

F. S. Cater, Department of Mathematics, Portland State University,
Portland, Oregon 97207, USA

ON THE CURVES OF MINIMAL LENGTH ON SPHERES IN REAL HILBERT SPACES

Abstract

In this note, we prove that the shortest curve joining two points on a sphere in a real Hilbert space of dimension greater than 2, lies in the great circle through these points. Our arguments will apply to all continuous curves on the sphere, not just to those that have tangent and principal normal vectors and curvature at each point.

1 Introduction and Notation

While constructing a real analysis proof that the shortest continuous curve joining points B and C on a sphere lies in the great circle through B and C , we noticed that our arguments work for any sphere in any real Hilbert space of dimension greater than 2. Our arguments will apply to all continuous curves on the sphere – we do not restrict ourselves to curves with tangent or principal normal vectors or curvature at each point (compare with [M], for example). Our arguments should be conceptually easy.

First we need to discuss our notation. In all that follows h is a real Hilbert space of dimension greater than 2 and S is the unit sphere of h ,

$$S = \{V \in h : \|V\| = 1\}.$$

If U and V are points (vectors) in h , define the segment UV to be the set

$$\{tV + (1-t)U : 0 \leq t \leq 1\}.$$

The length of the segment, $|UV|$, is $\|U - V\|$. We take the usual definition of the length of a continuous curve $X(t)$, $0 \leq t \leq 1$, to be the least upper bound of lengths of polygonal paths (see [A, 6.31]). We say that U and V are

Key Words: sphere, length of curve, Hilbert space.
Mathematical Reviews subject classification: 26B15, 26B30.
Received by the editors November 10, 1999

antipodal points if $U + V = 0$. If B and C are non-antipodal points on S , by the great circle through B and C we mean the set

$$\{(bB + cC)/\|bB + cC\| : b, c \text{ real, } b^2 + c^2 > 0\}.$$

This is essentially a circle on S with radius 1 and center at the point (vector) 0. By the arc \widehat{BC} we mean the range of the curve

$$X(t) = (tC + (1-t)B)/\|tC + (1-t)B\|, \quad 0 \leq t \leq 1,$$

and by the length of the arc, $|\widehat{BC}|$, we mean the length of this curve. Clearly \widehat{BC} lies in the great circle through B and C . In the unit ball of h , central angles will be measured in radians determined in the usual way by their subtended arcs. The symbol \angle of course means angle.

2 Elementary Geometry

In this section we prove the geometric lemmas we need to complete the work. Recall that a 2-dimensional (3-dimensional) subspace of h is identical to Euclidean 2-space (3-space). There are three lemmas.

Lemma 2.1. *Let \widehat{AB} , \widehat{CD} and \widehat{EF} be arcs on S such that*

$$\angle AOB + \angle COD > \angle EOF,$$

and each of these angles is less than π . Then $|\widehat{AB}| + |\widehat{CD}| > |\widehat{EF}|$.

PROOF. Without loss of generality we assume $|\widehat{AB}| + |\widehat{CD}| < \pi$. Choose point G on the great circle through A and B such that B lies in \widehat{AG} and $\angle BOG = \angle COD$. Then

$$\angle AOG = \angle AOB + \angle BOG = \angle AOB + \angle COD > \angle EOF.$$

It follows that $|\widehat{AG}| > |\widehat{EF}|$. But $|\widehat{BG}| = |\widehat{CD}|$ and

$$|\widehat{AB}| + |\widehat{CD}| = |\widehat{AB}| + |\widehat{BG}| = |\widehat{AG}| > |\widehat{EF}|. \quad \square$$

Our next result is sort of a triangle inequality for arcs.

Lemma 2.2. *Let B and C be non-antipodal points on S and let A be a point on S that is not on the great circle through B and C . Then*

$$|\widehat{BA}| + |\widehat{AC}| > |\widehat{BC}|.$$

PROOF. In view of Lemma 2.1, it suffices to prove

$$\angle BOA + \angle AOC > \angle BOC.$$

Without loss of generality we assume that $\angle AOC < \angle BOC$. We can (and do) select a point D on the segment BC such that $\angle AOC = \angle DOC$. (See Figure 1.) Select point A' on OA such that $|OA'| = |OD|$. Draw segments CA' , BA' and OD . Now $|OA'| = |OD|$, $|OC| = |OC|$, $\angle A'OC = \angle DOC$. So $A'OC$ and DOC are congruent triangles (side, angle, side). Hence $|A'C| = |CD|$ (corresponding parts of congruent triangles).

