

METRIC SPACES IN WHICH MINIMAL CIRCUITS CANNOT SELF-INTERSECT

DAVID SANDERS

ABSTRACT. Definitions are given for self-intersecting polygons and geodesic points in terms of betweenness, and then it is proved that the metric spaces in which minimal polygons on a finite number of distinct noncogeodesic points are not self-intersecting are completely characterized as those metric spaces which have the following betweenness property for any four distinct points: if b is between a and c and between a and d then either c is between a and d or d is between a and c .

1. Introduction. Minimal length circuits through a finite number of given noncollinear points in a euclidean plane cannot be self-intersecting. A proof of this theorem and also of the corresponding statement for points on a two dimensional euclidean sphere appears in [2]. This paper raises and gives an answer (cf. Theorem 2) to the question of how far this result can be generalized.

The first step is to determine a context in which the statement makes sense. A natural choice is *geodesic metric spaces*, i.e., metric spaces such that each pair of points is joined by an arc of length equal to the distance between the points. See [1] for a construction of examples of geodesic metric spaces.

However, the context for the problem of self-intersection of minimal circuits can be broadened to any metric space (M, d) by defining intersection in terms of betweenness (pq intersects rs if and only if there is some x between p and q and also between r and s), and relating betweenness to distance in the usual way (x is between p and q if and only if $d(p, x) + d(x, q) = d(p, q)$). The definition of cogeodesic points (i.e. points contained in some geodesic) can also be expressed in terms of betweenness so that the context of geodesic metric spaces is subsumed (see Theorem 3 and its proof). The essential condition for non-self-intersection of minimal circuits is a natural property of betweenness (if b is between a and c , and also between a and $d \neq c$, then either c is between a and d or d is between c and a).

Also note an extension of the context of the problem from geodesic to arbitrary metric spaces cannot be achieved by embedding a given metric space in a geodesic metric space (as can always be done) because whether or not a minimal circuit will self-intersect will depend on the embedding. For instance a metric space consisting of three points at unit distance from each

Received by the editors December 11, 1974 and, in revised form, August 29, 1975.

AMS (MOS) subject classifications (1970). Primary 05C35; Secondary 54E35.

Key words and phrases. Non-self-intersecting polygon, noncogeodesic points, geodesic metric space.

© American Mathematical Society 1976

other embeds in the euclidean plane so the minimal circuit on the three points is not self-intersecting. But when this three point metric space is embedded in a space consisting of three segments sharing a common endpoint with distances measured along the segments, then the minimal circuit on the three points must self-intersect.

2. Definitions and results. The distance between two points p and q will be denoted by \overline{pq} , and $(q_1 q_2 \cdots q_n)$ will mean that the points q_1, q_2, \dots, q_n are distinct, $n \geq 3$, and $\sum_{k=1}^{n-1} \overline{q_k q_{k+1}} = \overline{q_1 q_n}$. The term *polygon* will be used for orderings of $n \geq 3$ points (identified under cyclic permutations and their inverses) and will be denoted by $[p_1 p_2 \cdots p_n]$. Adjacent pairs of points in the ordering, i.e. $p_1 p_2, \dots, p_{n-1} p_n, p_n p_1$ will be called *edges* of the polygon. The term *circuit* will be reserved for any of the actual routes (unions of geodesics) that correspond to a particular polygon in a geodesic metric space.

In a polygon $[\dots ab \dots cd \dots]$ in a metric space M edge ab is said to *intersect* edge cd if there exists $x \in M$ so (axb) and (cxd) , or if (cad) , or (cbd) , or (acb) , or (adb) . Also adjacent edges ab, bc in a polygon $[\dots abc \dots]$ will be said to *intersect* if there exists $x \in M$ so that (axb) and (bxc) , or if (acb) or (bac) . A polygon in a metric space then is *non-self-intersecting* in that space if no pair of its edges intersect. Finally the distinct points q_1, \dots, q_n will be called *cogeodesic* if there is some ordering of them such that $(q_{i_1} \cdots q_{i_n})$.

THEOREM 1. *Given distinct noncogeodesic points p_1, \dots, p_n in a metric space M satisfying for any $x \in M$ and $j \neq k$ $(p_i x p_j)$ and $(p_i x p_k)$ imply $(p_i p_j p_k)$ or $(p_i p_k p_j)$, then any minimal polygon on p_1, \dots, p_n is not self-intersecting.*

The proof of this theorem is given in §3. Now the strength of the theorem and its consequences are discussed. First note that the converse to this theorem does not hold because the hypothesis really guarantees more than is concluded, viz., that a minimal polygon on any subset of the given noncogeodesic points is not self-intersecting. Thus in any example (such as the four corners of a square considered in a metric space consisting of the four sides of the square and an inscribed octagon with distance measured along the segments) in which the minimal polygon is not self-intersecting but a minimal polygon on a subset of the given points is self-intersecting, the converse to Theorem 1 must fail. The appropriate biconditional is formulated in the next theorem.

