PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 113, Number 1, September 1991

GEODESICS IN EUCLIDEAN SPACE WITH ANALYTIC OBSTACLE

FELIX ALBRECHT AND I. D. BERG

(Communicated by Jonathan M. Rosenberg)

ABSTRACT. In this note we are concerned with the behavior of geodesics in Euclidean n-space with a smooth obstacle. Our principal result is that if the obstacle is locally analytic, that is, locally of the form $x_n = f(x_1, \dots, x_{n-1})$ for a real analytic function f, then a geodesic can have, in any segment of finite arc length, only a finite number of distinct switch points, points on the boundary that bound a segment not touching the boundary.

This result is certainly false that for a C^∞ boundary. Indeed, even in E^2, where our result is obvious for analytic boundaries, we can construct a C^∞ boundary so that the closure of the set of switch points is of positive measure.

We denote by M the closure of the complement of the obstacle in Euclidean space E^n and by S the boundary of the obstacle. Thus M is an n-dimensional Riemannian manifold-with-boundary embedded in E^n and S is its boundary surface.

A geodesic on a Riemannian manifold-with-boundary, M, is defined to be a locally shortest path in M. In our context the geodesics are easy to visualize; in E^3 the geodesic is the path of a stretched string with the boundary considered as the surface of an obstacle around which the string must bend or into which the string must plunge and end. In [ABB1, ABB2] the properties of these geodesics are explored in the general setting of a Riemannian manifold-with-boundary.

We describe briefly the elementary properties of the geodesics in M.

A geodesic contacting the boundary in a segment is a geodesic of the boundary; a geodesic segment not touching the boundary is a straight line segment. A segment on the boundary joins a segment in the ambient space in a differentiable join. We call the endpoints on the boundary S of a segment not touching the boundary *switch points*. Cluster points of switch points, necessarily points at which the geodesic contacts the boundary, we call *intermittent points* or *chatter points*. As we will see, even for a C^∞ surface we can have sets of positive measure of such chatter points, and so it is reassuring to observe that the acceleration of a geodesic at a chatter point is 0. Indeed, acceleration is well defined and bounded everywhere except at the necessarily countable set of switch points; where it is defined, acceleration is normal and outward-pointing

Received by the editors November 27, 1989.
1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 53C22, 53A04.

or 0 when the geodesic is in contact, and 0 otherwise. Proof of these plausible statements is supplied in [ABB1].

V. I. Arnol'd has considered geodesics on Euclidean manifolds-with-boundary from a variational point of view [Ar]. In several works Aleksandrov and Gromov have considered generalized notions of length space, which in particular apply to Riemannian spaces with Riemannian boundaries, e.g. [Alv. G]. A variational inequalities approach such as that of Kinderlehrer and Stampacchia [K] is probably closest in spirit to the current work. In our case, however, we ask for one-dimensional minimizing manifolds, which in dimension three or more distinguishes our problem from theirs. In dimension two, where our problem is trivial, it represents a particularly accessible case of the situation discussed in Chapter V, Free boundary problems and the coincidence set of the solution, of [K]. These approaches are mentioned to give the reader some references for background and are not necessary for our particular problem.

Our theorem on analytic surfaces was announced in [ABB1]; however, both [ABB1] and [ABB2] showed that even for the C^∞ case in the presence of positive measure sets of chatter points a reasonable differential analysis of geodesics in manifolds-with-boundary was attainable. This current work is part of a general investigation of geodesics on manifolds-with-boundary conducted in concert with Stephanie B. Alexander and Richard L. Bishop whose contributions we gratefully acknowledge.

It is clear that a C^∞ boundary, as opposed to an analytic boundary, can allow chattering behavior, and even for $n = 2$ we can achieve a Cantor set of intermittent points of positive measure. Further, for $n = 2$, although it can be seen that second and higher order derivatives of the boundary curve and hence second derivatives of the geodesic γ are 0 at an intermittent point, we can arrange the set of positive measure so that any closed segment of γ that contacts S at an interior point of the segment lies properly above the secant connecting the endpoints of the segment. That is, γ genuinely bends when it touches the boundary.

