ILLINOIS JOURNAL OF MATHEMATICS
Volume 31, Number 1, Spring 1987

THE RIEMANNIAN OBSTACLE PROBLEM

BY

STEPHANIE B. ALEXANDER, I. DAVID BERG AND RICHARD L. BISHOP

1. Introduction

In this paper we consider the local distance geometry of Riemannian manifolds with boundaries. We view a boundary component as an obstacle around which a geodesic can bend, or at which a geodesic can end. Our emphasis is on the structure of fields of geodesics. In the presence of an obstacle, the description of such fields of geodesics in terms of differential equations is no longer feasible; as an alternative, we produce a key differential inequality which functions as a one-sided version of the Jacobi equation. In consequence we obtain local bipoint uniqueness and a geometric estimate on the cut distance, that is, the distance below which bipoint uniqueness holds. On the other hand, unique determination of geodesics by their initial tangents (Cauchy uniqueness) clearly fails; we have developed basic techniques to establish properties of the field of geodesics with a common tangent.

It is immediately clear that the Riemannian obstacle problem is natural from the variational and mechanical points of view. Consider, for example, a string in Euclidean 3-space, stretched around an obstacle, sometimes following the obstacle, sometimes travelling through the air. Or, the problem can be stated as that of analyzing the propagation of wavefronts (level surfaces of the distance function) around an obstacle. If the wavefronts are due to a point source of light or sound in a nonhomogeneous medium, then the appropriate geometry is Riemannian rather than Euclidean. Arnol'd has considered the obstacle problem in a series of recent papers. Arnol'd is carrying out a general program which identifies standard singularities related to the geometry of groups generated by reflections with normal forms for singularities occurring in variational problems. This investigation leads him to variational problems with one-sided constraints, and in particular to an analysis of the singularities of wavefronts for Euclidean obstacles in general position. (For a survey, see [5]; see also [6], [7], [8].) Arnol'd achieves an analysis in this case even though, as he states, "the problem of going around an obstacle has not yet been solved even in Euclidean 3-space" [5].

Received March 11, 1985.

Classical differential equations techniques cannot suffice for the obstacle problem. No matter how smooth the obstacle, we cannot assume the geodesics are C^2; they are not, in general, governed by a second order differential equation with Lipschitz conditions. Moreover, each geodesic is the union not only of boundary and interior segments, but also of a set of points which lie on no nontrivial boundary or interior segment. This set can have positive measure. Arnol'd rules out such points by assumptions of general position. However, since such "chattering" behavior can occur as the limit of finite switching behavior, it seems clear that a quantitative understanding of the latter would be tantamount to an understanding of the former. Also, it is uncertain which assumptions on the boundary guarantee finite switching behavior (see, for example, the question at the end of this paper). For our purposes, therefore, it is desirable to make no assumptions on the boundary beyond smoothness.

Note that general methods in the theory of variational problems with boundary constraints (see, for example, Almgren's book [3]) do not yield the regularity and uniqueness properties which we seek. As Antman remarks, on the subject of variational inequalities: "The analysis of regularity of solutions, still the main source of difficulty, is forced to accommodate the peculiarities of each special class of problems" [4]. Certainly the Cauchy uniqueness guaranteed to solutions of second order differential equations with Lipschitz conditions is violated with a vengeance here, since a geodesic might elect to hug the boundary or to peel off in a C^1 manner into the interior. Much of our work is devoted to considering just what uniqueness properties obtain.

Other authors also have considered regularity and uniqueness questions in the Riemannian obstacle problem. Wolter [19], [20] has shown that the distance function is C^1 at interior points in any neighborhood where bipoint uniqueness holds; and furthermore that the gradient vector field of the distance function at interior points is locally Lipschitz continuous at exactly those interior points for which the geodesic segment which realizes the distance exists and can be extended to be minimizing for a larger distance. He has examined several natural but, as he shows, distinct definitions of cut locus in a Riemannian manifold M with boundary, and related these cut loci to the set on which the distance function fails to be C^1. Wolter has introduced a hypothesis on the boundary which is sufficient for these results, namely, that M be locally C^1 diffeomorphic to a convex set.

Scolozzi [17] has given an independent proof of bipoint uniqueness. Marino and Scolozzi [13] have shown that geodesics have Lipschitz continuous derivatives, and that under suitable hypotheses on an obstacle in Euclidean space there exist infinitely many geodesics, the supremum of whose lengths is infinite, joining two given points. Scolozzi [18] has proved the existence of a nonconstant closed geodesic in this setting. These papers apply a theory of functionals, not necessarily C^2 or convex, on infinite-dimensional spaces which was initiated by De Giorgi, Marino, and Tosques [12].

Other references include [9], and for convex obstacles, [1], [14], [15].

Our methods in this paper are almost purely geometric. In Section 2 we give what we feel is a definitive statement of regularity of geodesics, involving a decomposition into tangential and normal parts. Specifically, the tangential part is smoother by one degree than the geodesic itself, and the normal part satisfies a convexity condition. In Section 3 we establish the differential inequality and consequently estimate the cut distance in terms of an extrinsic curvature invariant which we call the tubular radius. The differential inequality further proves itself as a powerful tool in Section 4, which considers the convergence of geodesics, and in Section 5, where we show the existence and continuity of Jacobi fields. In Section 6 we initiate an investigation of the Cauchy uniqueness question, by studying the initial tangent map which bipoint uniqueness provides. We show that on the intersection of the boundary with a sufficiently small distance ball centered at a boundary point p, this is a map of degree one which sweeps out the "horizon" visible from p along interior geodesics. Our conjecture about Cauchy uniqueness (now proved [2A]) is that this map is one-one. In particular, the injectivity of this map implies that if two geodesics have the same initial tangent, then one must be locally an involute of the other.

