A Historical Review of the Isoperimetric Theorem in 2-D, and its place in Elementary Plane Geometry

Alan Siegel[1]

Courant Institute of mathematical Sciences
New York University
New York

1 Introduction

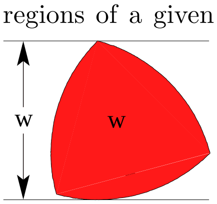
The isoperimetric theorem states:

Theorem A. Among all planar regions with a given perimeter p, the circle encloses the greatest area.

This result, which is also known as the isoperimetric inequality, dates back to antiquity.

The theorem has generalizations to higher dimensions, and even has many variants in two dimensions. For example, one version states that among all polygons with k sides and a fixed perimeter, those that are perfectly symmetric (i.e., regular) have the greatest area. These area and volume optimization theorems are especially appealing because they offer physical insights into nature. They tell us why a cat curls up on a cold winter night (to minimize its exposed surface area) [11]. They help us understand why honeybees build hives with cells that are perfectly hexagonal in shape. The isoperimetric theorem also helps explain why water pipes should have a round cross-section.

Of course, nature is complicated, and the underlying mathematics can be difficult. Even the simplest two-dimensional honeycomb conjecture, which dates back to Pappus, has an extensive literature [2, 4, 7], and was only recently established in its entirety [4]. Furthermore, many shapes are influenced by forces and other constraints as opposed to the efficient use of area and volume. The isoperimetric inequality does not really explain why manhole covers are round.[2] Similarly, if coins are required to present a constant width for vending machines, then many non-circular shapes are available to meet this standard. A result of J.E. Barbier states that all constant-width regions of a given width w have perimeter πw, whence the isoperimetric inequality shows that circular coins have the most metal, and are therefore the most expensive. Perhaps this fact explains why the British 50 and 20 pence coins are shaped as heptagonal analogs of the Rouleaux triangle.[5]



A Rouleaux triangle

In three dimensions, the sphere has the greatest volume for a given surface area, but this fact does not prove that eggs must be spherical. Sound modeling must not abstract away the hen and the process of egg laying, which will impose non-uniform forces on the egg. Still, the 2-D version of the theorem seems to help explain why eggs have a circular cross-section.

[1]This research was supported in part by NSF grant CCR-9503793

[2]The true reason is to ensure that the cover cannot fall into the manhole; the problem is, in a sense, $2\frac{1}{2}$ dimensional. On the other hand, the mathematical requirement that the cover be unable to "fall in" would be satisfied by any region of constant width. We must then appeal to either the isoperimetric inequality and the economics of manhole design or a lack of knowledge about constant width regions to explain why all such covers happen to be round.

[5]We resist the temptation to speculate about whether such cost-efficient engineering practices might have contributed to Great Britain's hesitance to adopt the Euro.

This remarkable theorem even has a literary history dating back some twenty-one centuries to Virgil's Aeneid and the saga of Queen Dido [13]. Apparently, the good Queen had more than her fair share of entrepreneurial skill and mathematical ability—as well as misfortune of epic proportion. Her legend recounts, among other tragedies, the murder of her father by her brother, who then directed his intentions toward her. She was obliged to assemble her valuables and flee her native city of Tyria in ancient Phoenicia.

In due course, her ship landed in North Africa, where she made the following offer to a local chieftain. In return for her fortune, she would be ceded as much land as she could isolate with the skin of an ox. The proposition must have seemed too good to refuse. It was agreed to, and a large ox was sacrificed for its hide. Queen Dido broke it down into extremely thin strips of leather, which she tied together to construct a giant semicircle that, when combined with the natural boundary imposed by the sea, turned out to encompass far more area than anyone might have imagined. And upon this land, the city of Carthage was born.

It is easy to understand why such legendary exploits would be preserved in the historical and literary record for all time. In terms of her business skills alone, Queen Dido must surely rank among the most remarkable individuals of her millennium. But as innovative and significant as her achievements may have been, one might wonder who among us—apart from Bill Gates and, perhaps, Donald Trump—might possess the capacity to master the Aeneid's teachings about barter and real estate negotiation.

On the other hand, Queen Dido's mathematics lessons are quite another matter. Evidently, she knew the isoperimetric inequality, and understood how to use this fact to find the best solution to her problem, which uses a semicircle rather than a circle. Yet despite the many insights and occasional power transactions attributable to this very special inequality, most high school geometry texts give this theorem rather short shrift. It is typically not even mentioned, and proofs are never offered.

Surely the reason for these omissions cannot be attributed to any deficiencies in the isoperimetric theorem or its applications, but rather in its proof.

