SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The Planar Isoperimetric Theorem and Related Results

av

Richard Säldebring

2018 - No K17

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

The Planar Isoperimetric Theorem and Related Results

Richard Säldebring

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Christian Gottlieb

2018

Abstract

If the length of the perimeter of a figure is given, what is the greatest area that can be enclosed? This age-old question is called the isoperimetric problem. Its origins date back to antiquity but a thorough and complete solution was not offered until the 19th century. In this thesis we will reveal the solution to the isoperimetric problem and present some distinctly different ways in which one can arrive at a conclusive answer.

We will also examine a few variations of this problem. For instance, one could look at a pentagon and ask oneself what type of pentagon would maximise the area when the length of the perimeter is given. This would then fall under the isoperimetric problem for polygons. Moreover, we will explore some results that bear a resemblance to the original problem. Lastly we take a brief look at the isoperimetric problem in higher dimensions.

Acknowledgements

I would like to extend my warmest thanks to my supervisor Christian Gottlieb for the unceasing enthusiasm which he has shown and for the support he has given me throughout the entire writing process. Moreover, I also wish to express my gratitude to my referee Ralf Fröberg for calling my attention to a few details in need of adjustment.

Contents

1 Introduction

Geometry, merely mentioning the word evokes a multitude of shapes taking form before our eyes. Ever since the release of Euclid's *Elements* in the 4th century B.C., countless generations have been introduced to the mathematical beauty of shapes and figures.

With analytic geometry and calculus making their debut in the 17th century, one was able to solve problems that had been unsolved hitherto. In modern mathematics it might be said that Euclidean geometry on its own is seen as slightly archaic. There is, however, no denying that Euclidean geometry has a certain charm of its own and that its seeming simplicity is part of its appeal.

One of the earliest problems facing mathematicians was the *isoperimetric problem*. The word isoperimetric means of equal perimeters. The isoperimetric problem entails solving the following: *Which figure encloses the greatest area out of all figures with a given perimeter?* An equivalent way of stating the isoperimetric problem is: *Which figure has the shortest perimeter out of all figures with a given area?*

That these two questions turn out to have the same answer may not be obvious at first sight. Nevertheless we will show that this is the case in a proof by contradiction. Let us suppose that the answer to the latter problem and the latter problem only, is figure A. It then follows that there has to be a figure that has the same perimeter as A but whose area is greater. Let us call it figure B. By rescaling B, we can turn it into a figure with the same area as A but it will then have a shorter perimeter than A. This is a contradiction since we started out by stating that figure A was the solution having *the shortest perimeter* out of all figures with a given area. Thus it follows that the two ways of formulating the isoperimetric problem are equivalent. It should be noted, however, that the former way of stating the problem is the most common.

Yet, the question remains; which figure is the solution to the isoperimetric problem? It turns out that *the circle* is the correct answer. Although one might be able to arrive at the right answer by sheer intuition, actually proving that the circle is the answer turns out to be a little more complicated than hazarding an educated guess. Still, numerous proofs exist. We formulate the isoperimetric theorem as follows:

Theorem 1.1 [The Isoperimetric Theorem]
Of all the figures with a given perimeter, the circle has the greatest area. Or equivalently: Of all the figures with a given area, the circle has the shortest perimeter.

One can state the isoperimetric theorem as an inequality as well. We know that out of all figures with a given perimeter, P, the greatest area is given by $A = \pi r^2 = \frac{(2\pi r)^2}{4\pi} = \frac{P^2}{4\pi}$. We are thus now able to present the following inequality.

Theorem 1.2 [The Isoperimetric Inequality]
For any figure with perimeter P and area A, the following inequality holds:

$$4\pi A \leq P^2.$$

Equality holds for the circle only.

In what follows we will have a look at so-called convex figures since they will turn out to be highly important to us. In sections 4 and 5 we present proofs of the isoperimetric theorem for triangles and quadrilaterals. We then move on to section 6 where we will give a few proofs of theorem 1.1. After that the isoperimetric theorem for n-gons will be proven. Following that, some problems that are of an isoperimetric nature will be solved and we will also take a closer look at figures of constant width. Lastly, we will make a brief foray into the topic of the isoperimetric theorem in higher dimensions.

2 Geometric Terminology

When reading a text on geometry, one might come across words that seem vaguely familiar. The purpose of this section is therefore to compile a list of explanations of some of the terminology used throughout the thesis which may cause some confusion.

Equilateral

An equilateral polygon is a polygon whose sides are all of equal length.

Equiangular

An equiangular polygon is a polygon whose angles are all of equal magnitude.

Regular Polygon

A polygon is regular if and only if it is both equilateral and equiangular.

Radius of a polygon

The distance to any vertex from the center of a convex regular polygon.

Apothem

The distance to the midpoint of any side from the center of a convex regular polygon.

Congruency

Two geometrical figures are congruent if and only if they can become identical figures by translation, rotation and/or reflection. If A and B are congruent, we write $A \cong B$. Two angles are said to be congruent if and only if they are of the same size.

Foci

The plural form of *focus*. The foci are two points situated on the longest axis (the major axis) of an ellipse, and the distances from both foci to the center are equidistant. If we choose a point on the ellipse and draw one ray to each focus, the sum of the length of the rays will always be constant, no matter which point on the ellipse we select. Thus the ellipse consists of all of these points.

3 Convexity

The notion of convexity is an important one. Let us say we have a region, R. If one is able to draw a straight line, L, between any two points a and b on R, without L appearing outside of R, the region R is said to be convex. Any region that is not convex, is concave. In figure 1, (I) and (II) are concave and (III) and (IV) are convex.

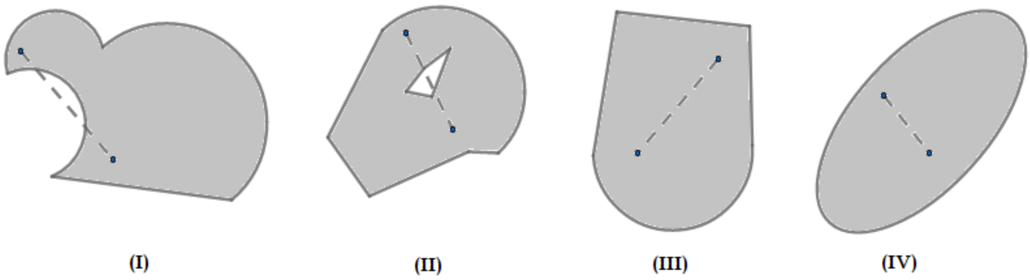


Figure 1

Any figure that is concave due to having a hole in its interior can be transformed into a convex figure with a shorter perimeter length but a greater area by simply filling in the hole. Additionally, any figure that is concave because of outer depressions can also be transformed into a convex figure. This is achieved by drawing a straight line between the two exact points at which the depression begins, resulting in a convex figure with a shorter perimeter length and a greater area. See figure 2.

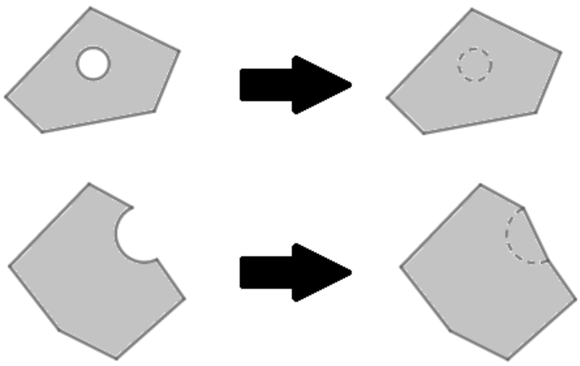


Figure 2

If we suppose that the figure that is the solution to the isoperimetric problem is concave, then making the figure convex as per the methods above will yield a figure with greater area and a shorter perimeter. Thus, in the search for the greatest area with a given perimeter for a given class of figures, be it quadrilaterals, n-gons or the standard isoperimetric problem, one only has to concern oneself with figures of a convex nature.