The authors acknowledge a debt of gratitude to their friend and colleague Professor Felix Albrecht, who has given his helpful criticism on a regular basis.

2. Regularity of geodesics

We consider a C^∞ Riemannian manifold M with C^∞ boundary B, and define geodesics to be locally shortest paths. Another reasonable approach, which reduces to the same thing, would be to require the expected differential properties; namely, a geodesic can be specified as an H^2 (Sobolev space) curve such that the acceleration where it exists is either 0 (at points of the interior) or outwardly normal (at points of B) [13], [17]. Actually even more is true: a geodesic fails to have an acceleration only at a countable number of points, and at those there are one-sided accelerations. These exceptional points are what we call *switch points*, where the geodesic switches from a boundary segment to an interior segment or vice-versa. (To be precise, by an interior segment we mean an open segment which has zero acceleration in M but which may include points of B; a boundary segment [also open] must have nonzero acceleration, necessarily normal to B, on a dense open subset.) Besides the switch points, boundary segments, and interior segments, one other kind of point is possible, an accumulation point of switch points, which we call an *intermittent point*. Examples of single geodesics are easily constructed which have a Cantor set of positive measure of intermittent points, although it seems clear that when the boundary is generic geodesics have none.

The existence of intermittent points makes the variational theory of geodesics difficult to analyze, but they do not cause the acceleration to fail to exist or even be discontinuous at the intermittent points themselves, only at numerous nearby points. Indeed, we observe below that the acceleration at an intermittent point vanishes.

The detailed arguments for the asserted regularity are rather intricate and can be approached either geometrically or analytically. There are neat geometric arguments that show geodesics are C^1 and have normal osculating planes at B [2]. In particular, the acceleration at a boundary point is normal to B when it exists. The step further, to show that velocities are absolutely continuous or Lipschitz continuous seems to require more careful estimates; see, for example, [13]. We have done this by a secant approximation argument which accounts for the bounded extrinsic curvature of B (this is where the smoothness of B enters) and a comparison of distances in M and in any Riemannian extension of M across B. This comparison is of independent interest: it asserts that the difference between the two distances is on the order of the cube of either one. We do not give the details in this paper. Wolter [20] and Marino and Scolozzi [13] use calculus of variations for all their considerations of regularity.

To show that the acceleration at an intermittent point vanishes, we argue by exclusion: B cannot be convex in that direction because that would allow us to shorten the geodesic by cutting across the interior. Nor can the boundary be concave in that direction, because then the normal projection onto the boundary would shorten curves which were not already on the boundary. Thus, the normal curvature of the boundary in that direction is 0. This normal curvature is identified as the acceleration of γ in the proof of Theorem 1 below.

To carry the regularity analysis a step further we split the geodesics into tangential and normal parts. We explain this decomposition in terms of special coordinates x_i adapted to the boundary. We let x_n be the distance from B. Starting with arbitrary coordinates x_i, $i < n$, on B, we extend them to be constant on ordinary geodesics normal to B. Denote the Christoffel symbols of the Levi-Civita connection of M by Γ_{ijk}. Because $\partial/\partial x_n$ is the unit normal field on B, the coordinate matrix of the second fundamental form of B with respect to $\partial/\partial x_n$ is just $-\Gamma_{ijn}$, where $i,\ j < n$. For a geodesic γ we let $\gamma' = \Sigma x_i' \partial/\partial x_i$ denote the velocity. On a boundary segment of γ, we denote the normal curvature of B in the direction of γ by κ; then

$$\kappa = -\sum\nolimits_{i,\,j<n} x_i' x_j' \Gamma_{ijn}.$$

When γ is not in B the same expression occurs in the differential equations for γ, so that if we define κ to be 0 off of boundary segments, then the differential equations can be unified to cover, in an integral sense, all points of γ, as follows:

(1) $x_k'' = -\sum_{i,j} x_i' x_j' \Gamma_{ijk}$ for $k < n$,

(2) $x_n'' = -\kappa - \sum_{i,j<n} x_i' x_j' \Gamma_{ijn}$.

Indeed, the curve on B with coordinates x_k, $k < n$, is the normal projection of γ to B, and we call it the *tangential part* of γ. The *normal part* of γ is simply x_n, the height of γ above B. Equations (1) and (2) are first of all obvious on open boundary and interior segments. Next, given the absolute continuity of γ' and the normality of γ'' on B, they hold everywhere in an integral sense. But then we see that the right side of (1) is continuous, so that (1) holds everywhere as it is. At the countably many switch points (2) can be interpreted as being valid in the limit from either side. The right side of (2) is continuous, and hence the acceleration of γ exists, except at switch points; in particular, the acceleration exists and is 0 whenever $\kappa = 0$. We state our conclusions about this regularity as a pair of theorems.

THEOREM 1. *The tangential part of a geodesic is C^2, and its second derivative is locally Lipschitz.*

The second theorem concerns the regularity of the normal part; once stated it is obvious. A convexity property compensates for the lack of differentiability. To describe it we use Theorem 1 to imbed the geodesic in a $C^{2,1}$ smooth surface, namely, the surface S swept out by the normal geodesics to B along the tangential part of γ (and hence along γ itself). This surface has a boundary $S \cap B$ and it is sensible locally to consider convex sets in S and convex hulls of subsets.

THEOREM 2. *Sufficiently short segments of γ are boundary segments of convex sets in S. Specifically, for $\gamma|[a, b]$ we take the convex hull of the set consisting of $\gamma(a)$, $\gamma(b)$, and the segment of $S \cap B$ running between the projections of $\gamma(a)$ and $\gamma(b)$.*

3. Local bipoint uniqueness

Locally we expect geodesic segments to be determined by their ends, but it is a delicate matter to prove such a result because it requires the boundedness of curvature. Bipoint uniqueness fails, for example, when the boundary consists of two spherical caps glued together along a common circle (neither great). In any neighborhood of the sharp edge there are numerous geodesic segments connecting a pair of points on the edge. At most two of these segments are minimal and the others oscillate back and forth across the edge. The edge itself is not a geodesic but it is a limit of geodesics. These features are retained when the edge is smoothed to make the surface C^1 but leaving infinite normal curvature.

