Cut Loci, Minimizers, and Wavefronts in Riemannian Manifolds with Boundary

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

This note investigates, in the setting of Riemannian manifolds with boundary, the ideas that in ordinary Riemannian manifolds fall under the heading of "cut locus". The new features, which we illustrate by examples, are due to the fact that minimizers can bifurcate and merge. Thus there may exist open sets each of whose points p has the property that every point q sufficiently far from p has more than one minimizer from p. We show (Theorem 1) that for any p and almost all q, there is a natural way to choose exactly one minimizer from p to q. This simple construction applies uniformly to all complete, connected Riemannian manifolds with boundary. In consequence, we may extend a substantial part of the classical theory to manifolds with boundary.

THEOREM 1. *Let p be a point of a complete, connected Riemannian manifold with boundary M. The set of points that have two minimizers from p with distinct terminal velocity vectors has measure zero in M. The complement of this set can be expressed as a union of "primary" minimizers displaying tree-like branching behavior.*

Throughout, M will denote a connected, metrically complete, and C^∞ Riemannian manifold with C^∞ boundary. *Geodesics* will be locally minimizing curves parameterized proportionally to arclength by $[0,1]$. Recall that geodesics are C^1 and that any two points of M are joined by a minimizer. For a general reference on geodesics in Riemannian manifolds with boundary, see [ABB1]. Cut loci in manifolds with boundary were previously investigated by Wolter in his dissertation [W2], to which we refer below.

1. Primary Minimizers

It is shown in [ABB2] that geodesics can bifurcate only by *involution,* that is, by unrolling from the boundary. The considerable difficulty of the proof is due to the fact that a geodesic may contact the boundary in, say, a Cantor-like set. We now state this theorem precisely; with a view to its use here, it is stated as a uniqueness theorem for given terminal position and velocity.

Received May 30, 1991.
Michigan Math. J. 40 (1993).

THEOREM 2 (Cauchy uniqueness for geodesics in manifolds with boundary) [ABB2]. *For a given point r on the boundary, there is a distance $b > 0$ such that if two geodesics in M of length b have the same terminal velocity vector at r and trivial common terminal segment, then one of them lies entirely (except for r) in the interior of M. For all the boundary points within a compact set, the distance b can be chosen uniformly.*

Theorem 2 allows us to select preferred minimizers. The intuition underlying the following definition is that of wavefront pinching; see Example 1 below.

DEFINITION 1. For two minimizing geodesics γ and σ from p to q, we say that γ is *preferred* to σ (or "$\gamma > \sigma$") if they have the same terminal velocity vector at q, and just before their common terminal segment (which may be trivial) γ has a nontrivial segment in the interior of M but σ does not. A *primary* minimizer from p to q is one that is preferred to all others having the same velocity vector at q.

THEOREM 3 (Existence and uniqueness of primary minimizers). *For each minimizer from p to q, there is exactly one primary minimizer from p to q having the same terminal velocity vector.*

Proof. It is easy to verify that the relation "$\geq$" is a partial ordering of the minimizers from p to q. Note that if $\gamma \geq \sigma$ and $\sigma \geq \tau$, then two possibilities may occur: the common terminal segment of γ and σ may contain that of σ and τ, or vice versa.

Theorem 2 implies that this partial order is linear on the family of minimizers from p to q having a fixed common terminal velocity. That is, for any two distinct minimizers from p to q with the same terminal velocity, one of them is preferred to the other. A primary minimizer is maximally preferred, and certainly unique because of the linearity.

To get the existence of these maximally preferred minimizers from p to q, we show more—namely, that in proceeding back from q, only a finite number of bifurcations need be considered. (There can be, however, continuous families of bifurcating minimizers; see Example 2 below.) Choose the constant b of Theorem 2 uniformly for all boundary points in the closed ball in M of radius $d(p,q)$ about p. We may also take b to be a uniform radius of bipoint uniqueness for this ball, so that any two points within distance b of each other determine a unique minimizer [ABB1].