4 The Isoperimetric Theorem for Triangles

While one most often hears of the original isoperimetric theorem as stated above, one can still happen upon intriguing, albeit lesser-known versions of it. Shortly, we will have a look at one of those; the isoperimetric theorem for triangles. Before that however, a few preliminaries are required. We begin by introducing and proving a theorem from ancient times by Heron [7].

4.1 Preliminaries

Theorem 4.1 [Heron's Formula]
For any triangle with sides a, b and c, its area is given by

$$A = \frac{1}{4}\sqrt{((a+c)^2 - b^2)(b^2 - (a-c)^2)}. \tag{4.1.1}$$

Proof. We look at any triangle with sides a, b and c. See figure 3. Ultimately, what we want is the area expressed as a function of the side lengths.

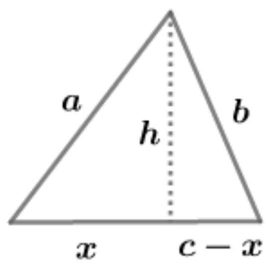


Figure 3

By the Pythagorean theorem we see that $a^2 = x^2 + h^2$ and $b^2 = (c-x)^2 + h^2$ hold for the left and right triangle, respectively. We thus acquire

$$b^2 = c^2 - 2cx + x^2 + h^2 = c^2 - 2cx + a^2. \tag{4.1.2}$$

Solving for x, yields

$$x = \frac{a^2 - b^2 + c^2}{2c}. \tag{4.1.3}$$

The area of the entire triangle is expressed as $A = \frac{1}{2}ch$. From this fact it follows that

$$4A^2 = c^2h^2 = c^2(a^2 - x^2) = c^2\left(a^2 - \left(\frac{a^2 - b^2 + c^2}{2c}\right)^2\right) = a^2c^2 - \frac{(a^2 - b^2 + c^2)^2}{4}. \tag{4.1.4}$$

Multiplying both sides by four, we get

$$\begin{aligned} 16A^2 = 4a^2c^2 - (a^2 - b^2 + c^2)^2 &= \Big(2ac + (a^2 - b^2 + c^2)\Big)\Big(2ac - (a^2 - b^2 + c^2)\Big) \\ &= \Big((a + c)^2 - b^2\Big)\Big(b^2 - (a - c)^2\Big). \end{aligned} \tag{4.1.5}$$

We finally obtain (4.1.1) by dividing each side by sixteen and then taking the square root of both sides. □

The following theorem will also prove useful. It is the well-known theorem of the arithmetic and geometric mean, often abbreviated as the AM-GM inequality. The proof is due to G. Pólya [4].

Theorem 4.2 [The AM-GM Inequality]
If $x_i \in \mathbb{R}^+$, $\forall i$ then the following relationship between the arithmetic and geometric mean holds:

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \prod_{i=1}^{n} x_i^{1/n}. \tag{4.1.6}$$

Equality holds only when $x_1 = x_2 = ... = x_n$.

Proof. We will make use of the function $exp(x-1) - x$. Its first derivative is $exp(x-1) - 1$ and differentiating twice yields $exp(x-1)$. Since the second derivative is greater than zero for all $x \in \mathbb{R}$, we know that the original function is convex everywhere. Furthermore, we see that the function attains its global minimum value of 0 when $x = 1$. Thus, the inequality
$0 \leq exp(x-1) - x \Leftrightarrow x \leq exp(x-1)$ holds.

Now let $f(x) = x$ and $g(x) = exp(x-1)$. The inequality can therefore simply be written as $f(x) \leq g(x)$. We now let a be the arithmetic mean and observe that the following inequality holds:

$$\prod_{i=1}^{n} \frac{f(x_i)}{a} \leq \prod_{i=1}^{n} \frac{g(x_i)}{a}. \tag{4.1.7}$$

The right-hand side can be written as

$$exp\left\{ \sum_{i=1}^{n} \left(\frac{x_i}{a} - 1 \right) \right\} = exp\left\{ \frac{1}{a} \sum_{i=1}^{n} x_i - n \right\} = exp(n-n) = 1. \tag{4.1.8}$$

Hence, (4.1.7) can be simplified to $\frac{\prod_n x_i}{a^n} \leq 1$. We thus finally end up with the AM-GM inequality:

$$\prod_{i=1}^{n} x_i^{1/n} \leq a. \tag{4.1.9}$$

□

4.2 Proving The Isoperimetric Theorem for Triangles

Now we are ready to prove the isoperimetric theorem for triangles. The following proof is due to Kazarinoff [7].

Theorem 4.3 [The Isoperimetric Theorem for Triangles]
Among all triangles with a given perimeter, the equilateral triangle has the greatest area.

Proof. Consider a triangle with perimeter p and sides s_1, s_2 and s_3. From theorem 4.1 we see that the greatest area is attained when

$$(p-2s_1)(p-2s_2)(p-2s_3) \tag{4.2.1}$$

is maximised. This can be seen by first observing that the factors under the radical sign in (4.1.1) can be written as

$$((s_1 + s_3) + s_2)((s_1 + s_3) - s_2)(s_2 + (s_1 - s_3))(s_2 - (s_1 - s_3)). \quad (4.2.2)$$

Secondly, we know that $p = s_1 + s_2 + s_3$ is fixed and therefore (4.2.1) follows. In order to maximise (4.2.1) we apply (4.1.6) with $n = 3$. We begin by letting $x_i = p - 2s_i$, $i = 1, 2, 3$. The AM-GM inequality thus states that $(x_1x_2x_3)^{1/3} \leq \frac{x_1+x_2+x_3}{3}$. The arithmetic mean can be rewritten as $p/3$. Hence we acquire the inequality $x_1x_2x_3 \leq (p/3)^3$. The left-hand side is clearly maximised when the two sides are equal, i.e. when $x_1 = x_2 = x_3 \Leftrightarrow p - 2s_1 = p - 2s_2 = p - 2s_3 \Leftrightarrow s_1 = s_2 = s_3$. That concludes the proof. □

5 The Isoperimetric Theorem for Quadrilaterals

Before proving the isoperimetric theorem for quadrilaterals we will prove the central angle theorem and two related results, both of which will be of use to us.

5.1 Preliminaries

Theorem 5.1 [The Central Angle Theorem and Some Related Results]
A central angle is defined to be any angle between two radii of a circle. Its vertex lies at the center of the circle. An inscribed angle is the angle between two chords that meet at a point on the circumference of the circle. The inscribed angle has its vertex on any point on the circumference of the circle, except on the circular arc on which it is subtended.

The central angle theorem states that the central angle has double the magnitude of any inscribed angle if they are subtended by the same circular arc. See figure 4.

As a consequence it follows that among all triangles sharing the same base and with an opposing angle of the same magnitude, the triangle with the greatest area is the one that is isosceles.

Additionally, as yet another consequence of the central angle theorem it follows that two opposite angles in an inscribed quadrilateral (a quadrilateral whose vertices lie on the circumference of a circle) always add up to π.

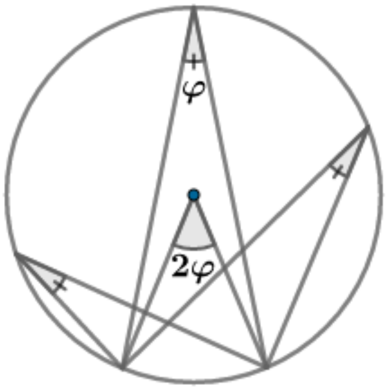


Figure 4

Proof. We will look at three different cases. See figure 5.