The first rigorous proofs date back to Weierstrass and to F. Edler, whose methods were founded in analysis and calculus. Since that time, many other proofs have been discovered. Some use calculus, and even the calculus is somewhat advanced. Among the calculus proofs, the shortest and one of the simplest is due to P.D. Lax [8]. There are also more sophisticated analytic studies of isoperimetric inequalities in higher dimensions, and even on surfaces [9]. These inequalities are all considerably more advanced than the problem we address.

The history of purely geometric proofs, however, is quite different. The theorem was known to the ancient Greeks, and was recorded by Pappus in the fourth century A.D. [10, Book V]. He, in turn, credited the isoperimetric results to Zenodorus, who lived during the second century B.C. But by modern standards, their proofs were severely incomplete; Pappus and his contemporaries seem to have just assumed that the nice pictures they drew must capture the essence of what had to be established. They left no historical record about the irregular cases which, of course, always turn out to have less area than the more natural figures they considered, but which must nevertheless be taken into account in any rigorous proof. Archimedes also studied the problem, but his work on the subject, like the original writings of Zenodorus, has been lost. No one knows if these studies might have contained a proof complete enough to meet the standards of modern mathematics.

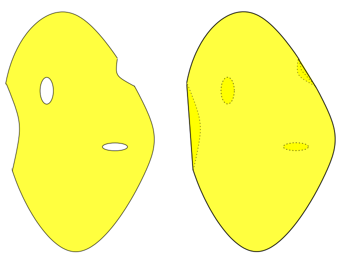
The modern-day search for a rigorous proof can be traced to Steiner, who realized that the ancient Greek arguments were inadequate. He reasoned that a better way to establish the inequality would be by showing how any figure that does not have a circular boundary can be transformed into a new region with the same perimeter and greater area. In 1841, Steiner published the first of five elegant improvement procedures, which lie within pure Euclidean geometry, and which he viewed as rigorous. However, Weierstrass, who was the strongest mathematician of the era, thought otherwise, and developed mathematics to a level that could formalize Steiner's error. He also advanced calculus and other areas of mathematical analysis to a point where the problem could be solved correctly. The final conclusion, which is indeed substantiated by modern mathematical understanding, is that Steiner's ideas comprise significant and insightful contributions to Euclidean geometry, but his proof of the isoperimetric theorem is fundamentally incomplete.

As of the mid 1960's, the question of finding an elementary geometric proof was, according to the literature, widely believed to be open (cf. [5, 3, 11]). Kazarinoff makes this point most emphatically: "no one has yet found a simple geometrical argument to show that the circle has a greater area than any other figure with the same perimeter. None is expected to be found [5, p. 61-2]."

Those are strong words. Our literature search found several proofs that lie within of elementary geometry. Yaglom, for example, sketches out one such approach [14], but it is definitely not simple. Benson presents an elegant monotonicity argument that exploits some intuitive notions from integral geometry plus a bit of clever algebraic manipulation [1, p.127-8] to establish Theorem 1 with elementary but subtle reasoning. An extended literature search even revealed that the approach we now present was actually discovered by Gary Lawlor about two years before we attained it through independent means [6].

1.1 Some geometric preliminaries

Many geometric problems are simplified by exploiting notions about convexity, which is an important concept in its own right. A thumbnail sketch of the idea is as follows.

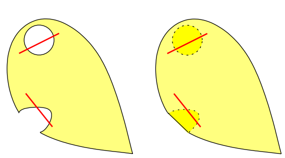


not convex convex

From an intuitive perspective, a region is *convex* if it has no holes and the boundary has no dents. We can fill in the holes and dents of a non-convex region F by putting a rubber band around it and taking the inside to be the new region. This new region is called the *convex hull* of F.

Observe that the convex hull of F will be precisely F if the region is already convex, and will otherwise have more area, since some hole or dent was filled in by the convex hull procedure. Moreover, the perimeter will decrease because the convex hull replaces dents by straight line segments—and a straight segment comprises the shortest path between its endpoints, as opposed to some piece of curvy boundary that defines a dent in F.

A more formal definition of convexity is as follows.

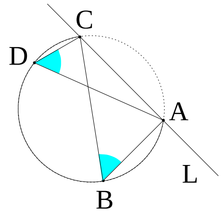


A region F is *convex* if for any pair of points A and B in F, the segment $\overline{AB}$ is fully contained in F.

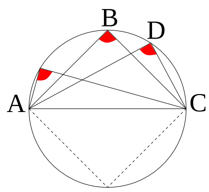
This definition prevents F from having any holes or dents.