We find it convenient to formulate the curvature estimate in terms of an isometric imbedding of M in some Euclidean space N. It would be desirable to have a more direct link with the natural intrinsic invariants, the sectional curvature of M, the normal curvature of B in M, and the injectivity radii of M and B, but our method has the advantage that it gives the estimate in terms of a single number. Specifically, we say that a positive number r is a *tubular radius* for M in N if every point at distance r or less from M is the center of a closed ball which meets M at a single point. Note that then a geodesic in M at every point has its radius of curvature in N extending beyond the tubular radius. The normal curvatures of M and those of B which belong to a normal vector whose M-component is outward from the interior of M are bounded above by $1/r$. Conversely, if we take an upper bound k of such normal curvatures and restrict our scope sufficiently to prevent global opposition, then $1/k$ will be a tubular radius for the restricted region of M.

The essence of any bipoint-uniqueness result is to obtain a lower bound on the distance two geodesics have to travel before they can rejoin, once parted.

THEOREM 3. *If r is a tubular radius for M, then two geodesics in M each extend more than πr from a point where they part to where they join again. (This is a sharp estimate: consider the case for which B is a sphere of radius $r' > r$ in Euclidean 3-space.)*

We prove Theorem 3 as a part of a more technical theorem, the heart of which is a differential inequality which we find useful for related purposes, because of both its technical and intuitive content.

THEOREM 4. *Let $r = 1/k$ be a tubular radius for M, and let γ and σ be geodesics in M having speed no more than one. Let $f(r) = |\gamma(s) - \sigma(s)|$ be the Euclidean displacement between corresponding points. Then except at the countably many points where f'' fails to exist, we have a differential inequality*

(a) $f'' \geq -k^2 f$ *(with strict inequality where $f > 0$).*

Consequently,

(b) *if $A\sin(ks+b)$, which we denote by $g(s)$, coincides with $f(s)$ at $s = u$ and $s = t$, where $|t - u| < \pi r$, then $f(s) < g(s)$ for $u < s < t$ (unless $f = g = 0$ identically on that interval);*

(c) *if we have $g(s) = A\sin(ks+b)$, and if for some u we have $g(u) = f(u) > 0$ and $g'(u) = f'(u)$, then $f(s) > g(s)$ for $s \neq u$ on the interval of length πr containing u on which $g(s) > 0$;*

(d) *if $\gamma(0) = \sigma(0)$, but $\gamma(s) \neq \sigma(s)$ for $s > 0$ arbitrarily near 0, then $\gamma(s) \neq \sigma(s)$ for $0 < s \leq \pi r$.*

Proof. Let $X = \gamma - \sigma$, the Euclidean displacement vector field from points on γ to points on σ. Thus, $X = fU$, where U is a unit vector function of s, defined except when $X = 0$. We have that X'', f'' exist everywhere as one-sided derivatives, with only countably many points where they are not two-sided. The same is true of U'' on the subset where $f > 0$. When $f(s) = 0$, the first derivative of f is one-sided unless it is 0, since we can calculate that

$$f'(s-) = -f'(s+) = \sqrt{\langle X'(s), X'(s)\rangle}\,;$$

but then f' has a positive jump at s.

At points where $f > 0$ and X'' exists we calculate:

$$X'' = f''U + 2f'U' + fU'',$$

and taking inner products with $X = fU$, using the facts that $\langle U', U\rangle = 0$ and $\langle U'', U\rangle = -\langle U', U'\rangle$, we get $\langle X'', X\rangle = f''f - f^2\langle U', U'\rangle$. That is,

$$f'' = \langle X'', X\rangle/f + f\langle U', U'\rangle. \tag{1}$$

The Euclidean accelerations of γ and σ can be written $\gamma'' = \kappa_\gamma N_\gamma$ and $\sigma'' = \kappa_\sigma N_\sigma$, where N_γ and N_σ are unit normals to M or B and the κ's are bounded by the corresponding normal curvatures, hence $\leq k$. (The κ's may not be normal curvatures because γ and σ may be slower than unit speed. We did not insist they have the same speed because we want (d) to be strong enough to prevent geodesics from rejoining at different lengths.)

Let $p = \gamma(s) + rN_\gamma(s)$, so that by the definition of r, $\sigma(s)$ is outside the sphere of radius r at p. Hence $\langle X + rN_\gamma, X + rN_\gamma\rangle > r^2$, that is $f^2 = \langle X, X\rangle > -2r\langle X, N_\gamma\rangle$. Similarly, $f^2 > 2r\langle X, N_\sigma\rangle$. Hence substituting $X'' = \kappa_\gamma N_\gamma - \kappa_\sigma N_\sigma$ into (1) gives

$$\begin{aligned} f'' &> -(\kappa_\gamma + \kappa_\sigma)f^2/2rf + f\langle U', U'\rangle \\ &\geq (-k^2 + \langle U', U'\rangle)f. \end{aligned} \tag{2}$$

Even at points where $f = 0$ the inequality $f'' \geq 0$ holds in the sense that either f' has a positive jump or, if $f' = 0$ too, f'' exists and is nonnegative (at a minimum!). Thus, we have proved a strong version of (a), which includes information about the exceptional points where f'' is one-sided.

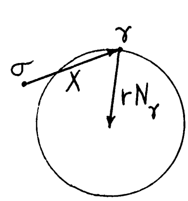

FIG. 1

The remaining parts, (b), (c), and (d), are standard. However, it is interesting to observe that they can be read off from standard convexity theory in conjunction with the following lemma, which has independent geometric interest as well as an application to the regularity properties of Jacobi fields (see Section 5).