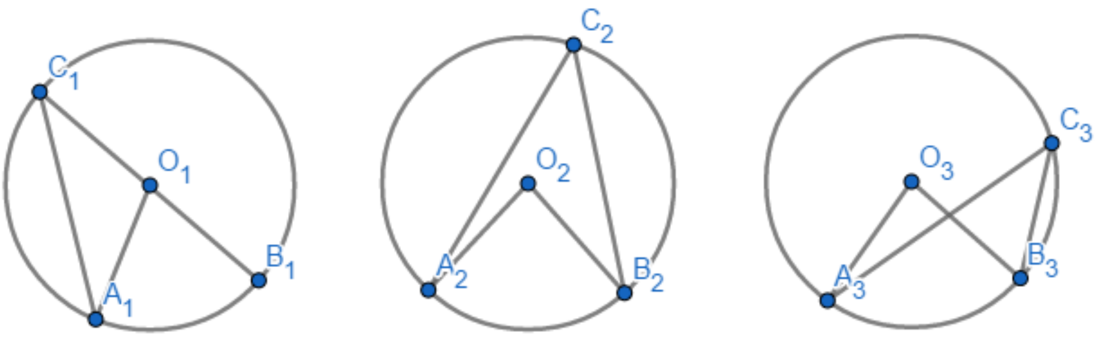


Figure 5

Case 1: We see that $\triangle O_1A_1C_1$ is isosceles since its legs are the radii of the circle. Thus $\angle O_1A_1C_1$ and $\angle O_1C_1A_1$ are congruent. We now observe that $\pi = \angle A_1O_1C_1 + \angle C_1A_1O_1 + \angle A_1C_1O_1 = \angle A_1O_1B_1 + \angle A_1O_1C_1$. Hence $\angle A_1O_1B_1 = \angle C_1A_1O_1 + \angle A_1C_1O_1 = 2\angle A_1C_1O_1$, which shows that the central angle has twice the magnitude of the inscribed angle.

Case 2: By drawing a diameter from C_2 we can use the same method as in the first case, twice. See figure 6. In other words we first show

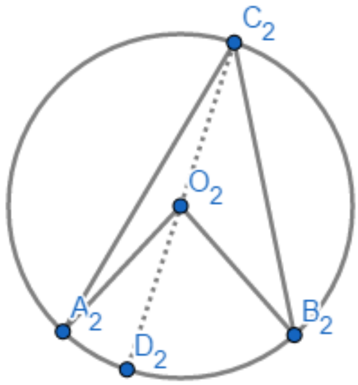


Figure 6

that $\angle A_2O_2D_2 = 2\angle A_2C_2O_2$ and then analogously show that $\angle D_2O_2B_2 = 2\angle B_2C_2O_2$. Finally, since $\angle A_2C_2B_2 = \angle A_2C_2O_2 + \angle B_2C_2O_2$, it holds that $\angle A_2O_2B_2 = \angle A_2O_2D_2 + \angle D_2O_2B_2 = 2\angle A_2C_2O_2 + 2\angle B_2C_2O_2 = 2\angle A_2C_2B_2$.

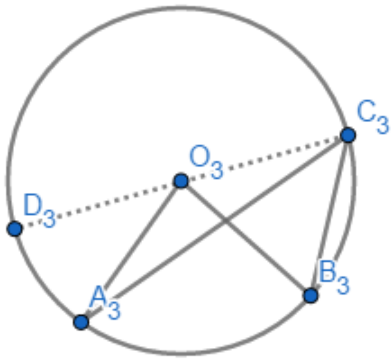


Figure 7

Case 3: We draw a diameter from C_3 and yet again use the method from case 1. See figure 7. First we show that $\angle D_3O_3B_3 = 2\angle D_3C_3B_3$ and then that $\angle D_3O_3A_3 = 2\angle D_3C_3A_3$ by using the preceding method. Hence $\angle A_3O_3B_3 = \angle D_3O_3B_3 - \angle D_3O_3A_3 = 2(\angle D_3C_3B_3 - \angle D_3C_3A_3) = 2\angle A_3C_3B_3$, which is exactly what we wanted to show.

Having shown all three cases, the central angle theorem is thus proved.

Furthermore, one can draw a chord and let it be the base of a triangle whose vertex is placed on the circumference of the circle that is not on the circular arc subtending the top angle of the triangle. Since the height of the triangle is at its peak when the highest point of the triangle lies directly above the central point of the circle, it follows that of all triangles having the same base length and opposing angle of the same magnitude, the isosceles triangle has the greatest area. See figure 8.

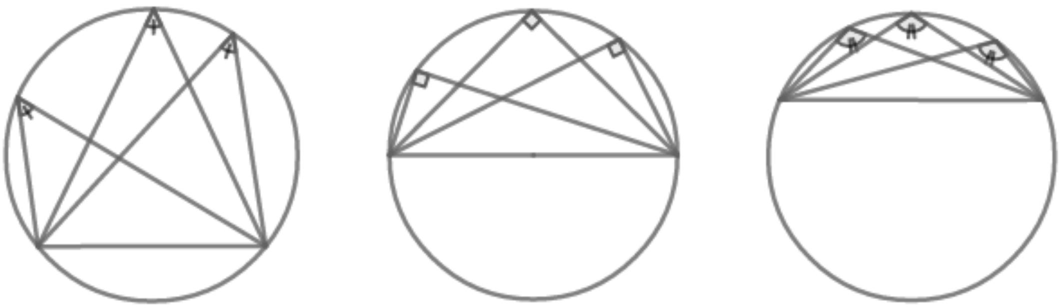


Figure 8

Now we will prove that two angles opposite of one another in an inscribed quadrilateral always add up to π. We inscribe the quadrilateral $ABCD$ in a circle and let O be the midpoint of the circle. See figure 9. Let $\angle BAD = \theta$. It then follows from the central angle theorem that the central angle subtended by the same circular arc equals 2θ. We can reason analogously for the angles ϕ and 2ϕ. Since $2\theta + 2\phi = 2\pi$ it follows that $\theta + \phi = \pi$. Thus the two remaining angles of the quadrilateral also add up to π. $\square$

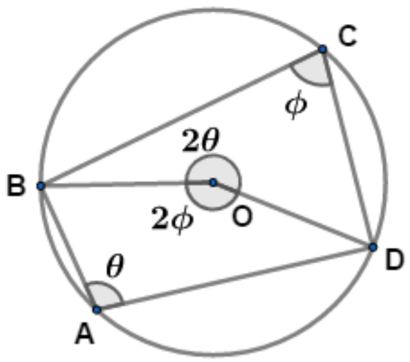


Figure 9

5.2 Proving The Isoperimetric Theorem for Quadrilaterals

Now we will present Kazarinoff's proof of the isoperimetric theorem for quadrilaterals [7]. It turns out that the square solves this isoperimetric problem, i.e. that the following theorem holds:

Theorem 5.2 [The Isoperimetric Theorem for Quadrilaterals]
Out of all quadrilaterals with equal perimeters, the square has the greatest area.

Proof. Let a convex quadrilateral be given as in figure 10 with sides of length s_1, s_2, s_3 and s_4. We also let its perimeter be P and its area be A. Now, let the line of length t divide the quadrilateral into two triangles. The height of the triangles are $s_1 sin(\varphi)$ and $s_4 sin(\theta)$, respectively. Thus the area of the leftmost triangle is $\frac{1}{2}s_1 s_2 sin(\varphi)$, and the area of the rightmost triangle is $\frac{1}{2}s_3 s_4 sin(\theta)$.