Some additional terminology is also helpful. A polygon with k sides is called a k-gon, and a circle plus its interior is called a disk.

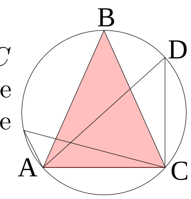
Lastly, the proof will depend on one more geometric fact, which is sometimes formulated as follows.



The Peripheral Angle Theorem. Let $\overline{AC}$ be a chord of the circle S, and suppose that B is another point on S. Let L be the infinite line through A and C as shown. Suppose that D is on the same side of L as B. Then $\angle ABC = \angle ADC$ if and only if D is on the circular arc $A\widehat{B}C$.

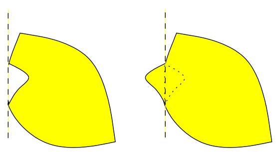


As a consequence, if $\overline{AC}$ is a diameter, then all such angles $\angle ADC$ will equal $90°$. Similarly, among all triangles with a common base and congruent opposing angles, the isosceles triangle will have the greatest altitude and, therefore, the greatest area.

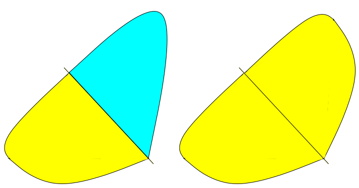


1.1.1 Steiner's Method

Steiner's approach was to characterize those regions F that do not have the maximum area for a given perimeter.

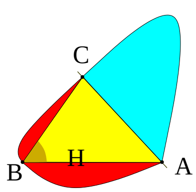


First, he observed that the area of F can be increased if F is not convex. In the leftmost figure, the region is non-convex because of an inward depression. This piece of boundary can be reflected across the support line as shown. The result is a new region with greater area and exactly the same perimeter.

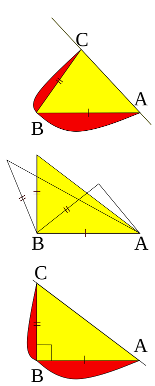


Now suppose that F is convex. Steiner suggested drawing a line that partitions F into two pieces with equal perimeters. If one of the halves has greater area, select that larger piece. Create a new region with this half and a mirror image of itself. The new figure must have a greater area than F and the same perimeter.

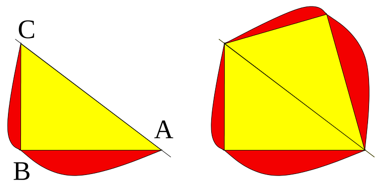
So suppose that F is convex, and suppose that this bisection line produces two halves with equal area and equal perimeters.



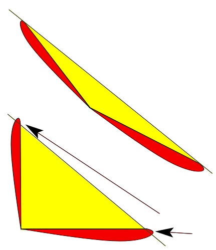
Let the boundary of F intersect the bisection line at points A and C. The Peripheral Angle Theorem says that F will be circular if and only if for every point B on its boundary (apart from A and C): $\angle ABC = 90°$. Consequently, we suppose that there is such a B where $\angle ABC \neq 90°$. Let H be the half region of F that contains B.



Steiner's idea was to view $\triangle ABC$ as having two sides with the fixed lengths $|\overline{AB}|$ and $|\overline{BC}|$, and to imagine a hinge at vertex B so that the angle at B is free to change. Thus, H is the union of three subregions: the adjustable $\triangle ABC$, the region bounded by $\overline{AB}$ and the portion of the boundary subtended by segment $\overline{AB}$, and the analogous region associated with the segment $\overline{BC}$. Changing $\angle ABC$ will change the area and shape of $\triangle ABC$, but not the other two pieces, which are essentially rigid shapes that are glued to their respective sides $\overline{AB}$ and $\overline{BC}$. Evidently, the triangle with base length $|\overline{AB}|$ and side length $|\overline{BC}|$ has the greatest area when its altitude is maximum. This occurs precisely when $\angle ABC$ is $90°$. So if the angle is not a perpendicular, the area can be increased by changing the angle to be $90°$.



Once angle B is upgraded to $90°$, the new figure can be combined with a mirror image copy to produce a new region that has the same perimeter as F and greater area.



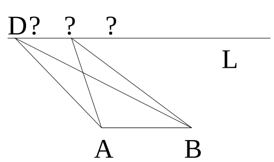
For completeness, we note that for some figures, decreasing $\angle ABC$ to $90°$ will cause portions of the region to spread across the bisection line L as shown. There are several ways to fix this problem; the details are left as an exercise for the interested reader.