LEMMA. *The graphs of the family of sinusoids $g(s) = A\sin(ks + b)$ are locally analytically equivalent to the family of lines in a Euclidean plane. Under this equivalence the solutions of the differential inequality $f'' > -k^2 f$ are transformed to locally strictly convex curves, with the origin on the convex side, with $\theta = ks$ as the central angle, and the points where $f = 0$ corresponding to the points where the curve goes to infinity.*

Proof. As a first step we consider $r = g(s)$ and $\theta = ks$ as polar coordinates, so that the sinusoids $r = A\ \sin(\theta + b)$ become circles through the origin. The second step is to make an inversion in the unit circle, under which our sinusoids are transformed to lines not through the origin. Thus, there is an analytic covering map from the upper half plane containing the graphs of the sinusoids $A\ \sin(ks + b)$ onto the punctured Euclidean plane, for which each sinusoidal arc is carried to a whole straight line.

At each point of a solution of $f'' > -k^2 f$ there is a sinusoid g which has first order contact with f at that point. Since $g'' = -k^2 g$, the difference $f - g$ has a positive second derivative at the point in question, from which we conclude that g is locally a support curve for f on the lower side. In the plane with polar coordinates (r, θ), f becomes a curve supported on the inside by circles through the origin, and under inversion, a convex curve as described in the lemma.

COROLLARY. *If the graph of a continuous function f satisfies either* (b) *or* (c) *of Theorem* 4, *then*

(a) *f has left and right derivatives everywhere,*

(b) *there are only countably many points where f' fails to exist and at these exceptional points f' has a positive jump, and*

(c) *f'' exists a.e. and satisfies $f'' \geq -k^2 f$.*

Proof. These are merely standard properties of convex curves, which f inherits through the transformations in the proof of the lemma.

Remark. The only properties of geodesics which are used in the proof of Theorem 4 are the differential properties; that is, the acceleration of a geodesic exists a.e. and when it is not 0 it is outwardly normal to B. Thus, if we chain together two geodesic segments with matching tangents at the join, then the resulting curve is a geodesic because it satisfies the differential properties. That is, for pairs of points on it near the join point it must coincide with the unique shortest geodesic between those pairs.

The most commonly used form of bipoint uniqueness incorporates the local existence of minimal segments as well. Local existence is well known, stemming from the local compactness of M [16].

THEOREM 5. *Every point of M has a neighborhood U such that for every p, q in U,*

(a) *there is a unique minimal geodesic segment joining p and q, and*

(b) *there is no other geodesic segment joining p and q and lying in U.*

(We shall show elsewhere that we can make U convex; that is, that the unique segments of (a) are contained in U.)

The proof simply requires successive shrinking of some initial neighborhood until it satisfies the condition for local existence (compact closure), and the existence of a tubular radius, and finally reduction to a size for which a segment leaving the previous neighborhood and returning would be too long to be minimal.

We call a neighborhood U which satisfies the conditions (a) and (b) of Theorem 5 a *neighborhood of bipoint uniqueness*. Henceforth when we specify a local result we assume that we are working within such a neighborhood.

4. Convergence of geodesics and their velocities

Consider a sequence of unitspeed geodesic segments γ_i, parametrized by $[0, l]$. In the following lemma assume that the segments are all contained in a compact region with tubular radius r and that $l < \pi r$.

LEMMA. *If $\gamma_i(0)$ and $\gamma_i(l)$ converge, then γ_i and γ_i' converge uniformly on $[0, l]$ to a geodesic segment γ and its velocity field γ'.*

Proof. It follows from Theorem 4(b) that if ε_{ij} is the larger of the distances $\rho(\gamma_i(0), \gamma_j(0))$ and $\rho(\gamma_i(l), \gamma_j(l))$, then there is a uniform bound

$$\rho\big(\gamma_i(s), \gamma_j(s)\big) \leq A\varepsilon_{ij}$$

for which A depends only on l and r. (In fact, a calculation shows that $A = \sec(l/2r)$.) By Cauchy's criterion γ_i converges uniformly to a function γ on $[0, l]$, which is continuous because each γ_i is continuous. The continuity of the distance function and the local minimizing properties of the γ_i in (possibly smaller) neighborhoods of bipoint uniqueness about the limit points imply that γ is also a geodesic.

We known that $k = 1/r$ is a uniform bound on the Euclidean acceleration of unitspeed geodesics: $\|\gamma''\| \leq k$. Hence k is a Lipschitz constant for their velocities, and velocities can be uniformly approximated by secants according to the formula

$$\|(\gamma(s) - \gamma(u))/(s - u) - \gamma'(u)\| \leq |s - u|k/2.$$

This holds for the γ_i as well as γ. Thus, if $\|\gamma_i - \gamma\| < \varepsilon$, we can add and subtract some secant approximations inside the expression $\|\gamma_i'(u) - \gamma'(u)\|$ and apply the triangle inequality to obtain

$$\|\gamma_i'(u) - \gamma'(u)\| < |s - u|k + 2\varepsilon/|s - u|.$$

Here s and u are arbitrary within $[0, l]$. For any u, given $\alpha > 0$ sufficiently small we can take s so that $|s - u|k = \alpha/2$ and take $\varepsilon = \alpha^2/4k$. For sufficiently large i we will have the presumed condition $\|\gamma_i - \gamma\| < \varepsilon$ and consequently $\|\gamma_i' - \gamma'\| < \alpha$. This proves the uniform convergence of velocities.