For a given nonprimary minimizer from p to q, we obtain the primary minimizer having the same velocity at q as follows. Consider the family $\mathfrak{F}$ of all minimizers from p to q having the given velocity at q. Proceed back from q to r along the maximal common terminal segment rq of this family. We show that some minimizer in $\mathfrak{F}$ has a segment of length b in the interior of M immediately preceding r. If not, then r is not a bifurcation point, by Theorem 2, and so moving back from r on members of $\mathfrak{F}$, bifurcations must occur arbitrarily close to r. By Theorem 2, there is a sequence of minimizers γ_n in $\mathfrak{F}$ such that γ_n has a segment of length b in the interior immediately preceding a point r_n, where r_n converges to r. Thus there exists a limit minimizer γ in $\mathfrak{F}$ that has rq as a terminal segment, and has a segment of length b with identically vanishing acceleration immediately preceding r. But then Theorem 2 and the maximality of rq imply that this segment also lies in the interior. Clearly, at most $d(p,q)/b$ repetitions will produce the desired primary minimizer. □

2. Examples

The first example illustrates the intuition underlying the definition of primary minimizer, and the second, the possible existence of continuous families of minimizers from p to q having the same terminal velocity.

EXAMPLE 1. *Let M be the Euclidean plane with an open circular disk centered at the origin and a smoothed tubular neighborhood of a vertical half-line removed, as pictured in Figure 1.*

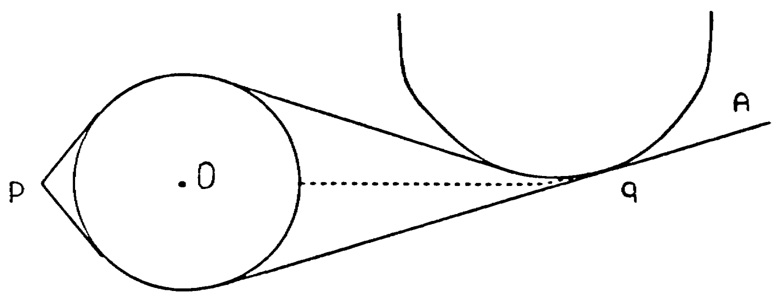

Figure 1

It is possible to position the second obstacle so that pairs of geodesics emanating from any point p on the negative x-axis and bearing on opposite sides of the disk come together tangentially and at equal lengths at a point q on the second obstacle. Extended beyond q, these geodesics bifurcate by involution to reach all points of the "shadow" region A of that obstacle. Thus, each point of A, including q, has two minimizers from p; these have the same terminal velocity and have minimizing extensions.

Running from q to the disk, there is a curve shown in dashes, each of whose points also has two minimizers from p. These have different terminal velocities and do not have minimizing extensions. Observe that the wavefront from p, that is, the distance-circle centered at p, is separated into branches by the disk; coming around either side of the disk, these two branches collide to form the dashed curve. The top branch of the wavefront is squeezed by the second obstacle and extinguished at q, leaving only the bottom branch, whose orthogonal trajectories are the primary minimizers, to sweep across A.

REMARK 1. In Example 1, p must be chosen carefully in order to obtain nonprimary minimizers from p. However, by reflecting Example 1 about the y-axis and letting A' be the reflected image of A, we obtain a space that contains an open set A' of points p, from each of which there are nonprimary minimizers. Furthermore, the property of containing nonprimary minimizers is a robust phenomenon. For instance, it is clear that any small perturbation of Example 1 contains such minimizers. Thus, if one wished to avoid working with nonprimary minimizers, the best one could hope is that—for most Riemannian manifolds with boundary M—most points p in M have the property that all minimizers from p are primary (equivalently, that distinct minimizers from p to any point have distinct terminal velocities). Our aim is, rather, to give a uniform analysis that works for all cases.

EXAMPLE 2. Let M be a standard 2-sphere with two *antipodal* small open disks removed, and let $\bar{M}$ be its universal cover. Then $\bar{M}$ has two boundary components, S and S', covering the two boundary circles of M. Choose a great semicircle in M joining two antipodal points on these two boundary circles, and consider an arc from $p \in S$ to $p' \in S'$ that covers it. Extending this arc through p' to q' by adding a nontrivial segment of S' gives a minimizing geodesic γ of M. Then γ belongs to a 1-parameter variation of minimizers from p to q', all of which have the same terminal velocity (see Figure 2). Within this 1-parameter family there exists one minimizer having a trivial segment on S', and that one is the primary minimizer from p to q'. In the figure, we view $\bar{M}$ as a strip in R^2 carrying on its interior a metric of constant positive curvature invariant under translation of the strip.