These constructions show that if F is not circular, then there is some other region that has the same perimeter and even greater area. Consequently, if there is a region with the maximum area, it must be a disk, and cannot be anything else.

But are we entitled to conclude that a maximum area region actually exists?

1.1.2 The gap in Steiner's argument

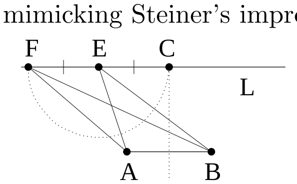
To see why Steiner's argument is incomplete, consider the following problem.



Given: A segment $\overline{AB}$ of unit length, and a line L that is parallel to $\overline{AB}$ and has unit distance from the segment.

Problem: Find a point D on L, so that $\angle ADB$ is as small as possible.

Of course, there is no such best point: by selecting candidate points that are located farther and farther to the left, we get a sequence of angles that approach $0°$, but there is no point on L that gives this angle. Nevertheless, the following argument pretends to find a best angle by mimicking Steiner's improvement scheme. Evidently, there must be an error.



Let C be the point on L that is equidistant from A and B, so that it lies at the intersection of L and the perpendicular bisector of $\overline{AB}$. Let E be a candidate solution that is not C. Use a compass to locate F on L so that E is the midpoint of segment $\overline{FC}$. Then F is an improvement over E. In particular, a standard extension of the Peripheral Angle Theorem will show that $\angle AFB < \angle AEB$. Evidently, this procedure will transform any point $E \neq C$ on L into a better point F, and the scheme will leave C unchanged.

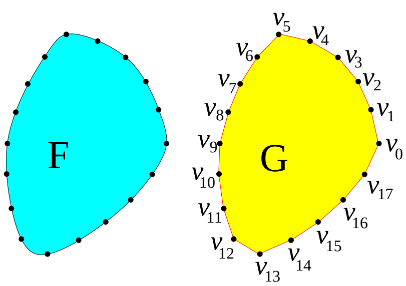
Now, $\angle ACB$ is the largest possible angle, not the smallest. So we cannot conclude that $\angle ACB$ is the smallest angle, despite the fact that no other point can be the solution.

From an intuitive perspective, the difference between this contrived improvement scheme and Steiner's elegant approach is obvious: his scheme makes figures more circular, in some sense, whereas the above scheme moves candidate solution points farther away from the fixed point C. But the underlying mathematical issue is: how do we formalize the notion that the Steiner improvements could lead to a circle? This question, it turns out, is resolved by the notion of compactness, which characterizes when and how limit solutions exist. Blashke was the first to show that compactness arguments can complete Steiner's proof, and more elegant justifications have also been discovered. However, these approaches lie outside of elementary Euclidean Geometry.

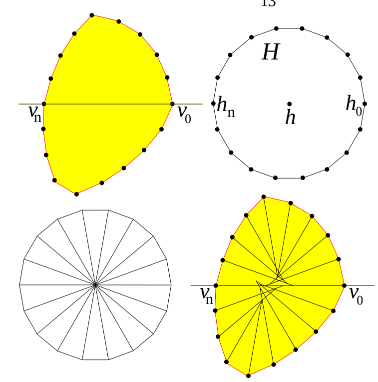
2 A classical dissection-style proof

The proof will assume that the region of interest F is convex. If not, we can replace F by a region that is convex, has greater area, and a perimeter that is no larger than that of F. There are many was to find such a region. One way is to use the convex hull of F. This notion was not featured in the ancient Greek works, but has certainly come to be viewed as elementary. A completely elementary alternative is presented after the following proof.

As is standard in studies of elementary geometry, we omit any formal development of measure theory, and simply assume that area and arclength are well defined. We also deliberately avoid formalizing the notion of a planar region. The reader who prefers more rigor may wish to suppose that F is a convex polygon. Then the notions of area and arclength need no elaboration.



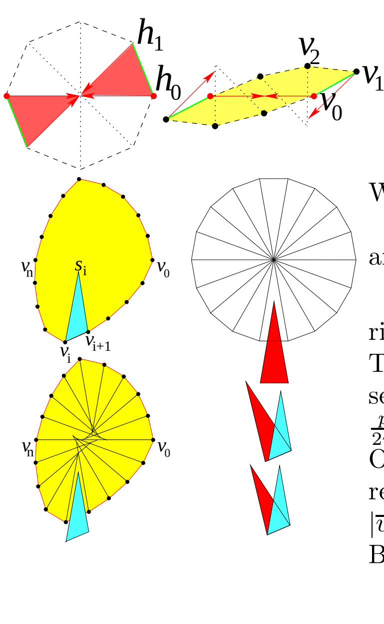
The Proof. [First discovered by Lawlor [6]]
Let F be a convex region. Let $n \geq 2$ be an integer.
• Place the points $v_0, v_1, \ldots, v_{2n-2}, v_{2n-1}$ along the boundary of F so that the length of boundary between any pair of consecutive points is exactly $\frac{p}{2n}$.
• Let G be the polygon formed by connecting edges between consecutive vertices.