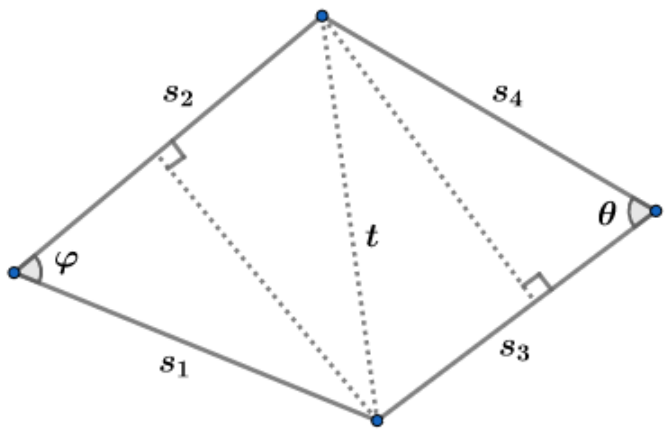


Figure 10

The area of the quadrilateral is therefore $A = \frac{1}{2}s_1 s_2 sin(\varphi) + \frac{1}{2}s_3 s_4 sin(\theta)$. It follows that $4A = 2s_1 s_2 sin(\varphi) + 2s_3 s_4 sin(\theta)$. Squaring both sides yields

$$16A^2 = 4s_1^2 s_2^2 sin^2(\varphi) + 8s_1 s_2 s_3 s_4 sin(\varphi)sin(\theta) + 4s_3^2 s_4^2 sin^2(\theta). \quad (5.2.1)$$

Through the law of cosines we see that

$$t^2 = s_1^2 + s_2^2 - 2s_1s_2cos(\varphi) = s_3^2 + s_4^2 - 2s_3s_4cos(\theta), \tag{5.2.2}$$

which means that

$$s_1^2 + s_2^2 - s_3^2 - s_4^2 = 2s_1s_2cos(\varphi) - 2s_3s_4cos(\theta). \tag{5.2.3}$$

Squaring both sides yet again, we see that

$$(s_1^2 + s_2^2 - s_3^2 - s_4^2)^2 = 4s_1^2s_2^2cos^2(\varphi) - 8s_1s_2s_3s_4cos(\varphi)cos(\theta) + 4s_3^2s_4^2cos^2(\theta). \tag{5.2.4}$$

Adding (5.2.1) and (5.2.4) together, we acquire

$$\begin{aligned}16A^2+(s_1^2+s_2^2-s_3^2-s_4^2)^2 = 4s_1^2s_2^2(sin^2(\varphi)+cos^2(\varphi))+4s_3^2s_4^2(sin^2(\theta)+cos^2(\theta)) \\ - 8s_1s_2s_3s_4(cos(\varphi)cos(\theta) - sin(\varphi)sin(\theta)).\end{aligned} \tag{5.2.5}$$

We now make use of the following two trigonometric identities:

$$sin^2(\alpha) + cos^2(\alpha) = 1,$$

and

$$cos(\alpha + \beta) = cos(\alpha)cos(\beta) - sin(\alpha)sin(\beta),$$

and thus obtain

$$16A^2 + (s_1^2 + s_2^2 - s_3^2 - s_4^2)^2 = 4s_1^2s_2^2 + 4s_3^2s_4^2 - 8s_1s_2s_3s_4cos(\varphi + \theta). \tag{5.2.6}$$

If the side lengths are all fixed, we see that the greatest area is acquired when $cos(\varphi + \theta) = -1$. This is the case when $\varphi + \theta = \pi$. This calls to mind an inscribed quadrilateral in which two opposing angles always add up to π. See theorem 5.1. Therefore we see that a quadrilateral whose side lengths are given, and which can be inscribed in a circle has the greatest area.

In order to prove the isoperimetric theorem for quadrilaterals it is thus sufficient to only look at the quadrilaterals that can be inscribed in a circle. So we keep the perimeter P of the quadrilateral fixed and let the two pairs of opposing angles add up to π, respectively. Now we permit the side lengths to vary, keeping in mind that the aforementioned constraints hold. Since $cos(\varphi + \theta) = -1$, (5.2.6) can be written as

$$
\begin{aligned}
&16A^2 + (s_1^2 + s_2^2 - s_3^2 - s_4^2)^2 = 4s_1^2 s_2^2 + 4s_3^2 s_4^2 + 8s_1 s_2 s_3 s_4 \\
&\Rightarrow 16A^2 = 4(s_1^2 s_2^2 + s_3^2 s_4^2) + 8s_1 s_2 s_3 s_4 - (s_1^2 + s_2^2 - s_3^2 - s_4^2)^2 = 4(s_1 s_2 + s_3 s_4)^2 - (s_1^2 + s_2^2 - s_3^2 - s_4^2)^2 \\
&= [2(s_1 s_2 + s_3 s_4) + (s_1^2 + s_2^2 - s_3^2 - s_4^2)][2(s_1 s_2 + s_3 s_4) - (s_1^2 + s_2^2 - s_3^2 - s_4^2)] \\
&= [(s_1 + s_2)^2 - (s_3 - s_4)^2][(s_3 + s_4)^2 - (s_1 - s_2)^2] \\
&= [(s_1 + s_2) + (s_3 - s_4)][(s_1 + s_2) - (s_3 - s_4)][(s_3 + s_4) + (s_1 - s_2)][(s_3 + s_4) - (s_1 - s_2)] \\
&= (P - 2s_1)(P - 2s_2)(P - 2s_3)(P - 2s_4). \quad (5.2.7)
\end{aligned}
$$

Using the AM-GM inequality for $n = 4$, we see that the following inequality holds

$$((P-2s_1)(P-2s_2)(P-2s_3)(P-2s_4))^{\frac{1}{4}} \leq \frac{(P-2s_1) + (P-2s_2) + (P-2s_3) + (P-2s_4)}{4}. \quad (5.2.8)$$

The right-hand side of (5.2.8) can be rewritten as $\frac{4P-2(s_1+s_2+s_3+s_4)}{4} = \frac{4P-2P}{4} = \frac{P}{2}$, yielding

$$((P-2s_1)(P-2s_2)(P-2s_3)(P-2s_4))^{\frac{1}{4}} \leq \frac{P}{2}. \quad (5.2.9)$$

Equality holds when

$$(P - 2s_1) = (P - 2s_2) = (P - 2s_3) = (P - 2s_4),$$

and that is the case when all sides are of equal length, i.e. when

$$s_1 = s_2 = s_3 = s_4.$$

This holds true when the quadrilateral is a square. Hence the proof is complete. $\square$

6 Proofs of The Isoperimetric Theorem

As mentioned earlier, there is an abundant number of proofs for the isoperimetric theorem. The isoperimetric problem was known to the Greeks of ancient times. Pappus wrote down a proof of the isoperimetric theorem in the 4th century A.D. He accredits Zenodorus, who lived in the 2nd century B.C., in regard to this result. These proofs were not, however, rigorous mathematical proofs by the standard of today.

The proofs that will be presented in the following are all from the 19th century onward. We have chosen a select few that are all interesting in their own way.

6.1 Steiner's Attempt At a Proof

In modern times, the pursuit to come up with a rigorous proof of the isoperimetric theorem was commenced by the Swiss mathematician Jakob Steiner (1796-1863). He eventually came up with five proofs, the first of which was published in 1841 [13]. Due to his aversion to calculus, he was adamant in seeking to come up with proofs that were purely geometrical. However, it turns out they were incomplete inasmuch as they failed to ascertain that a figure of maximum area actually exists. What Steiner proved was that *if* a figure of greatest area exists, it has to be the circle.

An analogy that shows just how absurd a theorem one can come up with if an extremum is presumed to exist is the following: We want to show that 1 is the largest integer. Now, for all integers $n \neq 1$, there is an integer $n^2 > n$. Therefore 1 is the largest integer.

In the case of the isoperimetric theorem, it turned out that the alleged assertion that the circle maximised the area for a given perimeter was actually true. The problem of proving the isoperimetric theorem and the existence of a maximum was solved in 1879 by Karl Weierstrass [14]. Today many proofs exist, displaying a variety of ingenious ways in which to prove the formerly so elusive theorem.