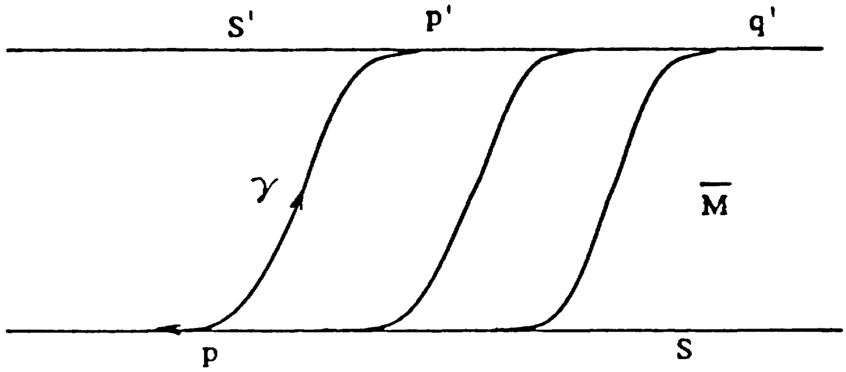

Figure 2

3. Proof of Theorem 1

We restate Theorem 1 as follows.

THEOREM 1. *For any given point p, let $M^*(p)$ consist of the points q all of whose minimizers from p have the same terminal velocity vector. Then the* *complement of $M^*(p)$ has measure zero, and $M^*(p)$ is the union of the unique primary minimizers from p to each of its points. The traces pq of these minimizers satisfy the branching condition*

$$pq_1 \cap pq_2 = pq_3 \quad \textit{for some } q_3. \tag{1}$$

Proof. It is proved in [ABB3] that $M^*(p) - \{p\}$ is precisely the subset of M on which the distance function from p is differentiable. Since the distance function is Lipschitz continuous (by the triangle inequality), it follows that its set of nondifferentiability has measure zero in M [F, 3.1.6].

By the existence and uniqueness theorem for primary minimizers (Theorem 3), $M^*(p)$ consists of those points that are joined to p by unique primary minimizers. Further, a primary minimizer γ from p lies entirely in $M^*(p)$ if its right-hand endpoint does. Indeed, if x is an internal point of γ then there cannot be two minimizers with distinct terminal velocities from p to x, because then it is easy to see (as in the classical case) that γ would not be minimizing. Thus $M^*(p)$ is expressible as the union of unique primary minimizers from p. If two of these have a common point, then—since any truncation of a unique primary minimizer from p again has this property—(1) follows. □

Our proof that $M^*(p) - \{p\}$ is the set of differentiability of the distance function d_p extended an earlier theorem of Wolter, who proved that the interior of $M^*(p) - \{p\}$ is the maximal open subset on which d_p is C^1 [W2]. Theorem 1 requires us to consider the remaining points of $M^*(p) - \{p\}$ because we do not know whether they form a set of measure zero.

4. Unique Minimizers and the Cut Locus

First we recall the classical situation, where M is a Riemannian manifold without boundary. Let G_p be the space of all geodesics from p in the uniform topology, and let G_p^* be the subspace consisting of minimizers from p that are uniquely determined by their right-hand endpoint. Then G_p^* is contractible, since a geodesic does not minimize past a point that has multiple minimizers. In addition, the right-hand endpoint map is a homeomorphism from G_p^* onto a set whose complement has measure zero in M, namely the points having unique minimizers from p. The manifold M itself is homeomorphic to the identification space under the endpoint map of the closure of G_p^* in G_p. Note that under the homeomorphism of G_p with its space of initial velocity vectors, the endpoint map becomes the exponential map and we obtain a more familiar description. However, the pathspace formulation given here is more suitable for extension to manifolds with boundary, which lack classical exponential maps.

How well does this paradigm apply if M is a manifold with boundary? As Example 1 makes clear, the set of points with unique minimizers from p need not be dense in M; indeed, the set having multiple minimizers might well contain most of M, both metrically and topologically. However, if we work instead with primary minimizers (causing no change in the classical case, where all minimizers are primary), the paradigm works rather well, as the following corollary shows. As before, G_p denotes the space of geodesics from p in the uniform topology.