Draw a regular (perfectly symmetric) $2n$-gon H with center h, vertices $h_0, h_1, \ldots, h_{2n}$ and each side of length $\frac{p}{2n}$ as shown.

Thus, the inscribed polygon G is an approximation of F, and H is an approximation of a circle with perimeter p. We will show that H encloses at least as much area as G. By letting n go to infinity, the areas of the approximating polygons converge to the areas enclosed by F and the circle as used in the statement of the isoperimetric theorem.

Let the figures be rotated so that $\overline{v_n v_0}$ and $\overline{h_n h_0}$ are parallel and horizontal. Draw a segment from each vertex of H to h. From each vertex of v_i of G, draw a ray $\vec{r}_i$ that is parallel to the axial ray that points from h_i through the center h. (Although the $\vec{r}_i$ are infinite, they have been drawn as finite segments to expose the underlying geometry.) We might expect, as the figure suggests, that the consecutive rays r_i and r_{i+1} will, in general, intersect.



However, it is possible that some pairs of consecutive rays (such as, in the contrived illustration, rays r_0 and r_1, which emanate from v_0 and v_1 as shown) might not intersect.
• But if $\vec{r}_i$ and $\vec{r}_{i+1}$ do intersect at some point s_i, let t_i be the triangular region $\triangle v_i v_{i+1} s_i$. Otherwise, it is convenient to define t_i as the segment $\overline{v_i v_{i+1}}$.

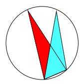
We claim:

(a) $Area(t_i) \leq Area(\triangle h_i h_{i+1} h)$,

and

(b) The "regions" $t_0, t_1, \ldots, t_{2n-1}$ cover G and all of its interior.

To establish (a), it suffices to assume that t_i is a triangle, since segments have zero area. The length of the base $\overline{v_i v_{i+1}}$ is at most $\frac{p}{2n}$, since the vertices v_i were placed at a distance of $\frac{p}{2n}$ along F. Of course, $\triangle h_i h_{i+1} h$ has a base length of $\frac{p}{2n}$. So we can shrink (or rescale) $\triangle h_i h_{i+1} h$ into a similar triangle whose base length equals $|\overline{v_i v_{i+1}}|$, and relocate the two triangles to share a common base. By construction, both triangles have the same angle at the vertex opposite their base, since the rays $\vec{r}_i$ and $\vec{r}_{i+1}$ were chosen to be parallel to sides $\overline{h_i h}$ and $\overline{h_{i+1} h}$.

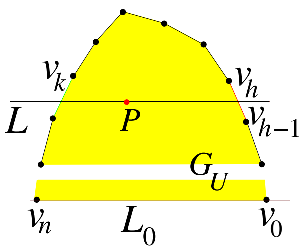


According to the Peripheral Angle Theorem, these two vertices will lie on a circle that goes through the endpoints of their common base. But since the triangle corresponding to (a possibly shrunken) $\triangle h_i h_{i+1} h$ is isosceles, it must have the larger altitude and area.

Claim (b)—which says that while individual "triangles" might seem somewhat problematic in definition and chaotic in behavior, they are collectively quite well behaved—is established next.

Lemma 1. The "triangles" $t_0,\ t_1, \ldots,\ t_{2n-1}$ cover G and its interior. Formally, let $\mathbf{G}$ be the finite polygonal region with boundary G. Then $\mathbf{G}$ is contained in the union of the triangular regions:

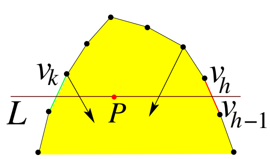
$$\mathbf{G} \subset \cup_{0 \le i < 2n}\, t_i.$$



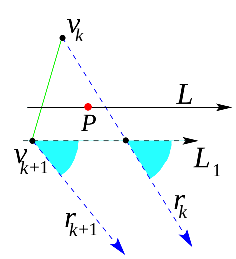
Proof of Lemma 1. The proof will be by contradiction. Suppose that P is a point in $\mathbf{G}$ that is not covered by any t_i. Draw a horizontal line L_0 through vertices v_0 and v_n. The line splits $\mathbf{G}$ into two pieces, which we call G_U and G_L. By construction, the rays $\vec{r}_0$ and $\vec{r}_n$ point at each other and lie on L_0. So we can suppose that P lies in G_U, since the proof for the alternative case is exactly the same. The definition of the t_i-s ensures that P must be in the interior of $\mathbf{G}$, since t_i contains $\overline{v_i v_{i+1}}$.