Shortly we will present one of Steiner's proofs from [2]. In order for the proof to make sense one has to be familiar with the following theorem.

Theorem 6.1 [The Greatest Area of a Triangle with Two Given Sides]
If two line segments of length s and s' are to be used as sides in a triangle, the greatest area is achieved when they form the catheti in a right triangle.

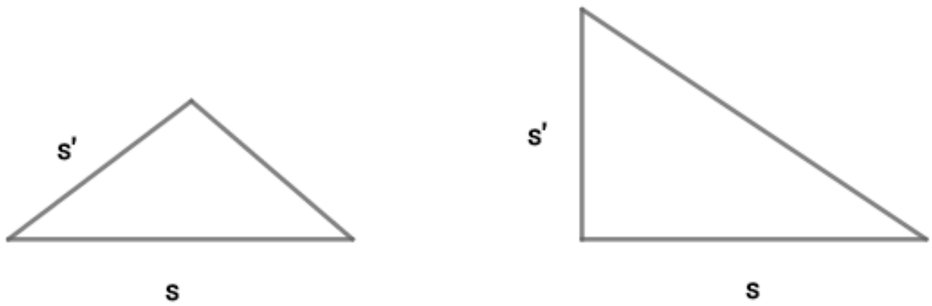


Figure 11

Proof. The area of a triangle is $\frac{1}{2}bh$ where b is the base and h is the height. In figure 11 we let s be the base and note that the height is maximised when the triangle is a right triangle which concludes the proof. □

Incomplete albeit elegant, only making use of Euclidean geometry, we now present one of Steiner's proofs:

Theorem 6.2 [Steiner's Attempted Proof]
Out of all the regions with a given perimeter that is not a circle, one can always find a figure that has the same perimeter but a greater area.

Proof. Let us assume that we already have the figure, F, with maximum area and that the length of its perimeter is l. F obviously has to be convex. We now proceed to draw a line, L, across F, which separates F into two figures, each having a perimeter of length $\frac{1}{2}l$. The two figures must necessarily have the same area. If that were not the case, we could just reflect the one figure with the greatest area across the aforementioned line, thereby creating a figure with the same perimeter l but a greater area. This would then contradict our assumption that F has the greatest area.

We now look at one of the halves of F and assume that it is not a semicircle. By drawing two lines, $\overline{AC}$ and $\overline{BC}$, there has to be a point on the perimeter on which $\angle ACB \neq \frac{\pi}{2}$. See figure 12. The resulting figure, G, consists of three regions: a triangle and two regions that can be thought of as being glued on to two of the sides of the triangle. Imagine sliding point A and B along L until $\angle ACB = \frac{\pi}{2}$. This enlarges the area of the triangle due to theorem 6.1 but keeps the area of the two glued on regions the same. Thus the area of G is increased.

By reflecting G over L, we now obtain a figure whose area is greater than F. This is a contradiction since F was assumed to be the figure with the greatest area. Therefore, G had to have been a semicircle and thus F must have been a circle.

□

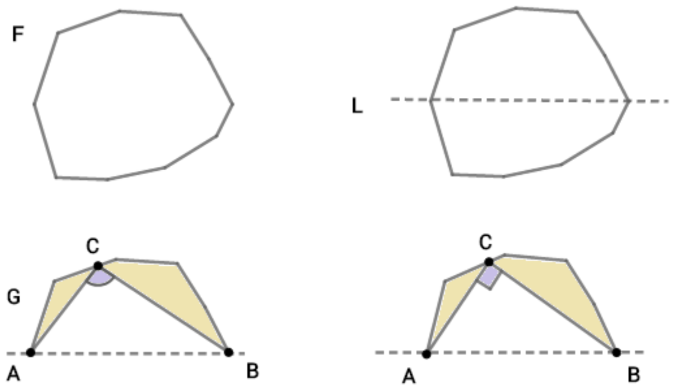


Figure 12

6.2 A Proof Using Elementary Geometry

The following proof is purely geometrical and it was published in 1998 by G. Lawlor [9]. We will look at the area of figures in the first quadrant, i.e. a figure whose area is hemmed in by the x- and y-axes and a smooth curve. The smooth curve will have the length $\pi/2$. The major part of the proof will consist of showing that out of all those figures, the quarter circle has a greater area than all the others. It will be sufficient to do so, since we can reflect this figure about the axes in two steps. At the very end, we reflect the quarter circle about the y-axis to end up with the semicircle. We then reflect the semicircle across the x-axis to finally get the full circle.

So, to start off we draw a curve between a point on the x-axis and a point on the y-axis. We let the figure thus encompassed by the axes and the smooth curve be F. It should be noted that we let F be a convex figure. We then place n points on the curve so that they are all equidistant from one another. Here, n is a large number. In total, we thus have $n+2$ points that are all equally distant from each other. We let the point on the x-axis be P_1. Traversing the curve towards the point on the y-axis, we let P_2 be the next point that we reach and finally let the point on the y-axis be P_{n+2}.

From all points except the ones lying on the axes we draw a ray; this is done so that the angle subtended by the x-axis and the ray emanating from P_i has the magnitude $\frac{\pi(i-1)}{2(n+1)}$. We have now partitioned F into $n+1$ sections, noting that two consecutive rays form what resembles a triangle. The ray emanating from P_k will be defined as R_k. As can be seen from the left figure in figure 13, the sections might overlap one another.

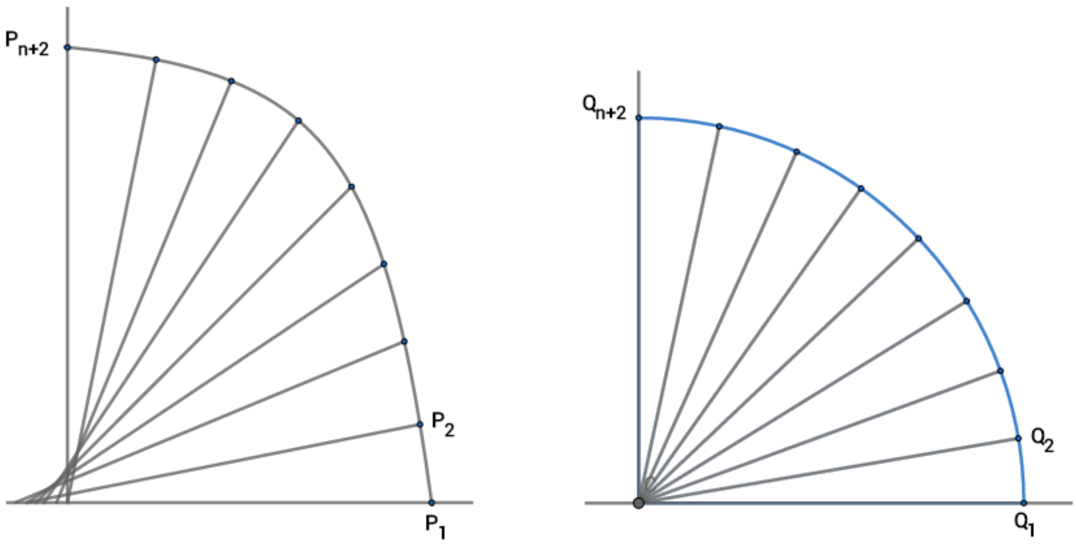


Figure 13

We now have to show that the triangle-like partitions cover all of F. In order to show this we will have to show that for any point in the interior of F, there is (at least) one partition covering it. It is clear that a point inside of F will be enclosed by the curve and the axes. Since F is convex and R_{i+1} subtends a greater angle with the x-axis than R_i, two consecutive rays eventually have to intersect one another, whether it be inside or outside of F. For any point p, the ray immediately above it, trapping p from above, has to be one of the rays R_j, $j = 2, ..., n+2$. As for the ray immediately below p; one of the rays R_k, $k = 1, ..., n+1$ must necessarily then be the ray that encloses p from beneath. That way, a point in the interior of F will always be covered by at least one of the partitions. Thus the partitions cover the entirety of F.