COROLLARY 1. *The space G_p^* of primary minimizers from p that are uniquely determined by their right-hand endpoints is contractible. The endpoint map carries G_p^* injectively onto a subset $M^*(p)$ of M whose complement has measure zero. The identification space under the endpoint map of the closure of G_p^* in G_p is homeomorphic to M.*

Proof. The first two statements follow from Theorem 1. Since M is finitely compact, any sequence of minimizers from p of bounded length has a subsequence converging to a minimizer [Bu, p. 24]. It follows that the restriction of the endpoint map to the closure of G_p^* is closed and hence is an identification onto its image; and that the image of the closure of G_p^* is the closure of its image, namely M by Theorem 1. □

Recall that the image $M^*(p)$ of G_p^* under the endpoint map is the set of points all of whose minimizers from p have the same terminal velocity vector. In Example 1, this set consists of q and the complement of the dashed curve. Here, taking the closure of G_p^* in G_p adjoins sets homeomorphic respectively to a closed interval (consisting of minimizers to points of the dashed curve including q that lie above the disk) and a half-closed interval (consisting of minimizers to points of the dashed curve not including q that lie below the disk). Note that, unlike the classical case, topological information may be lost by projecting the pathspace G_p^* to M; that is, the primary minimizers in G_p^* may not vary continuously with their endpoints, and so the injective map from G_p^* onto $M^*(p)$ may not be a homeomorphism. In particular, $M^*(p)$ need not be contractible in the topology induced from M, only in that inherited from G_p^*.

For manifolds without boundary, the cut locus of p is the closure of the set of points having more than one minimizer from p. Equivalently, it is the locus of points to which there is a minimizer from p having no minimizing extension. The equivalence of these two definitions involves the well-known theorem that geodesics do not minimize beyond a conjugate point and, to get the implication from the second to the first, the fact that the exponential map is never injective in a neighborhood of a conjugate point [Bi; W1]. The cut locus may also be described as the closure of the image of all many-to-one identifications by the endpoint map from $\mathrm{cl}(G_p^*)$ onto M; that is, as the *glueing locus.* The cut locus has measure zero in M.

Applying the first of these approaches to manifolds with boundary, we may define the cut locus of p as the closure of the set of points that have two primary minimizers from p. This definition agrees with the usual one when the boundary is empty, and can be formulated independently of primary minimizers as follows.

DEFINITION 2. The cut locus of p is the closure of the set of points having two minimizers from p with different terminal velocities.

Earlier, Wolter analyzed the relationships between Definition 2 and several variants [W2], all of which have similar behavior with regard to the properties and examples to be discussed below. Included are formulations in terms of nonminimizing extensions; these seem a little less natural in manifolds with boundary than in the classical case because a geodesic can end at the boundary, either transversely or tangentially. However, in the interior of M the cut locus as defined above coincides with the *closure* of the set of points that have a minimizer from p with no minimizing extension [W2]. Note that there certainly may be cut points that lie on extendible minimizers, for instance the point q in Example 1. Such cut points may even lie in the interior of M, as Example 3 below shows; this answers a question raised in [W2, p. 49].

In Example 1, the cut locus of p is the dashed curve, including q; this agrees with the glueing locus. Its complement is contractible, although not contractible along minimizers from p because these may pass through q. However, we shall construct an example in the next section (Example 3) for which the cut locus as defined above does not even have simply connected complement, and does not agree with the glueing locus. We know of no example where the complement fails to be connected. It is not known whether the cut locus must have measure zero, or even have empty interior.

5. Wavefronts

Now we investigate the ways in which the cut locus of Definition 2 diverges from the classical case. In particular, a key classical feature is the contractibility of the cut locus complement. Our analogue for manifolds with boundary is provided by the contractibility of the pathspace G_p^* (Corollary 1). Example 3 below shows that this analogy cannot in general be pushed further: specifically, that the complement of the cut locus need not be simply connected. Similarly, in the classical case the cut locus of p may be regarded as the locus of points near which the wavefront from p develops self-interference. Example 3 shows that in general the cut locus may be too small to satisfy this interpretation, in the sense that there may fail to be a continuous assignment of minimizers to its complement.

EXAMPLE 3. We start by rotating Example 1 about a vertical axis through p, to obtain a solid torus obstacle and a second, unbounded obstacle. We complete the example by inserting a ball blocking the lower path to q. This interferes with part of the lower branch of the wavefront between the torus and the unbounded obstacle. The vertical half-plane cross-sections that do not intersect the ball will be just as in Figure 1, while the others will require detours of the lower geodesics about the ball, as pictured in Figure 3.