We describe this construction: We define our boundary as the graph of a C^∞ function h obtained from its second derivative h''. We form our Cantor set of positive measure by extracting a sequence of open intervals $\{I_n\}$ from the unit interval. On $I_n = (a_n, b_n)$ we choose h'' to be a C^∞ function such that h'' and all its derivatives are 0 at a_n and b_n. We further choose h'' to be negative, 0, positive, 0, negative on successive nondegenerate subintervals of I_n and require both $\int_{I_n} h'' < 0$ and that the convex hull of h over I_n coincide with h on beginning and ending subintervals of I_n. We define h'' to be 0 on the complement of $\{I_n\}$. It is clear that by scaling h'' of magnitude sufficiently rapidly decreasing on successive I_n we can insure that h is C^∞. Then the C^∞ convex hull of h coincides with our geodesic γ, and γ has the required properties.

In the preceding construction we observe that the geodesic γ is C^∞ but can easily be made C^k but not C^{k+1} by an appropriate C^∞ tailoring of the joins of the positive to the negative intervals without the intervening 0 intervals. Indeed, γ is typically C^1 but not C^2.

Our theorem on analytic boundaries is trivial when the ambient space is E^2. Indeed, it is clear that for an analytic boundary in E^2 the geodesic cannot have an accumulation of switch points, since between any two such points there must be a point of zero curvature. An infinite set of points of curvature zero, necessarily clustering at a point p, implies that all the derivatives after the first vanish at the point and hence by analyticity the boundary is a straight line.

This simple argument fails in E^n for $n > 2$ because we may have vanishing directional second derivatives, but in different directions, and so we cannot hope for so easy a proof. Indeed it is easy to construct a curve lying on an analytic surface in E^3 except for intervals where it leaves and regains the surface in a tangential straight line an infinite number of times in a bounded set. This curve, however, by our theorem cannot be a geodesic. An attempt to construct such a curve as a putative counterexample is illuminating; the same turns out to require acceleration with a tangential component, hence cannot be a geodesic.

The following lemma is immediate by virtue of the Cauchy uniqueness theorem of [ABB2]. This theorem, however, is not obvious and the particular application we need here is easy enough to invite a direct proof.

Lemma. *Let S contain a straight line segment l originating at p. Let γ be a geodesic on M originating at p with initial tangent in the direction of l. Then γ contains l.*

Proof. It is clear that l is a geodesic and so, recalling that a geodesic is only a locally shortest path, we need show that there is no other. Suppose γ' is another geodesic and γ and γ' do not coincide close to p. Assume the normal to S is vertical at p and M lies above S near p. Then, since for any ε the only nonzero acceleration applied to γ' in a small enough neighborhood of p is within ε of vertical and γ is subject to 0 acceleration, we see that γ' lies almost under γ, indeed, within a wedge of vertical halfangle at most 2ε. But since γ lies on S and the normal is within ε of vertical in a neighborhood of γ we have γ' beneath the boundary surface S, which is impossible. □

We are ready to prove that there are no intermittent points for a geodesic in E^n with an analytic obstacle.

Theorem (Absence of intermittent points). *Let $n \geq 1$ and let M be an $(n+2)$-dimensional analytic manifold-with-boundary embedded in E^{n+2} and equipped with the induced Riemannian structure. Denote the boundary surface of M by S and let γ be a geodesic on M parametrized by arc length s, with $\gamma(0) = p \in S$. Then there exists an $\varepsilon > 0$ such that γ has no switch point for $0 < s < \varepsilon$.*