We have previously alluded to the fact that the partitions seem to resemble triangles. By letting the number of points, $n \to \infty$, the curved segment P_iP_{i+1} can be seen as a straight line. Thus the partitions have the line segment P_iP_{i+1} as their base and can be regarded as triangles, $\triangle T_i$, $1 \leq i \leq n+1$. In doing so, a tiny part of each partition, H_i, $1 \leq i \leq n+1$, will not be covered by the triangles. Since $\lim_{i\to\infty} \sum_{i=1}^{n+1} Area(H_i) = 0$, we now only look at the triangles. These triangles may overlap one another as stated above.

Now let C (the right-hand figure in figure 13) be the part of the unit circle centered at the origin that lies in the first quadrant. By partitioning C in the same way, by the points Q_i as outlined above including letting $n \to \infty$ we see that it now consists of $n+1$ isosceles triangles, $\triangle U_i$, $1 \leq i \leq n+1$.

The triangles covering F have the same base as the isosceles triangles covering C. Furthermore, the rays emanating from the points P_i and Q_i subtend the same angle with the x-axis. We know from theorem 5.1 that among all triangles sharing the same base and having an opposing angle of the same magnitude, the triangle that is isosceles has the greatest area. Therefore $\sum_{i=1}^{n+1} Area(\triangle U_i) > \sum_{i=1}^{n+1} Area(\triangle T_i)$. We know that the triangles $\triangle T_i$ cover at least all of F and as per the technique of reflection outlined above, the proof is complete.

6.3 A Proof Using Calculus

Another interesting proof, due to P.D. Lax [10], utilizes calculus to prove the isoperimetric theorem.

As we saw earlier in theorem 1.2, the isoperimetric inequality states that $4\pi A \leq P^2$, with equality only for the circle. The aforementioned inequality thus states that a closed curve of length 2π encompasses an area that is $\leq \pi$. Only when the curve forms the unit circle is the maximum area, π, attained. We now parameterize the curve, defining t to be the arc length and letting the curve be defined by the points $(x(t), y(t))$, $0 \leq t \leq 2\pi$. The curve is placed so that the coordinates $(x(0), y(0))$ and $(x(\pi), y(\pi))$ both lie on the x-axis.

The *area* of the resulting shape can be expressed, by the use of Green's theorem, as two integrals in the following way:

$$Area = \iint_C 1\, dA = \oint_C y\, dx = \int_0^{2\pi} y(t)x'(t)\, dt = \int_0^{\pi} y(t)x'(t)\, dt + \int_{\pi}^{2\pi} y(t)x'(t)\, dt = I_1 + I_2.$$

Since the area of the entire figure is proposed to be $\leq \pi$ it suffices to show that both I_1 and I_2 have to be $\leq \pi/2$. If, in addition, we can show that the equality holds solely when the parameterization is that of a circle, the proof will be complete.

We are now going to make use of the following inequality:

$$ab \leq \frac{a^2 + b^2}{2}.$$

This easily follows since $(a - b)^2 \geq 0 \Leftrightarrow a^2 - 2ab + b^2 \geq 0 \Leftrightarrow ab \leq \frac{a^2+b^2}{2}$. Looking at I_1 and implementing the inequality we get

$$I_1 = \int_0^{\pi} y(t)x'(t)\, dt \leq \frac{1}{2}\int_0^{\pi} (y(t)^2 + x'(t)^2)\, dt. \tag{6.3.1}$$

We now use the fact that t is the arc length to see that

$$(dt)^2 = (dx)^2 + (dy)^2 \Rightarrow \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 = 1 \Rightarrow x'(t)^2 + y'(t)^2 = 1.$$

By replacing $x'(t)^2$ with $1-y'(t)^2$ we rewrite the integrand on the right-hand side in (6.3.1) so that

$$I_1 \leq \frac{1}{2}\int_0^\pi (y(t)^2 + 1 - y'(t)^2)\, dt. \tag{6.3.2}$$

We now rewrite $y(t)$ as a product of the factors $g(t)$ and $sin\, t$. Of note is that g is a bounded and differentiable function. We get $y(t) := g(t)sin\, t$ and after differentiating we acquire $y'(t) = g'(t)sin\, t + g(t)cos\, t$.
Inserting both y and its derivative into (6.3.2) we get

$$I_1 \leq \frac{1}{2}\int_0^\pi (g(t)^2 sin^2 t + 1 - (g'(t)sint + g(t)cos\, t)^2)\, dt. \tag{6.3.3}$$

The integral in (6.3.3) can now be simplified. In what follows, note that the last equality follows by integration by parts of $\int_0^\pi g(t)^2 cos\, 2t\, dt$. This eliminates the two lattermost integrals.

$$\int_0^\pi (g(t)^2 sin^2 t + 1 - (g'(t)sint + g(t)cos\, t)^2)\, dt = \int_0^\pi (1 - g'(t)^2 sin^2\, t)\, dt +$$
$$+ \int_0^\pi g(t)^2(sin^2 t - cos^2 t)\, dt \;-\; 2\int_0^\pi g(t)g'(t) sin\, t\, cos\, t\, dt = \int_0^\pi (1 - g'(t)^2 sin^2\, t)\, dt -$$
$$- \int_0^\pi g(t)^2 cos\, 2t\, dt \;-\; \int_0^\pi g(t)g'(t) sin\, 2t\, dt = \pi - \int_0^\pi g'(t)^2 sin^2\, t\, dt. \tag{6.3.4}$$

Thus the inequality is reduced to

$$I_1 \leq \frac{\pi}{2} - \frac{1}{2}\int_0^\pi g'(t)^2 sin^2\, t\, dt. \tag{6.3.5}$$

The integrand, $g'(t)^2 sin^2\, t$, is composed of two factors that are both ≥ 0, $0 \leq t \leq 2\pi$. Hence we conclude that $I_1 \leq \frac{\pi}{2}$. Only when $g'(t) = 0$ do we get $I_1 = \frac{\pi}{2}$. Since we defined $y(t) := g(t)sin\, t$, we see that if $g'(t) = 0$ then $y(t) = k\, sin\, t$, where k is a constant.

Furthermore, only when $y(t) = x'(t) = \sqrt{1 - y'(t)^2}$ do we have equality in (6.3.1). This is the case solely when $k = \pm 1$ and so $y(t) = \pm sin\, t$. This, in turn, means that $x(t) = \pm cos\, t + c$, where c is a constant. This parameterization describes a semicircle whose diameter lies on the x-axis. In order to prove that $I_2 \leq \frac{\pi}{2}$ and that the equality only holds when $x(t)$ and $y(t)$ trace out a semicircle, one can reason analogously. This then completes the proof.

7 The Isoperimetric Theorem for n-gons

What if we look at polygons with more than four vertices? Having looked at the case when $n = 3$ and $n = 4$ one might suspect that the regular polygon is the one that maximises the area, given a certain perimeter. This actually turns out to be the case, as we will see shortly.

A most intriguing aspect of proving the isoperimetric theorem for n-gons is that no one seems to have managed to do so without the use of the original isoperimetric theorem [7]. The proof that will be presented shortly is no exception.

7.1 Preliminaries

First we will present three useful results. They will then be used to finally prove the isoperimetric theorem for n-gons. The proof of theorem 7.1 is due to [6].

Theorem 7.1
A regular polygon has an area that is equal or greater than an equilateral polygon with the same number of sides and with sides of equal length as the regular polygon.

Proof. Let P_1 be a polygon inscribed in the circle, O. Also, let P_2 be a polygon that has an equal number of sides, and whose sides are of equal length and with the sides in the same order as P_1. The area of O is comprised of the area of the polygon as well as the area of the circular segments. If we take the segments and fasten them to the corresponding sides of P_2 we obtain a new figure with the same perimeter as O. Let this new shape be called F.