Remark. The quadratic relation between ε and α in the above proof is not an accident of the technique. If we take B to be a circle in Euclidean 2-space, with M the outer region, then we can take as our geodesics γ_i an arc of B followed by a tangent straight line segment. Let the end $\gamma_i(l)$ converge to $p \in B$ along an involute of B. Then the relation between

$$\|\gamma_i(l) - p\| \quad \text{and} \quad \|\gamma_i'(l) - \gamma'(l)\|$$

is indeed asymptotically quadratic.

THEOREM 6. *If a sequence of geodesics γ_i converges pointwise, then the limit function is a geodesic γ and the convergence of both γ_i and γ_i' to γ and γ' is uniform on closed bounded segments.*

Proof. If we cover γ by compact regions having tubular radii, then the convergence to a geodesic and its velocity follows immediately from the

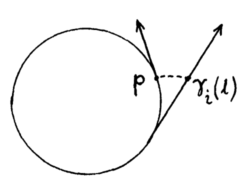

FIG. 2

lemma. For closed bounded segments we can use finitely many regions, so that the uniformity is obvious.

If U is a neighborhood of bipoint uniqueness, then for p in U we define the *radial vector field X from p* on $U - \{p\}$ by $X(q) = -\gamma_q'(0)$, where γ_q is the unique unitspeed minimizing segment from q to p. An obvious corollary of Theorem 6 is the following.

COROLLARY. *The radial field from p is continuous.*

More generally, we define a function F on triples (p, q, s) by

$$F(p, q, s) = \gamma_{p,q}'(s),$$

where $\gamma_{p,q}$ is the unitspeed segment from p to q. Clearly, F is continuous on an appropriate open subset of $(U \times U - \Delta) \times P$, where Δ is the diagonal of $U \times U$ and P denotes the nonnegative real numbers.

It is not difficult to show that the distance from a fixed p is a C^1 function on $U - \{p\}$ and that its gradient vector field is X. We have a proof based on the triangle inequality and the cubic distance approximation. Moreover, integral curves of X or $-X$ are geodesics, because they realize distance. As a consequence there is a form of Gauss's Lemma, namely, small geodesic spheres about a point are C^1 and the radial geodesics are orthogonal to the spheres. For $-X$ an integral curve is uniquely determined by its initial point q, since it must be the geodesic to p. But for X the integral curves can bifurcate, for example, at points of B where the normal curvature is nonnegative. Our study of involutes, Section 6, is concerned with the bifurcation of these integral curves.

5. Jacobi fields

Let γ_i be a sequence of geodesics defined on some interval $[0, l]$ and converging to a geodesic γ. We show that if the endpoints of the γ_i converge tangentially to vectors $J(0)$ and $J(l)$, then some subsequence of the γ_i converges tangentially to a continuous vector field J on the interior of γ. The argument uses for the first time the full differential inequality (2) of Section 3 for the distance between two geodesics. Until now we have been applying Theorem 4, thereby ignoring a term which reflects the extent to which the geodesics are skew to one another. We conjecture that the original sequence itself converges tangentially to J, but this seems to require a more advanced theory of Jacobi fields than we have yet developed. Since the argument is local, we suppose that M has a tubular radius $r = 1/k$ and that $l < \pi r/2$.

Let t_i be a parameter sequence of positive numbers converging to 0. If $X_i = \gamma_i(s) - \gamma(s)$ is the Euclidean displacement vector, set

$$J(s) = \lim t_i^{-1} X_i.$$

THEOREM 7. *If J exists for $s = 0$ and $s = l$, then, passing to a subsequence of the γ_i, J exists and is continuous for $0 < s < l$.*

We note that J may be discontinuous at the endpoints. Indeed, consider a geodesic γ which lies on the boundary and has nonvanishing normal curvature. Geodesics γ_i may be obtained from γ by lifting at the endpoints, producing a Jacobi field J along γ which vanishes on $(0, l)$ but not at 0 or l.

Proof. Set $f_i = \|X_i\|$, and $U_i = f_i^{-1}X_i$ where $f_i \neq 0$. The existence of $J(s)$ is equivalent to that of $f(s) = \lim t_i^{-1}f_i(s)$, and of $U(s) = \lim U_i(s)$ where $f(s) \neq 0$.

By Theorem 4(b), f_i is dominated by the sinusoid

$$F_i(s) = A_i \sin(ks + b_i)$$

which coincides with f_i at $s = 0$ and $s = l$. It follows that any limiting value of $t_i^{-1}f_i(s)$ is dominated by $F(s)$, where F is the sinusoid which takes the values $f(0)$ and $f(l)$ at $s = 0$ and $s = l$. Now that infinite velocities are ruled out, we can use a diagonal process to pass to a subsequence of the γ_i for which J exists on a countable dense subset S of $[0, l]$.

To show that f exists and is continuous on $(0, l)$, it suffices to show, for fixed s in $(0, l)$, that any limiting value $C < \infty$ of $t_i^{-1}f_i(s)$ as $i \to \infty$ is equal to any limiting value $D \leq \infty$ of $f(s_i)$ as $s_i \to s$, for $s_i \in S$. We may suppose s_i converges to s from the left. As before, the sinusoid F with values $f(s_1)$ and C at s_1 and s, respectively, dominates $f(s_j)$ for all $s_j > s_1$. Therefore $D \leq C$. Now choose $t > s$, $t \in S$. The sinusoid F which takes values $f(s_i)$ and $f(t)$ at s_i and t, respectively, satisfies $F(s) \geq C$. Choosing s_i arbitrarily close to s gives $D \geq C$.