Proof. Consider a coordinate system $(x, y_1, \dots, y_n, z) = (x, y, z)$ in $E^{n+2}(x, z \in E^1,\ y \in E^n)$ with the origin at p, the z-axis normal to S at p, and the x-axis tangent to the given geodesic at p. With respect to this coordinate system, S is defined near the origin by an analytic equation of the form $z = z(x, y)$. Observe that the (x, z)-plane is the osculating plane of γ at p, as defined in [AA]. Notice also that we can assume, without loss of generality, that $z(x, 0)$ is not identically zero (otherwise our lemma implies that the x-axis lies in the image of γ). Thus the equation defining S near p is of the form

$$z = x^N a(x, y) + x \sum_{i=1}^{n} y_i b_i(x, y) + \sum_{j,k\geq 1} y_j y_k c_{jk}(x, y), \tag{1}$$

where $N \geq 2$, the functions a, b_i, c_{jk} are analytic and $a(0, 0) \neq 0$. Choose the orientation of the coordinate system so that $a(0, 0) > 0$ and $\gamma'(0) = \partial/\partial x$.

Let us briefly sketch the scheme for E^3 before we present the proof. We can see that along the osculating plane at the initial point, p, of the geodesic, γ, the second derivative must be approximately $a_m x^m$ for some nonzero a_m where $m \geq 0$. Because the normal is vertical at p the transverse component of acceleration is so small compared with $a_m x^m$ that we are able to show that the directional second derivative along the osculating plane to γ, which plane is slightly rotated around the vertical axis but yet more slightly tilted, cannot change sign. Hence a geodesic leaving the boundary cannot rejoin the boundary nearby.

The proof of the theorem consists of several steps. For $y \in E^n$ we write $|y| = \max_{1\leq i\leq n} |y_i|$.

Step 1. There exist $\eta > 0$, $L > 0$ such that for every sufficiently small neighborhood U of $(0, 0)$ in the $(n+1)$-dimensional (x, y)-plane we have

$$z_{xx}(x_0, y_0) \geq L x_0^{N-2}$$

for all $(x_0, y_0) \in U$ satisfying $x_0 \geq 0$, $|y_0| \leq \eta x_0^{N-2}$.

Proof of Step 1. For U convex and sufficiently small and $(x, y) \in U$, one has

$$z_{xx}(x, y) = \int_0^1 \frac{d}{dt} z_{xx}(x, ty)\, dt + z_{xx}(x, 0) = \sum_{i=1}^{n} y_i H_i(x, y) + x^{N-2} r(x),$$

where the H_i are bounded on U and $r(0) > 0$. The desired inequality follows immediately. □

Let (x_0, y_0) be a point in the (x, y)-hyperplane close to the origin and let $T \in \mathbb{R}^n$. Consider the two-dimensional plane $y - y_0 = T(x - x_0)$ and its intersection with the surface $z = z(x, y)$. Set

$$f(x) = z(x, y_0 + T(x - x_0)).$$

Step 2. There exist $\eta > 0$, $\lambda > 0$, $\mu > 0$ such that for every sufficiently small neighborhood U of $(0, 0)$ in the (x, y)-hyperplane

$$(d^2 f/dx^2)(x_0) \geq \mu x_0^{N-2}$$

for all $(x_0, y_0) \in U$ and $T \in \mathbb{R}^n$ satisfying $x_0 \geq 0$, $|y_0| \leq \eta x_0^{N-2}$, and $|T| \leq \lambda x_0^{N-2}$.