In these latter cross-sections the region A is swept out by the upper branch of the wavefront, which collides with the lower branch along cut points that

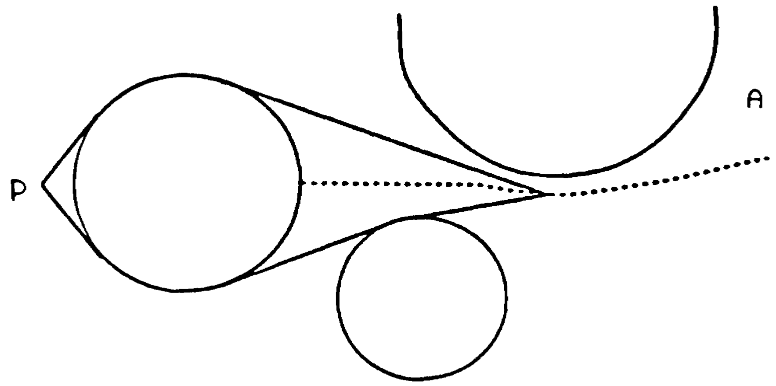

Figure 3

pass below the unbounded obstacle and extend beyond it. The lower geodesics will not necessarily be in vertical planes, but they are nevertheless elongated by the presence of the ball, so that the upper paths provide unique minimizers to points of A. The detour is illustrated in Figure 3 for the section through the center of the ball; the elongated geodesics are in this section because of symmetry. Note that there is a noncontractible loop encircling the cut points that extend beyond and below the unbounded obstacle in all sections intersecting the ball.

Now consider the sections next to the two vertical half-planes that are tangent to the ball. Figure 1 shows the pattern of the geodesics in such a section. On the side of this section away from the ball, region A is swept out by the lower branch of the wavefront, as in Example 1. On the other side, for which the lower geodesics are elongated by the presence of the ball, region A is swept out by the upper branch of the wavefront. In the transition section, these two branches of wavefront have a sheet of degenerate collision, that is, the minimizing geodesics which are their orthogonal trajectories coincide on the region A. Because the collision is degenerate, this transition region A is not included in the cut locus of Definition 2. However, in order to obtain a continuous map from the cut locus complement to the corresponding primary minimizers from p, one would have to enlarge the cut locus by adding the region A in each of the two transition sections. Similarly, it is clear that the cut locus would have to be enlarged substantially in order to achieve any continuous assignment whatever of minimizers from p to the points of the cut locus complement.

In general, continuity of the primary minimizers is achieved on the complement of the *glueing locus,* defined above as the closure of the image of all many-to-one identifications by the endpoint map from $\mathrm{cl}(G_p^*)$ onto M. The following proposition is easily verified.

PROPOSITION. *The complement of the glueing locus of p is the maximal open set on which the primary minimizers from p are single-valued and continuous.*

By omitting the words "and continuous" here, one obtains the complement of the cut locus instead. Example 3 shows that the two loci need not agree. We remark that in flat surfaces with boundary, the two may be shown to coincide.

In general, the glueing locus need not have measure zero. For instance, by a construction similar to that of Example 3 one can produce a Cantor set of sheets of degenerate wavefront collision, forming a set of positive measure. Moreover, the complement of the glueing locus need not be a connected set. This may be seen in Example 3 by moving an appropriately hollowed-out obstacle into position just tangent to one of the sheets of degenerate collision. It is an open question whether the complement must be contractible in manifolds with boundary for which the cut locus and the glueing locus agree.

In a general manifold with boundary, one might still ask whether there always exists a contractible dense subset S and a map from S to the space of minimizers from p that is continuous in the topology of the intrinsic metric on S. Examples similar to those mentioned above show that such a formulation is not obvious.

References

[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,* Differential geometry (Peñíscola, 1985), pp. 1–8, Lecture Notes in Math. 1209, Springer, Berlin, 1986.

[ABB3] ———, *Geometric curvature bounds in Riemannian manifolds with boundary,* Trans. Amer. Math. Soc. (to appear).

[Bi] R. L. Bishop, *Decomposition of cut loci,* Proc. Amer. Math. Soc. 65 (1977), 133–136.

[Bu] H. Busemann, *The geometry of geodesics,* Academic Press, New York, 1955.

[F] H. Federer, *Geometric measure theory,* Springer, Berlin, 1969.

[W1] F.-E. Wolter, *Distance function and cut loci on a complete Riemannian manifold,* Arch. Math. (Basel) 32 (1979), 92–96.

[W2] ———, *Cut loci in bordered and unbordered Riemannian manifolds,* Dissertation 249S, Technische Universität Berlin, 1985.

Department of Mathematics
University of Illinois
Urbana, IL 61801