Isoperimetric Inequality on the Sphere

Daniel Penner

In this paper I will present a proof of an extension of the classical isoperimetric inequality to curves on the unit sphere. The statement of the theorem is as follows.

Theorem 1. *Let γ be a simple closed curve of length L on the unit sphere such that γ separates the sphere into two regions. Let A be the area of the smaller region. Then*

$$L^2 \geq 4\pi A - A^2,$$

and equality holds precisely when γ is a circle.

We will first prove the inequality for the special case in which γ is convex and can be covered by an open hemisphere, and we will then extend this to the general case by geometrical means. The proof of the special case is due to Felix Bernstein [1], while the generalization is due to Tibor Rado [2].

Proposition 2. *Let γ be a simple closed convex curve on the unit sphere which can be covered by an open hemisphere. Then if L denotes the length of γ and A denotes the area of the smaller of the two regions of the sphere separated by γ, then*

$$L^2 \geq 4\pi A - A^2,$$

and equality holds precisely when γ is a circle.

Proposition 3. *If a simple closed curve γ on the unit sphere has length $L < 2\pi$, then γ can be covered by an open hemisphere.*

We first use these two results to prove the main theorem.

Proof of Theorem 1. By Proposition 2, we know that the inequality (and equality case) holds for any convex closed curve which can be covered by an open hemisphere. This generalizes to non-convex curves easily, since if γ_1 is not convex, and γ_2 is a curve which is equal in most parts but is convex, then γ_1 has greater length, but γ_2 has greater interior area, so in fact the inequality is weaker for non-convex curves than for convex curves.

Proposition 3 then implies that all curves which *cannot* be covered by an open hemisphere must have length at least 2π. But letting $f(x) = 4\pi x - x^2$, we can easily see that f attains its maximum at $x = 2\pi$, so that since $L \geq 2\pi$, we obtain that

$$L^2 \geq (2\pi)^2 \geq 4\pi A - A^2,$$

as desired.

Now we tackle the equality case for curves that have $L \geq 2\pi$. Clearly, equality holds precisely when $A = L = 2\pi$. But then γ must contain two points x, y which are endpoints of the same diameter of the unit sphere (we prove this below as part of the proof of Lemma 6). However, since $L = 2\pi$ and γ is closed, this implies that γ must be a union of two great arcs connecting x and y (since a great arc is the only path of length π connecting x and y). Finally, since $A = 2\pi$, γ divides the sphere into two portions with equal area. This implies that the two great arcs comprising γ must form a great circle, so that γ is a circle. $\square$

Next we collect some facts for the proof of Proposition 2.

Definition 4. *We define a transformation T_ε on curves on the unit sphere. Given a simple, closed, convex curve γ on the unit sphere with a fixed tangent t and fixed geodesic curvature $\frac{d\tau}{ds} = \frac{1}{\rho}$, which is contained in the unit hemisphere bounded by its tangent, T_ε sends γ to a "parallel" curve γ_ε (that is, a curve whose tangent lines at every point are parallel to those of γ), so that there is a gap of length ε between the points of γ and those of γ_ε.*

Lemma 5. *The transformation T_ε, when applied to an arbitrary such curve γ, acts as a rotation about the origin of angle ε. In particular, if A and L are the interior area and length of γ, respectively, and A_ε and L_ε are the interior area and length of γ_ε, respectively, the following equalities hold, where $\varepsilon \in [0, \pi/2)$:*

$$2\pi - A_\varepsilon = (2\pi - A)\cos\varepsilon - L\sin\varepsilon,$$
$$L_\varepsilon = (2\pi - A)\sin\varepsilon + L\cos\varepsilon.$$

Proof. By definition we have that $\frac{\partial A_\varepsilon}{\partial \varepsilon} = L_\epsilon$ and $\partial\frac{2\pi - A_\varepsilon}{\partial\varepsilon} = -L_\epsilon$ [1, p. 125]. We also have that $\frac{\partial}{\partial\varepsilon} ds_\varepsilon = d\tau_\varepsilon$, and that $\frac{\partial}{\partial\varepsilon} L_\varepsilon = 2\pi - A_\varepsilon$ [1, p. 121]. Here $L_\varepsilon = \int ds_\varepsilon$ and $2\pi - A_\varepsilon = \int d\tau_\varepsilon$. Noting also that

$$\frac{\partial}{\partial\varepsilon}(2\pi - A_\varepsilon + iL_\varepsilon) = i(2\pi - A_\varepsilon + iL_\varepsilon),$$