THEOREM 2. *In a metric space M , (abc) and (abd) , $c \neq d$ imply (acd) or (adc) for any four points a, b, c, d in M , if and only if any minimal polygon on a finite number of distinct noncogeodesic points is not self-intersecting.*

PROOF. The only if part follows from Theorem 1. To establish the converse assume for some distinct four points (abc) , (abd) , not (acd) , and not (adc) . Then also not (cad) because otherwise (cad) , (abc) , (abd) give $\overline{cd} = \overline{ca} + \overline{ad} = \overline{cb} + \overline{ba} + \overline{ab} + \overline{bd} \geq \overline{cd} + 2\overline{ab}$ contradicting $a \neq b$. However not (acd) , not (adc) and not (cad) means a, c, d are not cogeodesic and the minimal, in fact only, polygon on these three points has an intersection of edges ac and ad at b .

The specialization of Theorem 1 to the context of geodesic metric spaces is expressed in the last theorem.

THEOREM 3. *In a geodesic metric space, given distinct noncogeodesic points p_1, \dots, p_n such that if E is any geodesic joining p_i and p_j and E' is any geodesic joining p_j and p_k , then either $E \cap E' = \{p_j\}$ or p_i, p_j, p_k are cogeodesic; then any minimal circuit on the given n points cannot be self-intersecting.*

PROOF. This is a consequence of Theorem 1, the argument given in the proof of Theorem 2 that (abc) and (abd) with $c \neq d$ imply not (cad) , and the following two facts. First, in a geodesic metric space if (abc) then there exists a geodesic from a to c containing b . This follows from a slight modification of the argument used to verify that the triangle inequality is satisfied by $d(p, q) = \text{length of any minimal length arc joining } p \text{ to } q$ in spaces having such geodesics, cf. [1, §III]. The second fact is that if a polygon is not self-intersecting, then none of the circuits corresponding to it are self-intersecting. Though it is not required in the proof of Theorem 3, note that if a polygon is self-intersecting, some, but not necessarily all, corresponding circuits self-intersect.

3. Proof of Theorem 1. The contrapositive will be demonstrated; i.e., assuming that a polygon Q on the vertices p_1, \dots, p_n is self-intersecting, show that there is another polygon P on p_1, \dots, p_n of strictly shorter length. Three cases are considered:

Case 1. Suppose $Q = [\dots ab \dots cd \dots]$ has an intersection of edges ab and cd where a, b, c, d are noncogeodesic. Let $P = [\dots a \langle b \dots c \rangle d \dots]$ where this denotes the polygon resulting from Q when the order of the vertices from b to c is reversed while the order of the other vertices is unchanged. That length $Q > \text{length } P$ will follow from $\overline{ab} + \overline{cd} > \overline{ac} + \overline{bd}$, which is established in the following two subcases.

Case 1 (i). The intersection point x of ab and cd is distinct from a, b, c, d so that (axb) and (cxd) . Then $\overline{ab} + \overline{cd} = \overline{ax} + \overline{xb} + \overline{cx} + \overline{xd} \geq \overline{ac} + \overline{bd}$. If this inequality were actually equality then (axc) and (bxd) . But then (axc) and (axb) by hypothesis give (abc) or (acb) and, therefore $(axbc)$ or $(axcb)$. Similarly from (bxd) and (cxd) follow $(bcxd)$ or $(cbxd)$. It is impossible to have both $(axbc)$ and $(bcxd)$ since (xbc) and (bcx) are incompatible. Similarly $(axcb)$ and $(cbxd)$ cannot hold simultaneously. But $(axbc)$ and $(cbxd)$ imply (axc) and (cxd) which imply, using the theorem hypothesis, $(adxc)$ or $(daxc)$ which leads to $(adxbc)$ or $(daxbc)$ respectively, contradicting the initial assumption for Case 1. Similarly $(axcb)$ and $(bcxd)$ yield a, b, c, d cogeodesic.

Case 1 (ii). The intersection point of ab and cd is one of the four points a, b, c, d ; say a to be specific so (cad) . Then $\overline{ab} + \overline{cd} = \overline{ab} + \overline{ca} + \overline{ad} \geq \overline{ca} + \overline{bd}$. However, again the inequality must be strict, because otherwise (bad) with (cad) gives (bcd) or (cbd) which implies $(bcad)$ or $(cbad)$ again contradicting a, b, c, d noncogeodesic.