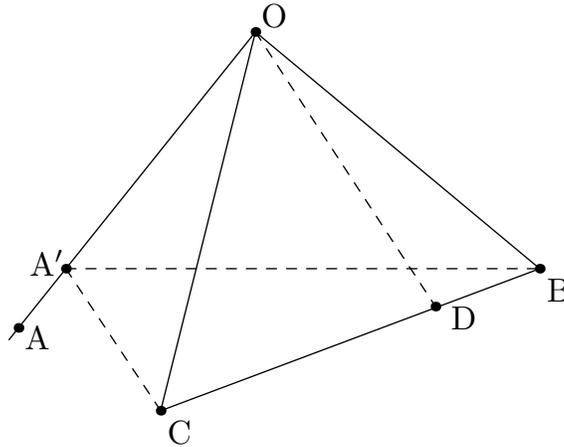


Figure 1:

The point A' is not on the line through B and C ; otherwise A would lie on the great circle through B and C . In the triangle $A'BC$, the sum of the lengths of two sides is greater than the length of the third:

$$|BA'| + |A'C| > |BC| = |BD| + |DC|.$$

But $|A'C| = |DC|$ so $|BA'| > |BD|$.

We apply the law of cosines to the triangles BOC and BOA' :

$$|BD|^2 = |BO|^2 + |DO|^2 - 2|BO| \cdot |DO| \cos \angle BOD,$$

$$|BA'|^2 = |BO|^2 + |A'O|^2 - 2|BO| \cdot |A'O| \cos \angle BOA'.$$

But $|BD|^2 < |BA'|^2$, $|DO| = |A'O|$, and it follows that $\cos \angle BOA < \cos \angle BOD$.

Because the cosine decreases from 0 to π , it follows that $\angle BOA > \angle BOD$.

Finally, $\angle BOC = \angle BOD + \angle DOC < \angle BOA + \angle AOC$. □

Lemma 2.3. *Let A, B, C be points on S . Then $|\widehat{BA}| + |\widehat{AC}| \geq |\widehat{BC}|$.*

PROOF. If A, B and C lie on the same great circle, the conclusion is clear. Otherwise, apply Lemma 2.2. \square

3 Real Hilbert Spaces

In the Existence Theorem we prove that if B and C are non-antipodal points on S , no curve on S from B to C is shorter than the curve

$$X_0(t) = (tC + (1-t)B) / \|tC + (1-t)B\|, \quad 0 \leq t \leq 1.$$

Recall that the length of this curve is $|\widehat{BC}|$.

Existence Theorem. *Let $X(t)$, $0 \leq t \leq 1$, be a continuous curve on S with $X(0) = B$, $X(1) = C$. Then the length of this curve is greater or equal to $|\widehat{BC}|$.*

PROOF. If U, V are points on S with $r = |\widehat{UV}|$, then $|\widehat{UV}|/|UV| = r/(2 \sin(r/2))$. But $\lim_{r \rightarrow 0} r/(2 \sin(r/2)) = 1$. Take any $\epsilon > 0$. There is a $\delta > 0$ such that if $|UV| < \delta$, then $|\widehat{UV}|/|UV| < 1 + \epsilon$ and $|\widehat{UV}| < (1 + \epsilon)|UV|$.

By the uniform continuity of $X(t)$ we can choose points

$$0 = w_0 < w_1 < w_2 < \dots < w_n = 1 \quad \text{such that} \quad |X(w_{i-1})X(w_i)| < \delta$$

for each index i . By Lemma 2.3 we have

$$|\widehat{X(w_0)X(w_1)}| + |\widehat{X(w_1)X(w_n)}| - |\widehat{X(w_0)X(w_n)}| \geq 0,$$

$$|\widehat{X(w_1)X(w_2)}| + |\widehat{X(w_2)X(w_n)}| - |\widehat{X(w_1)X(w_n)}| \geq 0,$$

$$|\widehat{X(w_2)X(w_3)}| + |\widehat{X(w_3)X(w_n)}| - |\widehat{X(w_2)X(w_n)}| \geq 0,$$

$$|\widehat{X(w_3)X(w_4)}| + |\widehat{X(w_4)X(w_n)}| - |\widehat{X(w_3)X(w_n)}| \geq 0,$$

\vdots

$$|\widehat{X(w_{n-2})X(w_{n-1})}| + |\widehat{X(w_{n-1})X(w_n)}| - |\widehat{X(w_{n-2})X(w_n)}| \geq 0.$$