[1, p. 125] we can integrate the above from 0 to ε to obtain that

$$2\pi - A_\varepsilon + iL_\varepsilon = e^{i\varepsilon}(2\pi - A + iL),$$

for $\varepsilon \in [0, \pi/2)$. But taking real and imaginary parts gives us precisely the pair of formulas stated in Lemma 5, concluding our proof. $\square$

It follows immediately from Lemma 5 that the quantity $(2\pi - A)^2 + L^2$ is invariant under the transformation T_ϵ, so that

$$(2\pi - A_\varepsilon)^2 + L_\varepsilon^2 = (2\pi - A)^2 + L^2$$

for any curve satisfying our assumptions. Additionally, if we let $\varepsilon = (2\pi - A)/L$, the equations of Lemma 5 give us that $A_\varepsilon = 2\pi$ (that is, γ_ε splits the sphere into two parts of equal area) so that the invariance property gives the equality

$$L_\varepsilon^2 = (2\pi - A)^2 + L^2. \tag{1}$$

We have now reduced the problem to showing that any curve of our given qualities which splits the unit sphere into two parts of equal area has length at least 2π, which would give us that $L_\epsilon \leq 2\pi$, so that (1) becomes our desired inequality.

Lemma 6. *If γ is a simple closed curve on the unit sphere which divides the sphere into two parts of equal surface area, the length of γ, L, is at least 2π, and $L = 2\pi$ precisely when γ is a great circle.*

Proof. First we will show that γ must contain at least one pair of diametrically opposite points. We denote by γ' the curve which consists of all points diametrically opposite to those of γ, and note that γ' also splits the sphere into two parts of equal surface area.

Assume γ and γ' do not intersect. Then they would split the sphere into three separate parts, which we call T_1, T_2 and T_{12}, where T_1 and T_2 are enclosed by γ and γ', respectively, and T_{12} is bounded between γ and γ'. But this is not possible, since γ and γ' must both enclose sections of area 2π, so that if the area of T_{12} is nonzero, the sum of the three distinct regions' areas would be greater than 4π, which is impossible.

So γ and γ' must intersect at least at a point R. However, its diametric opposite, R', must clearly also be an intersection point of γ and γ', so that γ contains at least one pair of diametrically opposite points.

Now, R and R' split γ into two separate segments, γ_1 and γ_2, which have lengths L_1 and L_2, respectively. If either γ_1 or γ_2 is not a great semicircle, its length is greater than π, so that the length of γ is greater than 2π. If the length of γ is 2π, then γ_1 and γ_2 must both great semicircles, and indeed γ must be a great circle, for otherwise it could not split the unit sphere into two sections of equal area. □

Proof of Proposition 2. The result follows immediately from the formula (1) derived from Lemma 5, and the conclusion of Lemma 6 that the length L_ε of γ_ε must be at least 2π, so we have that

$$(2\pi - A)^2 + L^2 = L_\varepsilon^2 \geq (2\pi)^2,$$

or in other words that

$$L^2 \geq 4\pi A - A^2.$$

□

Finally, we gather a few lemmas for the proof of Proposition 3.

Lemma 7. *If γ is a simple closed curve on the unit sphere of length strictly less than 2π and C is a great circle of the unit sphere, then C has an open sub-arc of length strictly greater than π which does not intersect γ.*

Proof of Lemma 7. If γ and C do not intersect, then the lemma is a trivial statement, so we may assume that C intersects γ at a set of points we call S, which is clearly closed since the curves are continuous (and thus any point of non-intersection has a neighborhood which contains no points of intersection). Then the complement of S on the curve C is a set of distinct sub-arcs separated by the points of S. Call these sub-arcs $c_1, c_2, \ldots, c_n$. Since there are finitely many, we can pick c_1 to be one of maximal length. Then the endpoints x and y of c_1 are points of γ. Since γ has length less than 2π, this implies that any two points on γ, particularly x and y, cannot be diametrically opposite (because the shortest closed path containing two such points is a union of two great arcs, which has length $2\pi > L(\gamma)$). Therefore the length of c_1 is not equal to π.

Assume $L(c_1) < \pi$. Then if x' and y' are the points diametrically opposite to x and y, respectively, and c_1' is the sub-arc of C joining them, we can see that c_1' and c_1 may not overlap, since $L(c_1') = L(c_1) < \pi$. By our earlier observation that any two points of γ cannot be diametrically opposite, it follows that neither x' nor y' are points of γ, and thus not points of S, the set of points of intersection of γ and C. Since S is closed, we can be sure that x' and y' are not limit points of S, so that there are small neighborhoods about each point which do not contain intersection points of γ and C.