Now draw a horizontal line L through P. Let the vertices of G_U that lie above L be v_h, v_{h+1}, …, v_k, so that L intersects $\overline{v_{h-1} v_h}$ to the right of P, and $\overline{v_k v_{k+1}}$ to the left. Since P lies in the interior of $\mathbf{G}$, v_h and v_k exist (but could be the same).

While the proof will be completed soon, we momentarily pause to recommend that the reader try to draw a counter-example before proceeding any further. The difficulty is that P will wind up being trapped between some pair of consecutive rays. The following definition formalizes a rotational sense for the rays to simplify the proof of this trapping property.



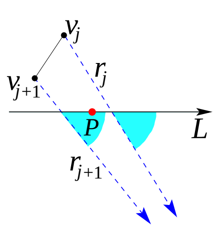
- For $h \le j \le k$, let ray r_j be called *right-oriented* if it intersects L to the right of P, and *left-oriented* if it intersects to the left. A ray that intersects P will be both left-oriented and right-oriented. By construction, the rays $\vec{r}_1$ through $\vec{r}_{n-1}$ all have a downward direction, so rays $\vec{r}_h$ through $\vec{r}_k$ will intersect L and be classifiable. All other rays are left unclassified.



Bolstered by these classifications, we return to the proof. There are three cases.

Case 1: Ray r_k is right-oriented. We show that P must be contained within triangle t_k. Draw a horizontal line L_1 through v_{k+1}, and view it as directed to the right. Ray r_k and segment $\overline{v_k v_{k+1}}$ must intersect L_1 in the same order as they intersect L. Hence r_k intersects L_1 to the right of v_{k+1} as shown. Since r_{k+1} forms an angle with L that is (in absolute value) less than that of r_k, the two rays must intersect below L. Hence s_k exists and P is trapped inside of $\triangle v_k v_{k+1} s_k$.

Case 2: Ray r_h is left-oriented. The reasoning is analogous to Case 1.

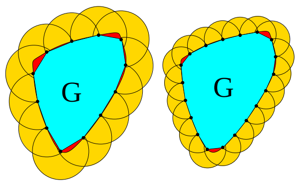


Case 3: Ray r_h is exclusively right-oriented and r_k is exclusively left-oriented. Let r_j be the ray that, among the right-oriented rays within r_h, r_{h+1}, r_{h+2}, …, r_k, has the largest possible index. Since r_h is right-oriented, there is such an r_j. Moreover, $j < k$ since r_k is not right-oriented. Hence r_{j+1} is classified, and it must be left-oriented. Evidently, $\overline{v_{j+1} v_j}$ lies in the halfplane above L; r_j intersects L to the right of P; and r_{j+1} intersects to the left. Since r_{j+1} forms an angle with L that is (in absolute value) less than that of r_j, the two rays must intersect below L. Hence $\triangle v_{j+1} s_j v_j$ exists and must contain P. ∎

So $Area(G) \le Area(t_0) + Area(t_1) + \cdots + Area(t_{2n-1}) \le 2n Area(\triangle h_0 h_1 h) = Area(H)$.

If we pass to the limit by increasing n to infinity, the area of the approximating G-s approaches the area of F, and the area of the regular $2n$-gons approaches the area of a disk with perimeter p by standard arguments. The proof is complete. ∎

Error bounds. The preceding construction only showed that the area of the inscribed polygon G is bounded by $Area(D_p)$, where D_p is a disk with perimeter p. The proof relied on the fact that the area of G should be, somehow, as close to the area of F as we like, for suitably large n. Of course, this must be true, as otherwise we do not have a satisfactory theory of area for general regions. On the other hand, it is worth remarking that this reasoning can be made rigorous by appealing to a simple covering argument.