Now suppose that $f(s) \neq 0$; choose a subinterval $[a, b]$ of $[0, l]$ on which f is never 0 and with $s \in (a, b)$. Note first that there are a slightly smaller interval I about s and constants A_1, A_2, C and N such that

(1) $A_1 \geq f_i(t)/f_i(u) \geq A_2 > 0$,

(2) $|f_i'(t)|/f_i(u) \leq C$

for all t, u in I and $i > N$. For (1), arguments like those in the preceding paragraph show that $C_1 \geq t_i^{-1}f_i(t) \geq C_2 > 0$, for $i > N$ and t in $[a, b]$. (2) follows from Theorem 4(c), according to which f_i lies above the sinusoid with the same value and derivative at t as f_i. If $f_i'(t) \geq 0$, evaluating at b yields $r|f_i'(t)|\,|\sin k(b - t)| \leq f_i(b)$; if $f_i'(t) < 0$, evaluating at a yields the same inequality with b replaced by a. Together with (1), these inequalities give (2) for all t, u in $[a + \varepsilon, b - \varepsilon]$.

It remains to show that any limiting value of the unit vectors $U_i(s)$ as $i \to \infty$ is equal to any limiting value of the $U(s_i)$ as $s_i \to s$, $s_i \in S$. After passing to a subsequence of the γ_i, it suffices to show that the Euclidean angle between $U_i(s)$ and $U_i(s_i)$ converges to 0 as $s_i \to s^+$. But we have

$$\begin{aligned}
C &\geq f_i'(s_i)/f_i(s_i) \\
&= \left[f_i'(s) + \int_s^{s_i} f_i''(t)\,dt\right]\Big/ f_i(s_i) \\
&\geq f_i'(s)/f_i(s_i) + \int_s^{s_i} A_2[f_i''(t)/f_i(t)]\,dt \\
&\geq -C + A_2\int_s^{s_i} f_i''(t)/f_i(t)\,dt \\
&\geq -C + A_2\int_s^{s_i}[-k^2 + \|U_i'(t)\|^2]\,dt \\
&\geq -C - A_2k^2(s_i - s) + A_2(s_i - s)^{-1}\left[\int_s^{s_i}\|U_i'(t)\|\,dt\right]^2
\end{aligned}$$

The last step is by the Schwarz inequality and the preceding step is by the differential inequality. Therefore the integral, which is not less than the Euclidean angle between $U_i(s)$ and $U_i(s_i)$, approaches 0 as $s_i \to s^+$.

In the above proof we have observed that the length f of a Jacobi field satisfies the sinusoidal convexity property (b) of Theorem 4. Hence, by the corollary in Section 3, f'' exists a.e. and $-f''/f \leq k^2$. This leads us to *intrinsic* numerical curvatures of M in terms of the ratios $-f''/f$, which should be regarded as $-\infty$ at a point where f' has a positive jump. Thus, we may consider the supremum of $-f''/f$ for all Jacobi fields vanishing at p, on all geodesics through p in a neighborhood N, and then take the infimum as N shrinks to p. The resulting function agrees on the interior of M with the maximum sectional curvature function [11, p. 178], and is everywhere bounded above by $1/r^2$, where r is a tubular radius for an arbitrary Euclidean isometric imbedding of M. If, on the other hand, we take the supremum of infima on shrinking neighborhoods, we obtain the minimum sectional curvature function on the interior of M, and $-\infty$ at any boundary point at which B has positive outward normal curvature. The infinitesimal form of the Cauchy uniqueness result described in the next section is that this supremum is finite if we restrict to Jacobi fields determined by variations with endpoints on B.

6. Involutes

Involutes arise in the study of families of geodesics having the same initial conditions. Suppose N is a Riemannian extension of M with the same dimension and without boundary, and let γ, parametrized by $[0, l]$, be a geodesic. (By "geodesic" we always mean a geodesic of M.) A *lift* of γ is a C^1

curve which has the same length and initial tangent vector as γ, and which consists of an initial segment of γ (possibly trivial) followed by an N-geodesic segment. By the remark in Section 3, if a lift lies in M, it is a geodesic.

The lift endpoints trace out the *involute* curve σ of γ, namely,

$$\sigma(u) = \exp(l - u)\gamma'(u), \quad 0 \le u \le l, \tag{1}$$

where exp is the exponential map of N. Thus, σ travels from the right endpoint of the N-geodesic with the same initial tangent vector and length as γ to the right endpoint of γ. For points q in B, locally the exponentiated tangent spaces $\exp(T_qB)$ form hypersurfaces of N which we call H_q. The fact that, locally, involutes descend with respect to the height over B and over H_q gives the following result.

THEOREM 8. *Let p be a point of B. Then p has a neighborhood U in M such that for any geodesic γ in U,* (a) *the lifts of γ lie in M and no lift except possibly γ itself has its endpoint on B and* (b) *if γ is tangent to B at some q in B, then γ does not enter the inward side of H_q.*

Proof. Let σ be the involute of a geodesic γ, as in (1). Consider the variation of N-geodesics $F(u, s) = \exp(s - u)\gamma'(u)$ for $0 \le u, s \le l$. By the one-sided differentiability of γ', the N-geodesic $\gamma_a(s) = F(a, s)$ carries an N-Jacobi field $J_a^+(s) = (\partial F/\partial u)(a^+, s)$. By definition,

$$\sigma'(a^+) = J_a^+(l). \tag{2}$$

The Jacobi field J_a^+ is determined by the conditions

$$J_a^+(a) = 0, \quad (DJ_a^+/ds)(a) = D_X X(a^+), \tag{3}$$

where $X = \gamma'$ and D is covariant differentiation in N. Both equations follow from $F(t, t) = \gamma(t)$ and the chain rule. To avoid differentiability problems, we may clearly suppose that γ' is differentiable at a, so that F and $\partial F/\partial s$ are differentiable at (a, a); for the second equation, we may write $J_a^+(s) = (\partial F^{\sim}/\partial u)(a, s)$ where $F^{\sim}$ is a C^2 variation satisfying $\partial F^{\sim}/\partial u = 0$ and $\partial^2 F^{\sim}/\partial u \partial s = \partial^2 F/\partial u \partial s$ at (a, a). Similarly, if J_a^- is the N-Jacobi field along γ_a determined by equations analogous to (3), then $\sigma'(a^-) = J_a^-(l)$.