It follows from the original isoperimetric theorem (theorem 1.1) that $Area(O) > Area(F)$. Furthermore, since the segments enclosing O and F are identical it follows that $Area(P_1) > Area(P_2)$. We know that a regular polygon can be inscribed in a circle and thus the proof is complete. □

Theorem 7.2
Out of two regular polygons that are of the same perimeter, the one with the most number of sides has the greatest area.

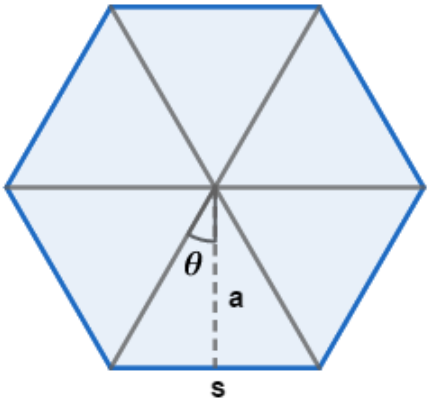


Figure 14

Proof. We consider a regular n-gon with perimeter of length p that has been partitioned into n congruent isosceles triangles with base length s. The apothem, a, is equal to the height of the triangles. Moreover we observe that $p = ns$. See figure 14. As for the indicated angle we see that $\theta = \frac{2\pi}{n} \cdot \frac{1}{2}$. The height of the triangles can therefore be expressed as $a = \frac{s}{2\,tan\,(\pi/n)}$. The area of one triangle is then equal to $\frac{s^2}{4\,tan\,(\pi/n)}$. Since we have n congruent triangles, the area of the whole n-gon is

$$n\frac{s^2}{4\,tan\,\frac{\pi}{n}} = \frac{p^2}{4} \cdot \frac{1}{n\,tan\,\frac{\pi}{n}}. \tag{7.1.1}$$

Now let $x = \pi/n$, $0 < x \leq \pi/3$. Thus the area can be expressed as

$$\frac{p^2}{4\pi} \cdot \frac{x}{tan\,x}. \tag{7.1.2}$$

It is clear that as n increases, x decreases. Hence, if we show that (7.1.2) increases as x decreases, the proof will be complete.
It suffices to only look at the latter factor in (7.1.2), i.e. $f(x) = \frac{x}{tan\,x}$. Differentiating this function yields

$$f'(x) = \frac{tan\,x - \frac{x}{cos^2 x}}{tan^2 x} = \frac{sin\,x\,cos\,x - x}{sin^2 x} = \frac{sin\,2x - 2x}{2sin^2 x}. \tag{7.1.3}$$

The denominator is a square and is therefore always positive. The numerator, on the other hand, is always negative for all x, $0 < x \leq \pi/3$. Thus $f'(x) < 0$ for all x in $0 < x \leq \pi/3$. In other words, as x decreases (n increases) the area increases. That concludes the proof. □

Theorem 7.3

Of two triangles having the same base and the same perimeter, the one whose legs have the smallest difference in their lengths has the largest area. Thus, of all triangles having the same base and the same perimeter, the isosceles triangle has the greatest area. Furthermore, out of two triangles whose bases are identical and whose perimeters are of equal length, the one whose base angles have the smallest difference in magnitude has the greatest area.

Proof. An ellipse can be defined as all the points to which we draw two rays, with the sum of the rays being constant, from the foci. Each pair of rays, together with the line segment AB, create a triangle. All of these triangles share the same base, AB. In addition, the triangles thus created, have perimeters of equal length.

We clearly see from figure 15 that $|AD - BD| < |AE - BE|$, since $|AD| < |AE|$ and $|BD| > |BE|$. One can reason analogously when comparing $\triangle ABD$ with the isosceles $\triangle ABC$ to see that $|AC - BC| < |AD - BD|$. In fact, as the points on the ellipse move closer towards C, we see that the difference of the side lengths decrease and eventually at C, $|AC - BC| = 0$. Hence, the smaller the difference between the sides, the greater the area of the triangle.

As for the angles, the same reasoning holds. Thus, we see that the smaller the difference between the base angles, the greater the area of the triangle.

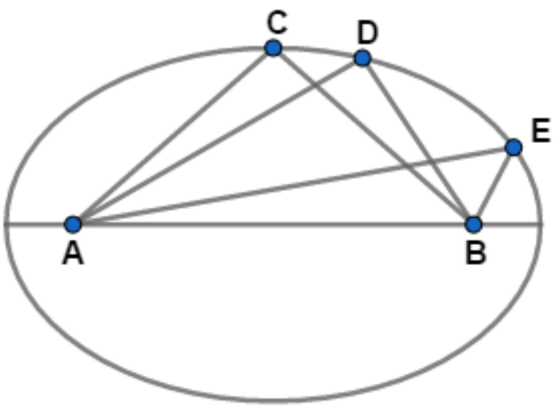


Figure 15

For the isosceles triangle, $|AC - BC| = |\angle ABC - \angle BAC| = 0$ and thus it has the greatest area out of all triangles with the same base and equal perimeter lengths. □

7.2 Proving The Isoperimetric Theorem for n-gons

Now we are finally ready to prove the isoperimetric theorem for n-gons. The proof comes from [6].

Theorem 7.4 [The Isoperimetric Theorem for n-gons]
Out of all n-gons with fixed perimeter, the regular n-gon has the greatest area.

Proof. This will be a proof by induction. The base case, i.e. $n = 3$, has been verified in theorem 4.3. Our induction hypothesis is that the area of any convex n-gon with a given perimeter is smaller than the area of a regular n-gon of the same perimeter.

The outline of the proof is as follows: Firstly we will prove that for any convex (n+1)-gon, one of the following two assertions hold.
i) The area of any convex (n+1)-gon of a given perimeter is not greater than the area of some equilateral (n+1)-gon of the same perimeter.
ii) The area of any convex (n+1)-gon of a given perimeter is not greater than the area of some n-gon of the same perimeter.

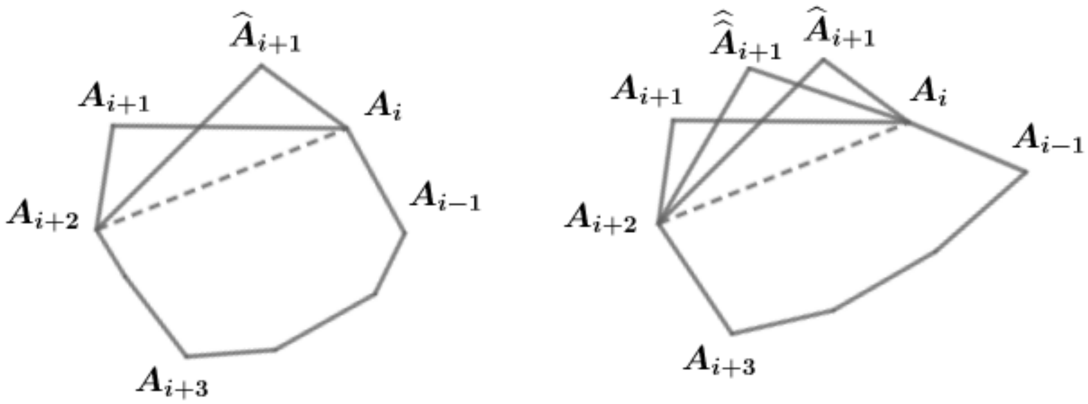
When we have proved that either *i)* or *ii)* holds, we will make use of theorem 7.1 and theorem 7.2. In the end it will follow that the area of a regular (n+1)-gon of a given perimeter is greater than the area of any convex (n+1)-gon of the same perimeter. Thus we will have proven exactly what we set out to prove from the outset.