Recall that consecutive vertices of G are defined to partition F's boundary into pieces of length $\frac{p}{m}$, where m is the number of vertices used. Consequently, no part of the boundary that lies between vertices v_i and v_{i+1} can be at a distance from v_i that exceeds $\frac{p}{m}$. So by centering a disk of radius $\frac{p}{m}$ on each vertex, we ensure that F will be completely covered by G and the m disks. (It is not difficult to see that a covering is achieved by disks with half this radius, but this hardly matters.) Since these disks increase the total area of the covering by at most $m \times \pi(\frac{p}{m})^2 = \pi\frac{p^2}{m}$, we get the following overestimate:

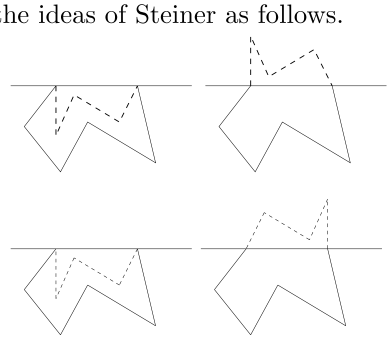
$$Area(F) \leq Area(G) + \pi\frac{p^2}{m} \leq Area(H_{m,p}) + \pi\frac{p^2}{m},$$

where $H_{m,p}$ is a regular m-gon with perimeter p. Letting m go to infinity shows that

$$Area(F) \leq Area(D_p). \quad \blacksquare$$

As presented, the proof shows that no planar region with perimeter p can have an area that exceeds $Area(D_p)$. So D_p does indeed have the largest area, and it now follows from Steiner's results that the disk is the only such figure with this maximum area: otherwise there would be even larger regions with perimeter p.

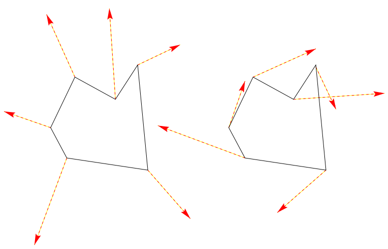
Convexity revisited. To keep the argument squarely within the genre of Euclidean geometry, it is convenient to suppose that F is a polygon. In this case, the convex hull of F is easily defined, but the notion lies outside of the world of Euclid. An attractive alternative can be extracted from the ideas of Steiner as follows.



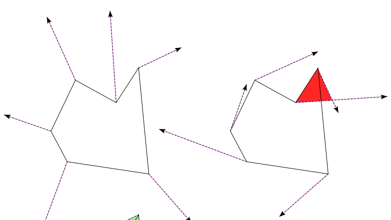
Suppose that the line ℓ seals off some pocket of F as shown. Technically, all of F should lie in one of the halfplanes bounded by ℓ, and there should be distinct vertices v_i, v_j such that $F \cap \overline{v_i v_j} = \{v_i, v_j\}$. Steiner symmetrization flips (reflects) one boundary subpath connecting v_i to v_j about ℓ to increase the area of the resulting figure while preserving the edge lengths. Unfortunately, this mirroring scheme presents a new problem: how do we prove that the process terminates? As a practical matter, it is simpler to reverse the sequencing of these edges, which effectively rotates the boundary portion about the midpoint of $\overline{v_i v_j}$. While iterations of either operation lead, eventually, to a convex region with increased area, the virtual rotation operation is easier to analyze. It simply rearranges the segment ordering without introducing any rotations. Since this procedure can produce no more than $(n-1)!$ such arrangements, one of these polygons must have a maximal area. This polygon must be convex, since otherwise the procedure would further increase the area. Of course, the figure becomes convex when the segments are sorted according to their directed orientations, where the orientations are determined by a traversal around the region.

3 Extensions

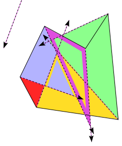
The covering property stated in Lemma 1 has a natural but somewhat surprising generalization [12], which we state without proof.



Theorem B. Let P be a simple polygon that is not necessarily convex. Let the vertices of P be in, counterclockwise order, $v_1, v_2, \ldots, v_n$. Let, for $i = 1, 2, \ldots, n$ the ray r_i emanate from v_i and form an the angle θ_i with a horizontal ray emanating to the right from v_i, and suppose that $\theta_1 < \theta_2 < \ldots < \theta_n < \theta_1 + 2\pi$.



Whenever the rays from two consecutive vertices intersect, let them induce the triangular region defined by the two vertices and the intersection point.



Then there is a fixed α such that if all of the assigned angles are increased by α, the triangular regions induced by the redirected rays cover the interior of P.

4 Conclusions

The basic isoperimetric theorem for two dimensions can be established with very elementary geometry, and the reshaping ideas of Zenodorus and Pappus can indeed be completed to give solid proofs. Who knows? The day may yet come when an archaeological dig will prove that the ancient Greeks had found comparable constructions, or maybe ones that were even better.

