action of Z_n on $S(M)$ (the total space of the unit circle bundle) by choosing an orientation on M (or its 2-fold cover), letting the generator of Z_n map each normalized vertex to the next in the positive direction, and extending this map linearly in θ between vertices. Now we note that locally the vertices determine smooth sections of the tangent bundle, since they are determined by the intersections of transversely intersecting pairs of equihypersurfaces. Also $f \equiv n$ assures us that a vertex cannot "split". Thus it is clear that the normalized vertices are bounded away from each other since $g: M \to \boldsymbol{R}$ defined by the minimum angle between adjacent vertices at each point is continuous and greater than 0 on compact M, and hence bounded away from 0. Thus the action is clearly properly discontinuous and $Q: S(M) \to S(M)/\boldsymbol{Z}_n$ is a covering projection. Now the function $\sigma: M \to S(M)/\boldsymbol{Z}_n$ which assigns to each point the orbit of vertices of the cut locus at that point is a section of the bundle $\pi': S(M)/\boldsymbol{Z}_n \to M$. It is not difficult to see that the primary obstruction cocycle for finding a section of the circle bundle with map π' is n times that of the unit tangent bundle (i.e., the Euler class). Thus the existence of σ implies $n\chi(M) = 0$, and hence $\chi(M) = 0$.

Proof of the Corollary. If there are no points p and q, $p \neq q$, with 4 or more length-minimizing geodesics joining them, then every vertex of every cut locus, no matter what point of M we choose, has order 3. We have already seen in proving Theorems A and A′ that this can be the case only if f (of Theorem B) assumes the maximal value of $12G-6$ everywhere (or $6G$ in the non-orientable case). Thus f is constant and $\chi(M) = 0$ by Theorem B.

5. An interesting proof of a known result. In Lemma 1 we proved that, for appropriate (M, g), P was a covering projection. For M compact with $\chi(M) \neq 0$, there are no covering transformations other than the identity, since a non-trivial covering transformation applied to the 0-section would yield a nowhere-vanishing vector field. This observation leads to the following previously known result:

THEOREM. *If (M, g) is compact and without conjugate points, then the center of $\pi_1(M, p)$ is trivial provided $\chi(M) \neq 0$.*

Proof. There is a natural isomorphism

$$\varphi\colon \pi_1(M \times M, (p, p)) \to \pi_1(M, p) \times \pi_1(M, p).$$

Let Δ be the subgroup of $\pi_1(M, p) \times \pi_1(M, p)$ consisting of the diagonal. The claim is that

$$\varphi^{-1}(\Delta) = \operatorname{im}[P\#\colon \pi_1(TM, 0_p) \to \pi_1(M \times M, p \times p)].$$

Just note that any loop in TM is homotopic to a loop in the 0-section of TM and then projects to a loop of the form $\sigma \times \sigma$. Conversely, every loop of the form $\sigma \times \sigma$ in $M \times M$ lifts to a loop in TM, namely $\tilde{\sigma}(t) = 0_{\sigma(t)}$. Now the group of covering transformations is isomorphic to $N(\Delta)/\Delta$, where $N(\Delta)$ is the normalizer of Δ in $\pi_1(M, p) \times \pi_1(M, p)$. Thus, by our observation, $N(\Delta) = \Delta$. Now $(h, g) \in N(\Delta)$ if $(h, g)(c, c)(h^{-1}, g^{-1})$ is of the form (d, d) for all $c \in \pi_1(M, p)$, or $hch^{-1} = gcg^{-1}$ or $(g^{-1}h)c(g^{-1}h)^{-1}$

$= c$ for all c, that is if $g^{-1}h$ belongs to the center C of $\pi_1(M, p)$. Thus, if $f \in C$, then $(f, 1) \in N(\Delta)$, and so $f = 1$.

6. Concluding remarks. Theorems A, A′, and B suggest looking further into the structure of the sets $f^{-1}(n)$ with $4G(M) \leqslant n \leqslant 12G(M)-6$, where $f(p)$ is the number of vertices of L_p. Alan Weinstein has suggested that there may be some relation between $f^{-1}(n)$, for small n, and Weierstrass points. Any results in this direction would be of significant interest.

In his thesis, Buchner [1] proved that generically the cut locus L_p, at a fixed point p, is stable under variations of the metric on M (for $\dim M \leqslant 5$). For surfaces without conjugate points, it follows from his result that (for a fixed p) L_p will have only vertices of order 3 generically. Theorem B implies that, for $\chi(M) \neq 0$, L_q must have vertices of order greater than 3 at some $q \in M$. A conjecture is that generically these points q are isolated or perhaps they form a 1-dimensional subcomplex of M (**P 1210**). To answer this problem, a study of the subset $C \subset M \times M$, consisting of pairs (p, q) such that $q \in \exp_p(L_p)$, should be made. Specifically, what are the crucial generic properties of C under variations of the metric on M?

Finally, the reader may wish to consult the work of Weinstein [6] and Eberlein [2] for additional sources pertaining to the cut locus.

REFERENCES

[1] M. Buchner, *Stability of the cut locus*, Thesis, Harvard University, 1974.
[2] P. Eberlein, *The cut locus of noncompact finitely connected surfaces without conjugate points*, unpublished (available from the author at the University of North Carolina).
[3] S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, Vol. I, New York 1963.
[4] S. B. Myers, *Connections between differential geometry and topology: I. Simply connected surfaces*, Duke Mathematical Journal 1 (1935), p. 376-391.
[5] — *Connections between differential geometry and topology: II. Closed surfaces*, ibidem 2 (1936), p. 95-102.
[6] A. Weinstein, *The cut locus and conjugate locus of a Riemannian manifold*, Annals of Mathematics 87 (1968), p. 29-41.

Reçu par la Rédaction le 23. 11. 1977