Next consider a neighborhood V of p in N on which the inward-pointing normal N-geodesics to B determine a vector field E_n. Let $W(\theta)$ consist of all nonzero tangent vectors to V which make an angle of less than θ with $-E_n$. Given any θ in $(0, \pi/2]$, there is a neighborhood U of p such that for all geodesics in U, the nonvanishing tangents of their involutes lie in $W(\theta)$. To see this, let U be a convex N-ball at p of diameter c. Choose c so that for any N-geodesic α with length less than c and initial point in U, the N-Jacobi field

J along α which satisfies at its left endpoint $J(a) = 0$ and $(DJ/ds)(a) = -E_n$ takes its right endpoint value in $W(\theta)$. (This is possible because the rate of turning of J with respect to the parallel translate of E_n is governed by curvature.) Since the acceleration vectors $D_X X(a \pm)$ of γ are either zero or positive multiples of $-E_n$, it follows from (2) and (3) that if γ lies in U, then the nonvanishing tangents of its involute lie in $W(\theta)$.

Taking $\theta = \pi/2$ gives a neighborhood U of p satisfying (a). Indeed the signed inward distance z from B is nonincreasing on the involute of any geodesic in U, and decreasing whenever the involute has nonzero derivative.

For (b), restrict the original neighborhood V so that the outward pointing normal geodesics to each H_q determine a vector field which lies in $W(\varphi)$, for some φ in $[0, \pi/2)$. Taking $\theta = \pi/2 - \varphi$ gives a neighborhood U of p satisfying (b). This is because the inward distance from any H_q is nonincreasing on the involute of any geodesic in U.

Now let $V(q, l)$ denote the set of unit tangent vectors at $q \in M$ of all N-geodesics of length l from q which lie entirely in M. If none of these N-geodesics has its right endpoint on B, then the boundary $\partial V(q, l)$ in the unit tangent $(n-1)$-sphere to N at q corresponds to the horizon visible from q along N-geodesics of length not exceeding l. In this case, the following corollary of Theorem 8(a) implies that for l sufficiently small, no geodesic γ of length l from q has an initial tangent which points below the horizon. (The figure represents the scene visible to an observer at q who looks along N-geodesics of length not exceeding l, and hence no geodesic of length l from q can pass through r. This can be interpreted as saying that a sufficiently long geodesic through r cannot bend enough along its boundary segments to avoid striking the boundary transversely beyond r.)

COROLLARY. *Let p be a point of B. For all q in a neighborhood of p in M and all l less than a positive constant, $V(q, l)$ contains the initial tangents of all geodesics of length l from q.*

THEOREM 9. *Let p be a point of B. For all q in a neighborhood U of p in B and all l less than a positive constant C, $\partial V(q, l)$ consists of the initial tangents of all geodesics of length l from q whose right endpoints lie on B.*

Proof. Let $S(q, l)$ denote the points of B at M-distance l from q. We may choose U and C so that for q in U and $l < C$ there is a well defined and

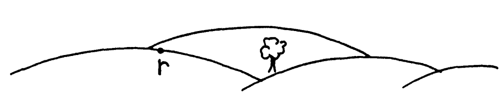

FIG. 3

continuous initial tangent map $f: S(q,l) \to V(q,l)$. Here, $f(r)$ is the initial unit tangent to the unique geodesic of length l from q to r. It is clear that f takes its values in $\partial V(q,l)$, provided we work in neighborhoods which are foliated by the N-geodesics normal to B. It suffices to show, for some choice of U and C, that there are homeomorphisms of $\partial V(q,l)$ and $S(q,l)$ with the $(n-2)$-sphere, with respect to which f has degree one.

A unit vector v in T_qB determines a half plane $\{av + bN_q : a \geq 0\}$ normal to B in T_qN. Let S_q denote these halfplanes carrying the structure of the standard $(n-2)$-sphere inherited from the unit sphere in T_qB. Let the M-ball of radius C at each q in U lie in a normal coordinate neighborhood on N, and the corresponding preimage of B under $\exp_q$ be the graph of a function over T_qB. Then the map $h_1 : \partial V(q,l) \to S_q$ for which $h_1(v)$ is the normal halfplane through v is easily seen to be a homeomorphism.

To define a homeomorphism $h_2: S_q \to S(q,l)$, consider the normal slice curves $(\exp_q P) \cap B$, for P in S_q. Since geodesics have Lipschitz continuous velocities with uniform Lipschitz constant, we may assume that for any r in $S(q,l)$, the geodesic and the normal slice curve from q to r meet at acute angles at both q and r. Define h_2 by mapping each normal halfplane in S_q to the intersection of the corresponding slice curve with $S(q,l)$. It may be assumed that h_2 is well defined because, locally, distance in M is monotonic along sufficiently short slice curves. (This fact is immediate from the C^1 regularity of the distance function; for completeness we include a brief independent proof below.) Clearly h_2 is one-one and bicontinuous, so it only remains to verify that the self-map $h_1 \circ f \circ h_2$ of S_q has degree one. Under this map, the normal halfplane P in S_q whose slice curve passes through $r \in S(q,l)$ is mapped to the normal halfplane tangent to the geodesic from q to r. By assumption, the angle between P and its image is acute. But any continuous self-map of the standard sphere which moves every point through a distance of less than π is homotopic to the identity.