If c_1' does not intersect γ, then this means we can extend c_1' past x' and y' without collecting any points which intersect γ, so that c_1' is a proper subset of one of the distinct sub-arcs c_i from earlier ($i \neq 1$), say, c_k. But then $L(c_k) > L(c_1') = L(c_1)$, which contradicts the maximality of c_1. So $L(c_1) > \pi$.

On the other hand, if c_1' does contain a point $z \in S$, that is, one which intersects γ, then the three distinct sub-arcs of C separated by x, y, and z necessarily have lengths less than π (since c_1 has length less than π by assumption, and the arc bounded by x and z, for instance, must be of length strictly less than that bounded by x and x', which is π). However, this is not possible, since x, y, and z also separate the curve γ into three distinct parts, γ_1, γ_2, and γ_3, also bounded by x, y and z, and since the sub-arcs of C which connect those three points must be the shortest paths between them, $L(\gamma) = L(\gamma_1) + L(\gamma_2) + L(\gamma_3) \geq L(\tilde{c}_1) + L(\tilde{c}_2) + L(\tilde{c}_3) = L(C) = 2\pi$ (where the $\tilde{c}_i$ are the sub-arcs separated by x, y and z), contradicting that $L(\gamma) < 2\pi$. So $L(c_1) > \pi$. □

Definition 8. *Given a set S of points on the unit sphere, we say that S is OH if there is some open hemisphere containing S, and that S is CH if there is some closed hemisphere containing S.*

Lemma 9. *Let γ be a simple closed curve on the unit sphere whose length is strictly less than 2π. Then if some closed sub-arc Γ of γ is CH, Γ is also OH.*

Proof of Lemma 9. Without loss of generality, we assume that Γ is contained in the closed southern hemisphere of the unit sphere, whose boundary is the equator E. By Lemma 7, since E is a great circle, there is an open sub-arc C of E which does not intersect γ (and thus Γ) and has length strictly greater than π. We take a closed sub-arc C' of C which has length π, and clearly still does not intersect Γ. Since it has length π and is contained in a great circle, its endpoints x and y are diametrically opposite. Let φ be the isometry which rotates the sphere about the diameter connecting x to y such that it rotates C' into the southern hemisphere by a small enough angle that the curve $\varphi(C')$ still does not intersect Γ' (possible since both C' and Γ are closed sets and do not intersect). Then under φ, the other half of the equator E, which we denote by $E - C'$, is rotated into the northern hemisphere, and thus $\varphi(E - C')$ does not intersect Γ, which is contained in the closed southern hemisphere.

Therefore the great circle $\varphi(E)$, since it is comprised of $\varphi(E - C')$ and $\varphi(C')$, does not intersect Γ, and Γ is covered by the open hemisphere beneath this great circle. Thus Γ is OH. □

Proof of Proposition 3. First we assign γ a positive sense of direction, and a "starting" point $p \in \gamma$, such that $\gamma(s)$ is the point on γ of distance s along γ in the positive direction from the point p (where $s \in [0, L(\gamma)]$, noting that $L(\gamma) < 2\pi$). We write S for the set of values of s for which the sub-arc of γ extending from p to $\gamma(s)$ is CH. We can see that S is non-empty (since it contains at least 0 and many small values) and closed (since we are considering coverings by closed hemispheres). Therefore S contains a maximal element s'. Assume $s' < L(\gamma)$. Then the arc $\gamma(s), s \in [0, s']$ is CH, and thus by Lemma 9, also OH. But then we can pick some small ε such that $\gamma(s), s \in [0, s' + \varepsilon]$, is also OH, since the covering hemisphere is open. But then we can take the closure of this hemisphere to show that $\gamma(s), s \in [0, s' + \varepsilon]$ is also CH, which implies that $s' + \varepsilon \in S$. But this contradicts the maximality of s'. Then $s' = L(\gamma)$, which means that $\gamma(s), s \in [0, L(\gamma)]$ (that is, the whole curve γ) is CH, and thus also OH by Lemma 9. This concludes the proof. □

References

[1] F. Bernstein. *Über die isoperimetrische Eigenschaft des Kreises auf der Kugeloberfläche und in der Ebene*, Mathematische Annalen, Vol. 60 (1905), 117-136.

[2] T. Radó. *The Isoperimetric Inequality on the Sphere*, American Journal of Mathematics Vol. 57, No. 4 (1935), 765-770.