5 Exercises and further questions

1. Dido's problem is usually presented as follows. Let s be a string of length ℓ, and let L be a straight line. Among all planar regions with a boundary that can be formed by part of L and all of s, find those with the largest area.

a. Solve Dido's problem.

There are many variations to this problem. Suppose we have the same problem where L is not a line but instead is the following:

b. L is a finite line segment. Characterize the best solution.

c. L is an angle $\alpha < \pi$, and s must be placed in the interior of the region bounded by L. Characterize the best symmetric solution. Is this solution the best possible without the requirement of symmetry?

d. L is an angle $\alpha > \pi$. characterize the best symmetric solution. Is this solution the best possible without the requirement of symmetry?

e. L is a parabola, and s is to be placed in its interior. characterize the best symmetric solution. Show that no non-symmetric solution can have the greatest area. Conclude that the symmetric solution has the greatest area.

f. L is a parabola, and s is to be placed in its exterior. Characterize the best symmetric solution. Can this solution the be improved by dropping the symmetry requirement?

g. L is a circle, and s is to be placed in the exterior of the circle. Show that in the best solution, the two curves are perpendicular to each other.

2. Prove Barbier's theorem.

3. Use theorem B to prove the following standard variant of the isoperimetric inequality.
Let F be a polygon. Then among all polygons with edges congruent to those of F, those polygons whose vertices lie on a circle have the greatest area.
Suggestion: First assume that the inscribed polygon contains the center of the circumscribing circle.
Then address the case where the center is not inside the polygon.

4. Use theorem B to prove the following.
Let P be an n-gon with side lengths of $\ell_1, \ell_2, \ldots, \ell_n$. Let $\theta_1, \theta_2, \ldots, \theta_n$ be nonnegative values that sum to 2π. Then

$$\sum_{j=1}^{n} \frac{1}{4} \ell_j^2 \cot(\frac{\theta_j}{2}) \geq Area(P),$$

and equality holds if and only if P is inscribed in a circle and each θ_j equals the radial angle of the arc subtended by a chord of length ℓ_j.

5. Use Theorem B to prove the following. Let P be a simple (but not necessarily convex) n-gon, and let Q be an n-gon with that has the same side lengths as P, and is inscribed in a circle. Suppose that the interior of Q contains the center of its circumcircle. Then Q can be partitioned into $2n$ pieces that can be rearranged to cover all of P and its interior. (This covering will, in general, have overlaps and portions that extend beyond the borders of P.)

6. Prove Theorem B.

7. (Open question) Is there a three dimensional analog to Theorem B?

A lovely set of classical geometric optimization problems and related developments can be found in Courant and Robbins [3, p. 346–79].

References

[1] Benson, R.V. Euclidean Geometry and Convexity, McGraw-Hill, Inc., 1966.

[2] Croft, H.T., K.J. Falconer and R.K. Guy, Unsolved Problems in Geometry, Springer-Verlag New York, Inc., 1991.

[3] Courant, Richard, and Herbert Robbins. What is Mathematics? Oxford University Press, 1960.

[4] Hales, T.C. The honeycomb conjecture, Preprint, available at http://xxx.langl.gov/abs/math.MG/9906042

[5] Kazarinoff, N.D., Geometric Inequalities. Random House, 1961.

[6] Lawlor, G. A New Area Maximization Proof for the Circle, *Mathematical Intelligencer*, 20 (1999) 29–30.

[7] Morgan, F. The hexagonal honeycomb conjecture *Trans. of the AMS*, 351(5) (1999) 1753–1763.

[8] Lax, P.D. A Short Path to the Shortest Path, *American Mathematical Monthly*, 158–159, Feb., 1995.

[9] Osserman, R. The isoperimetric inequality, Bulletin of the American Mathematical Society, 84 (1978), 1182–1238.

[10] Pappus, of Alexandria. La collection mathematique; oeuvre traduite pour la premiere fois du grec en francais par Paul Ver Eecke. Paris, Bruges, Desclee, de Brouwer et cie, 1933.

[11] Rademacher, H. and O. Toeplitz. The Enjoyment of Mathematics, Princeton University Press, 1994.

[12] Siegel, A. An isoperimetric theorem in plane geometry. *J. Discrete and Computational Geometry*, 29(2) 239–55 (2003).

[13] Virgil (A.K.A. Publius Vergilius Maro). Aeneid I, 365–368. (19 B.C.)

[14] Yaglom, I.M., and V.G. Boltanskiĭ. Convex Figures, translated by P.J. Kelly and L.F. Walton, Holt, Rinehart and Winton, 1961.