Lastly we verify that M-distance from q in U is monotone on normal slice curves σ to points of $S(q,l)$, $l < C$. In effect, we shall prove a version of Gauss's Lemma; that is, curves in $S(q,t)$ must be orthogonal to radial geodesics. The only properties we use of σ are that it is C^1 and makes acute angles with radial geodesics from q. It is not hard to show that the difference between distance in M and in N is bounded by a constant times the square of the distance in N. Let γ be the shortest geodesic from q to a point r on σ. Consider points r' on σ converging to r along σ from the right; then the radial geodesics from q to r' converge to γ. By assumption the angles $rr'q$ between σ and those radial geodesics are acute, and, in fact, bounded away from $\pi/2$. Hence, there will be points q' on the segment from q to r' so that $r'rq'$ is a right triangle in N. The discrepancy between the lengths of the hypotenuse $q'r'$ and the leg $q'r$ is bounded below by a positive multiple of either of them, first in the sense of lengths in N, but then also in the sense of lengths in M because the error is quadratic. From the triangle inequality for $qq'r$ and this

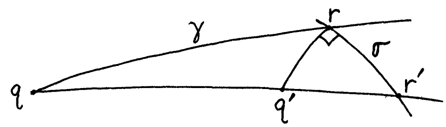

FIG. 4

hypotenuse-leg discrepancy it then follow that the distance from q to r is greater than that from q to r', which is what we wanted to show.

By Theorem 9, each vector v in $\partial V(q, l)$ is the initial tangent of at least one geodesic of length l from q to B. In a subsequent work we prove that for l sufficient small, each v corresponds to exactly one such geodesic: to be precise, that every boundary point p has a neighborhood in which two geodesics coincide if they have the same initial tangent vector and length and if their right endpoints lie on B. It follows from Theorem 9 that p has a neighborhood in which every geodesic which is somewhere tangent to B is a lift of a geodesic whose right endpoint lies on B. Therefore an equivalent formulation of our subsequent result is the following: *Cauchy uniqueness for manifolds with boundary*. Every boundary point has a neighborhood U such that for any two geodesics in U with the same initial tangent vector and length, one of the geodesics is a lift of the other.

We conclude by recalling our remark in the introduction that it is not yet clear what conditions on the boundary imply the piecewise C^2 behavior of all geodesics. Indeed, if we consider an analytic boundary in Euclidean 2-space, it is clear that there can be no intermittent points; moreover, we have established the considerably more difficult result that an analytic boundary always excludes the possibility of intermittent points.

REFERENCES

1. S. ALEXANDER, *The expontial map in Riemannian spaces with convex obstacles*, preprint.
2. RALPH ALEXANDER AND S. ALEXANDER, *Geodesics in Riemannian manifolds-with-boundary*, Indiana Univ. Math. J., vol. 30 (1981). pp. 481–488.
2A. S.B. ALEXANDER, I.D. BERG AND R.L. BISHOP, "Cauchy uniqueness in the Riemannian obstacle problem" in *Differential Geometry*, Peñíscola 1985, Lecture notes in Mathematics, vol. 1209, Springer-Verlag, Berlin, pp. 1–7.
3. F.J. ALMGREN, JR., *Existence and regularity almost everywhere of solutions to elliptic variational problems with constraints*, Mem. Amer. Math. Soc., no. 165 (1976).
4. S.S. ANTMAN, *The influence of elasticity on analysis*: *Modern developments*, Bull. Amer. Math. Soc., vol. 9 (1983), pp. 267–291.
5. V.I. ARNOL'D, *Singularities in the calculus of variations*, Itogi Nauki i Techniki, Vol. 22 (1983), pp. 3–55 (in Russian); J. Soviet Math., vol. 27 (1984), pp. 2679–2712 (in English).
6. ______, *Lagrange varieties, asymptotical rays and the open swallowtail, J. Functional Analysis, vol.* 15 (1981), *pp.* 1–14.
7. ______, *Evolution of singularities of potential flows in collision free media and caustics metamorphosal in 3-space*, Trudy Sem. im. I. G. Petrovskogo, vol. 8 (1982), pp. 21–57 (in Russian).

8. ______, *Singularities of Legendre varieties of evolvents and of fronts at an obstacle*, Ergodic Theory and Dynamic Systems, vol. 2 (1982), pp. 301–309.
9. I.D. Berg, *An estimate on the total curvature of a geodesic in Euclidean 3-space-with-boundary*, Geometriae Dedicata, vol. 13 (1982), pp. 1–6.
10. I.D. Berg and R.L. Bishop, *Local uniqueness of geodesics*, Abstracts Amer. Math. Soc., vol 4 (1983), p. 501.
11. R. L. Bishop and R. J. Crittenden, *Geometry of manifolds*, Academic Press, N. Y., 1964.
12. E. DeGiorgi, A. Marino and M. Tosques, *Problemi di evoluzione in spazi metrici e curve di massima pendenza*, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur., vol. 68 (1980), pp. 180–187.
13. A. Marino and D. Scolozzi, *Geodetiche con ostacolo*, Boll. Un. Mat. Ital., vol. (6) 2–B (1983), pp. 1–31.
14. A.D. Milka, *Shortest arcs on convex surfaces*, Soviet Math. Dokl., vol. 20 (1979), pp. 949–952.
15. ______, *Geodesics and shortest lines on convex hypersurfaces I*, Ukrain. Geometr. Sb., no. 25 (1982), pp. 95–110 (in Russian).
16. W. Rinow, *Die innere Geometrie der metrischen Räume*, Springer, Berlin, 1961.
17. D. Scolozzi, *Un risultato di locale unicità per le geodetiche su varietá con bordo*, preprint, Universitá di Pisa.
18. ______, *Un teorema di esistenza di una geodetica chiusa su varietá con bordo*, preprint, Universitá di Pisa.
19. F.-E. Wolter, *Interior metric, shortest paths and loops in Riemannian mainfolds with not necessarily smooth boundary*, Diplomarbeit, Freie Universität Berlin.
20. ______, *Cut loci in bordered and unbordered Riemannian manifolds*, doctoral dissertation, Technische Universität Berlin, 1985.

University of Illinois
Urbana, Illinois