Case 2. Suppose $Q = [\dots ab \dots cd \dots]$ has an intersection of edges ab and cd where a, b, c, d are cogeodesic. By cyclically permuting the ordering of Q , the possible cogeodesic orderings of a, b, c, d can be assumed to be exhausted by the following three subcases.

Case 2 (i). $(abcd)$ or $(abdc)$. This case cannot occur because each of the two possible orderings is incompatible with the assumption that ab and cd intersect. For instance if $(abcd)$, whether the intersection point x of ab and cd is distinct from a, b, c, d or not, then

$$\overline{ad} = \overline{ab} + \overline{bc} + \overline{cd} = \overline{ax} + \overline{xb} + \overline{bc} + \overline{cx} + \overline{xd} > \overline{ad} + \overline{bx} + \overline{bc} + \overline{cx}$$

which implies $\overline{bc} = 0$ contradicting $b \neq c$.

Case 2 (ii). $(acbd)$ or $(acdb)$. Then $P = [\dots a \langle b \dots c \rangle d \dots]$ has length strictly less than Q , because if $(acbd)$ then (acb) and (cbd) and it follows: $\overline{ab} + \overline{cd} = \overline{ac} + \overline{cb} + \overline{cb} + \overline{bd} > \overline{ac} + \overline{bd}$. Similarly for $(acdb)$.

Case 2 (iii). $(adbc)$ or $(adcb)$. Discussion of this situation is postponed until the one remaining possibility is examined.

Case 3. Suppose $Q = [\dots abc \dots]$ has an intersection point $x \neq b$ on edges ab and bc . If $x \neq a$ and $x \neq c$ then (axb) and (bxc) which imply (acb) or (bac) . If $x = a$ then (bac) and if $x = c$ then (acb) . Since by assumption all the vertices of Q are not cogeodesic there must be another vertex which can be indicated by $Q = [\dots abcv \dots]$ if (acb) or by $Q = [\dots uabc \dots]$ if (bac) . In the first situation edges ab and cv intersect in c ; in the second situation edges ua and bc meet in a . Assuming, as may be done, that all intersections considered in Cases 1 and 2(ii) have been already eliminated the intersection just produced must belong in Case 2(iii). Thus all that remains is to return and argue that case.

Suppose then that all $Q = [\dots r_i abq_1 \dots q_s cdr_1 \dots]$, that all intersections of the types considered in Cases 1 and 2(ii) have been previously removed, and that ab and cd intersect as in Case 2(iii)—say $(adbc)$. Then bq_1 and cd intersect in b and since only intersections of type 2(iii) remain, either (dbq_1c) or $(dbcq_1)$. If (dbq_1c) then q_1q_2 intersects cd at q_1 and therefore either (dq_1q_2c) or (dq_1cq_2) . If, say the former, then cd meets q_2q_3 in q_2 and either (dq_2q_3c) or (dq_2cq_3) . Assuming the latter, q_2q_3 and $q_s c$ intersect in c giving $(q_2cq_3q_s)$ or $(q_2cq_3q_s)$. Continuing in this manner until the q 's are exhausted, regardless of which choice holds at each step, an order of the vertices is established like $adbq_1q_2cq_3q_s \dots q_v$. Also running the same procedure through the r 's produces an ordering of all the vertices, which might look like

$$[r_u \dots r_2 r_{t-1} r_t a r_1 dbq_1 q_2 cq_3 q_s q_{s-1} q_{s-2} \dots q_v].$$

Call this polygon P . Then by the construction length

$$P = \sum_{k=u}^{t-1} \overline{r_k r_{k+1}} + \overline{r_t a} + \overline{ab} + \overline{bq_1} + \sum_{k=1}^{v-1} \overline{q_k q_{k+1}} + \overline{q_v r_u}.$$

Comparing with length Q shows length $P < \text{length } Q$ if and only if

$$\overline{q_v r_u} < \sum_{k=v}^{s-1} \overline{q_k q_{k+1}} + \overline{q_s c} + \overline{cd} + \overline{dr_1} + \sum_{k=1}^{u-1} \overline{r_k r_{k+1}}.$$

But again by the construction of P , equality could hold here only if all the vertices were cogeodesic between q_v and r_u . This completes the proof.

REFERENCES

1. K. Menger, *Untersuchungen über allgemeine Metrik*, Math. Ann. **103** (1930), 466–501.
2. L. V. Quintas and F. Supnick, *On some properties of shortest Hamiltonian circuits*, Amer. Math. Monthly **72** (1965), 977–980. MR **32** #6304.

DEPARTMENT OF MATHEMATICS, BROOKLYN COLLEGE (CUNY), BROOKLYN, NEW YORK 11210

Current address: Actuarial Division, Metropolitan Life Insurance Company, New York, New York 10010