Proof of Step 2. One has

$$\frac{d^2 f}{dx^2}(x_0) = z_{xx}(x_0, y_0) + 2\sum_{i=1}^{n} T_i z_{xy_i}(x_0, y_0) + \sum_{j,k\geq 1} T_j T_k z_{y_j y_k}(x_0, y_0).$$

Let η and L be the constants obtained in Step 1. Our assumptions imply that

$$\frac{d^2 f}{dx^2}(x_0) \geq L x_0^{N-2} - 2\lambda x_0^{N-2} \sup_{(x,y)\in U} \sum_{i=1}^{n} |z_{xy_i}(x, y)| - \lambda^2 x_0^{2N-4} \sup_{(x,y)\in U} \sum_{j,k\geq 1} |z_{y_j y_k}(x, y)|$$

and the desired inequality holds for λ and U sufficiently small. □

Consider a geodesic segment $\{\gamma(s) | 0 \leq s \leq \varepsilon\}$ and its projection Γ onto the (x, y)-hyperplane. For ε sufficiently small, Γ is defined by an equation of the form $y = \alpha(x)$, with $\alpha(x) = (\alpha_1(x), \dots, \alpha_n(x))$.

Step 3. Let η, λ, μ be the constants defined in Step 2. If ε is sufficiently small, then every point $(x_0, y_0) \in \Gamma$ satisfies

$$|y_0| \leq \eta x_0^{N-2}, \qquad |(d\alpha/dx)(x_0)| \leq \lambda x_0^{N-2}. \tag{2}$$

Proof of Step 3. Let

$$\gamma(s) = x(s)\frac{\partial}{\partial x} + \sum_{i=1}^{n} y_i(s)\frac{\partial}{\partial y_i} + z(s)\frac{\partial}{\partial z} \in S$$

for some sufficiently small $s > 0$. The vector

$$N(s) = z_x(x(s), y(s))\frac{\partial}{\partial x} + \sum_{i=1}^{n} z_{y_i}(x(s), y(s))\frac{\partial}{\partial y_i} - \frac{\partial}{\partial z}$$

is normal to S at $\gamma(s)$. From (1) it follows that $N(s)$ is of the form

$$N(s) = \left[x(s)g(x(s), y(s)) + \sum_{i=1}^{n} y_i(s)h_i(x(s), y(s))\right]\frac{\partial}{\partial x} + \sum_{i=1}^{n}[x(s)k_i(x(s), y(s)) + y_i(s)l_i(x(s), y(s))]\frac{\partial}{\partial y_i} - \frac{\partial}{\partial z}.$$

Since $\alpha(0) = \frac{d\alpha}{dx}(0) = 0$ one has $y(s) = o(x(s))$ and hence $N(s) = x(s)V(s) - \partial/\partial z$, where the vector $V(s)$ is a linear combination of $\partial/\partial x$ and $\partial/\partial y_i$, $1 \leq i \leq n$, and hence is orthogonal to $\partial/\partial z$.

Now let s be such that $\gamma''(s)$ exists, with $\gamma(s)$ not necessarily in S. Then $\gamma''(s) = 0$ if $\gamma(s) \notin S$ and $\gamma''(s) = \varkappa(s)N(s)$ otherwise, where $\varkappa(s) = -z''(s)$. In this case

$$y_i''(s) = x(s)\varkappa(s)\langle V(s), \partial/\partial y_i\rangle = x(s)\varkappa(s)q_i(s),$$

where $1 \le i \le n$ and $\langle\ ,\ \rangle$ denotes the scalar product.

Set $\varkappa(s) = q_i(s) = 0$ if $\gamma(s) \notin S$ and recall that $\gamma'(0) = \partial/\partial x$. Notice that x is increasing, q_i is bounded on $[0, \varepsilon]$ for all i, and $z'(0) = z(0) = y'(0) = y(0) = 0$. Furthermore z'' (and hence also z') does not change sign on the interval $[0, \varepsilon]$. Indeed, otherwise $\gamma''(s_0)$ is directed toward the interior of M at some point $\gamma(s_0) \in S$ and therefore, the boundary segment of the geodesic containing $\gamma(s_0)$ can be shortened by cutting across the interior.