We add and find that the sum telescopes, and

$$\sum_{i=1}^n |\widehat{X(w_{i-1})X(w_i)}| - |\widehat{X(w_0)X(w_n)}| \geq 0.$$

But $X(w_0) = B$, $X(w_n) = C$ and

$$|\widehat{X(w_{i-1})X(w_i)}| < (1 + \epsilon)|X(w_{i-1})X(w_i)|,$$

for each index i . It follows that

$$|\widehat{BC}| \leq (1 + \epsilon) \sum_{i=1}^n |X(w_{i-1})X(w_i)| < (1 + \epsilon)L$$

where L is the length of the curve $X(t)$, $0 \leq t \leq 1$. Finally, ϵ was arbitrary, so $L \geq |\widehat{BC}|$. □

In the Uniqueness Theorem we see that any continuous curve of length $|\widehat{BC}|$ on S joining B to C has range \widehat{BC} . In this sense there is essentially one continuous curve on S joining B to C that is shorter than all the others.

Uniqueness Theorem. *Let B and C be non-antipodal points in S . Let $|\widehat{BC}|$ be the length of the continuous curve $Y(t)$, $0 \leq t \leq 1$, on S with $Y(0) = B$, $Y(1) = C$. Then the range of $Y(t)$ is precisely the set \widehat{BC} .*

PROOF. Let $Y(t_0)$ be a point in the range of $Y(t)$. Put $A = Y(t_0)$. Let L_1 be the length of the curve $Y(t)$, $0 \leq t \leq t_0$, and let L_2 be the length of the curve $Y(t)$, $t_0 \leq t \leq 1$. Then $L_1 + L_2 = |\widehat{BC}|$. By the Existence Theorem $|\widehat{BA}| \leq L_1$ and $|\widehat{AC}| \leq L_2$. Consequently $|\widehat{BA}| + |\widehat{AC}| \leq |\widehat{BC}|$. We deduce from Lemma 2.2, that A must lie on the great circle through B and C . It follows that this circle contains the range of $Y(t)$, which in turn contains an arc of the circle. We leave the rest. □

Observe that in our arguments the range of the curve X need not be contained in any finite dimensional subspace of h . Moreover, the curves X and Y need not have tangent or normal vectors at every point. In this way our work is more general than that in [M], for example, where only everywhere twice differentiable curves are treated.

Let H denote the Hilbert space of sequences of real numbers (x_n) such that

$$\|(x_n)\| = \left(\sum_{n=1}^{\infty} x_n^2 \right)^{\frac{1}{2}} < \infty.$$

(Consult [N, chapter VII, Section 4].) Tangent vectors did not enter our arguments, and it is fitting that we conclude with an example of a continuous curve $W(t)$, $0 \leq t \leq 1$, of finite length in H such that there is no tangent vector at any rational value of t and such that the restriction of W to each component is a polynomial in t .

Let $n(1) < n(2) < n(3) < \dots$ be an increasing sequence of positive integers such that at each index j , $\int_{-1}^1 (1-x^2)^{n(j)} dx < 2^{-j}$. Put

$$F_j(x) = (1-x^2)^{n(j)} \quad \text{for } -1 \leq x \leq 1.$$

Let $g(1), g(2), g(3), \dots$ be a sequence of rational numbers in $[0, 1]$ such that each rational number in $[0, 1]$ occurs an infinite number of times in the sequence. (For example, a diagonalizing process could be used here.) Put

$$W(t) = \left(\int_0^t F_1(x-g(1)) dx, \int_0^t F_2(x-g(2)) dx, \int_0^t F_3(x-g(3)) dx, \dots \right).$$

Note that each component is a polynomial in t . It follows that $\int_0^t F_j(x-g(j)) dx$ is an increasing function on $0 \leq t \leq 1$ with total variation less than 2^{-j} . Routine arguments show that $W(t)$, $0 \leq t \leq 1$, is a continuous curve of finite length in H . Let r be any rational number in $[0, 1]$, and let $g(k) = r$. It follows that the k -th component of the vector $(W(r) - W(r+q))/q$ tends to 1 as q tends to 0. Thus there are infinitely many indices k for which the k -th component of $(W(r) - W(r+q))/q$ tends to 1 and hence every component of

$$(W(r) - W(r+q))/\|W(r) - W(r+q)\|$$

tends to 0 as q tends to 0. It follows that W can not have a tangent vector at $t = r$.

References

- [A] T. M. Apostol, *Calculus, I*, Blaisdell, New York, 1961.
- [M] Z. A. Melzak, *Invitation to Geometry*, Chapter 10, Wiley, 1983.
- [N] I. P. Natanson, *Theory of functions of a real variable, I*, Ungar, New York, 1955.