For every $s \in [0, \varepsilon]$, one gets

$$(3) \qquad |y'(s)| = \left|\int_0^s y''(\sigma)\,d\sigma\right| \le Ax(s)\int_0^s |z''(\sigma)|d\sigma = Ax(s)|z'(s)|,$$

where $A = A(\varepsilon) \to 0$ as $\varepsilon \to 0$. Integrating again one obtains

$$|y(s)| \le Ax(s)|z(s)|.$$

For $\gamma(s) \in S$ the last inequality and (1) imply $|z(s)| \le Bx^N(s) + C|z(s)|$, hence $|z(s)| \le Dx^N(s)$ for some constants B, C, D $(C < 1)$. Finally

$$(4) \qquad |y(s)| \le Ex^{N+1}(s),$$

where $E = E(\varepsilon) \to 0$ as $\varepsilon \to 0$. This proves the first inequality (2) for $\gamma(s) \in S$. For the second inequality differentiate (1) with respect to s and substitute (3) and (4) to get $|z'(s)| \le Fx^{N-1}(s)$, and hence $|y'(s)| \le Gx^N(s)$, where again $G = G(\varepsilon) \to 0$ as $\varepsilon \to 0$. It is immediate that (2) holds for points of Γ corresponding to $\gamma(s) \notin S$. □

Step 4. If ε is sufficiently small the curve Γ has at most one point that is the projection of a switch point $\gamma(s_0)$ of γ for some $s_0 \in [0, \varepsilon)$.

Proof of Step 4. Let $(x_0, y_0) \in \Gamma$ be the projection of a switchpoint $\gamma(s_0)$ of γ at which γ enters the interior of M for increasing s.

The geodesic arc beyond this point (for $s > s_0$ and $s - s_0$ sufficiently small) is a straight line segment contained in the plane

$$y - y_0 = (d\alpha/dx)(x_0)(x - x_0).$$

If this line segment meets S again then it does so at a point on the trace $z = f(x)$ of S in that plane. However, this is impossible by Steps 2 and 3. □

We should point out a natural question left unanswered. A little thought will convince the reader that for $m \ge 2$ we can construct an m-dimensional manifold-with-boundary M in E^m with analytic boundary surface S and a point $p \in S$ such that for any $\varepsilon > 0$, there is a geodesic starting at p that leaves S and rejoins S in an ε neighborhood of p. Indeed for any n we can construct an analytic surface S and a point p thereon such that for any ε there exists a geodesic starting at p with n distinct intervals in the ambient space in an ε neighborhood of p. We conjecture that for a fixed surface S and fixed point p this n is bounded. That is, we have shown that for analytic S no geodesic has an infinite number of intervals in the ambient space near p;

we conjecture that there is a uniform bound, depending on the lowest degree terms in the Taylor expansion of f, on the number of such intervals. □

REFERENCES

[AA] R. Alexander and S. B. Alexander, *Geodesics in Riemannian manifolds—with—boundary*, Indiana Univ. Math J. **30** (1981), 481–488.

[ABB1] S. B. Alexander, I. D. Berg, and R. L. Bishop, *The Riemannian obstacle problem*, Illinois J. Math. **31** (1987), 167–184.

[ABB2] ——, *Cauchy uniqueness in the Riemannian obstacle problem*, Lecture Notes in Math., vol. 1209, Diff. Geom. Peñiscola, 1985, Springer-Verlag, New York, pp. 1–7.

[Alv] A. D. Aleksandrov, *A theorem on triangles in a metric space and some of its applications*, Trudy Mat. Inst. Steklov **38** (1951), 5–23.

[Ar] V. I. Arnol'd, *Singularities in the calculus of variations*, J. Soviet Math. **27** (1984), 2679–2712.

[G] M. Gromov, *Hyperbolic manifolds, groups and actions, Riemann surfaces and related topics*, Proc. of the 1978 Stony Brook Conf., Princeton Univ. Press, Princeton, NJ, 1980, pp. 183–213.

[K] D. Kinderlehrer and G. Stampacchia, *An introduction to variational inequalities and their applications*, Academic Press, New York, 1980.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, ILLINOIS 61801