We now start off the proof, our aim being to prove that for any convex (n+1)-gon, either *i)* or *ii)* holds. We regard the convex (n+1)-gon $A_1A_2...A_{n+1}$. See (I) in figure 16. Now we look at two adjacent sides of the polygon, say A_iA_{i+1} and $A_{i+1}A_{i+2}$. We proceed to draw the line segment A_iA_{i+2} and let this be the base of $\triangle A_iA_{i+1}A_{i+2}$. Now we create a new triangle, $\triangle A_i\widehat{A}_{i+1}A_{i+2}$, that has the same base as the aforementioned triangle. In addition to that we let $\triangle A_iA_{i+1}A_{i+2} \cong \triangle A_i\widehat{A}_{i+1}A_{i+2}$. In other words, we have acquired an (n+1)-gon where the two sides A_iA_{i+1} and $A_{i+1}A_{i+2}$ have swapped places with each other.

What if the resulting (n+1)-gon ends up being concave? See (II) in figure 16. In that case we will transform it into a convex n-gon by eliminating one vertex. We start by extending one of the two sides next to A_iA_{i+1} or $A_{i+1}A_{i+2}$ that formed part of the original (n+1)-gon. In other words, we either extend $A_{i+2}A_{i+3}$ or $A_{i-1}A_i$. Let us extend the side $A_{i-1}A_i$ by adding the segment $A_i\widehat{\widehat{A}}_{i+1}$ on to it. The length of the segment is chosen so that $A_i\widehat{\widehat{A}}_{i+1} + \widehat{\widehat{A}}_{i+1}A_{i+2} = A_iA_{i+1} + A_{i+1}A_{i+2}$. Thus the perimeter of $\triangle A_iA_{i+1}A_{i+2}$ and $\triangle A_i\widehat{\widehat{A}}_{i+1}A_{i+2}$ are equal since they share the same base. The extended side will end up crossing $\triangle A_i\widehat{A}_{i+1}A_{i+2}$, since if that were not the case $\triangle A_i\widehat{\widehat{A}}_{i+1}A_{i+2}$ would end up fully contained within $\triangle A_i\widehat{A}_{i+1}A_{i+2}$ and they could not then have equal perimeters.

The base angles of $\triangle A_i\widehat{\widehat{A}}_{i+1}A_{i+2}$ are smaller than $\angle A_iA_{i+2}A_{i+1} = \angle A_{i+2}A_i\widehat{A}_{i+1}$ and greater than $\angle A_iA_{i+2}\widehat{A}_{i+1} = \angle A_{i+2}A_iA_{i+1}$. Therefore $|\angle A_iA_{i+2}\widehat{\widehat{A}}_{i+1} - \angle A_{i+2}A_i\widehat{\widehat{A}}_{i+1}| < |\angle A_iA_{i+2}A_{i+1} - \angle A_{i+2}A_iA_{i+1}|$ and hence from theorem 7.3 we see that $Area(\triangle A_i\widehat{\widehat{A}}_{i+1}A_{i+2}) > Area(\triangle A_iA_{i+1}A_{i+2})$. So by replacing $\triangle A_iA_{i+1}A_{i+2}$ with $\triangle A_i\widehat{\widehat{A}}_{i+1}A_{i+2}$ we have now transformed the (n+1)-gon into an n-gon of the same perimeter but with greater area.

To summarize, in the end we either end up with an (n+1)-gon that is convex and of the same area with two of its sides interchanged *or* we acquire a convex n-gon whose area is greater. In the latter case we stop. If the former



(I) **(II)**

Figure 16

scenario happens, we can rearrange the sides of the (n+1)-gon until we have obtained an (n+1)-gon with its largest and smallest side adjacent to one another. In the case that there is no smallest side and/or no greatest side, we simply rearrange the sides so that one of the smallest and/or one of the greatest sides are placed next to each other.

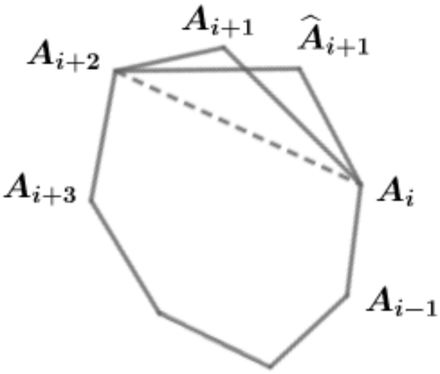


Figure 17

Having done that we now look at the (n+1)-gon in which the smallest and largest side are next to each other (alternatively one of the smallest or one of the greatest sides). We assume that these sides are $A_{i+1}A_{i+2}$ and A_iA_{i+1}. See figure 17. We create the line segment A_iA_{i+2} and let this be the base of $\triangle A_iA_{i+1}A_{i+2}$. Then we create $\triangle A_i\widehat{A}_{i+1}A_{i+2}$ which shares the same base as the triangle just mentioned and also let it be of the same perimeter. The perimeter of the (n+1)-gon being p, one of the sides of $\triangle A_i\widehat{A}_{i+1}A_{i+2}$ is created so that its length is $p/(n+1)$.

The smallest side of $\triangle A_iA_{i+1}A_{i+2}$ must be less than $p/(n+1)$ and the greatest side must be greater than $p/(n+1)$ since the two triangles would not have equal perimeters if that were not the case. This means that $|A_i\widehat{A}_{i+1} - A_{i+2}\widehat{A}_{i+1}| < |A_iA_{i+1} - A_{i+2}A_{i+1}|$ and by theorem 7.3 we thus see that $Area(\triangle A_i\widehat{A}_{i+1}A_{i+2}) > Area(\triangle A_iA_{i+1}A_{i+2})$. Hence the (n+1)-gon in which we have replaced $\triangle A_iA_{i+1}A_{i+2}$ with $\triangle A_i\widehat{A}_{i+1}A_{i+2}$ has a greater area than the (n+1)-gon we started out with. If, when creating $\triangle A_i\widehat{A}_{i+1}A_{i+2}$, we get a concave (n+1)-gon we repeat the same process as delineated above and obtain a convex n-gon of the same perimeter but with greater area than the original (n+1)-gon.

In the former case we now have an (n+1)-gon with (at least) one side of length $p/(n+1)$. Repeating this exact same process we can create a new (n+1)-gon with (at least) two more sides of length $p/(n+1)$. Continuing this process, the area increasing each time, we eventually end up with one of two figures; either an equilateral (n+1)-gon with all sides of length $p/(n+1)$ and the same perimeter but with greater area *or* an n-gon with the same perimeter as the original (n+1)-gon but with greater area.

So to summarize, we have now proved that for any convex (n+1)-gon, *i)* or *ii)* as described above, holds. Using theorem 7.1, we see that if *i)* holds, then the convex regular (n+1)-gon with perimeter p has a greater area than any convex (n+1)-gon with perimeter p. When *ii)* holds we see that if the induction hypothesis holds true, theorem 7.2 also yields the same conclusion. $\square$

8 Dido's Problem

Legend has it that Dido was the founder of Carthage during the first millennium B.C. and that she later became queen of the said city. The legend of Dido has been recited by many writers throughout the ages, but Virgil's version as retold in *The Aeneid* is surely the most well known. Dido, the daughter of the king of Tyre, found herself forced to abandon the city of Tyre together with a steadfast entourage due to her deceitful brother Pygmalion.

In due course they arrived in northern Africa. Here they met up with Hiarbas, a local king. Looking for a piece of land, they reached an agreement with the king, that they could have as much soil as could be enclosed by the hide of an ox. Upon hearing this, Dido proceeded to cut the hide into exceptionally thin stripes. Using the coastline as a boundary, she thereby managed to surround a substantial piece of land with the chopped-up pieces of the ox hide tied into a long string. Here, the city of Carthage was eventually erected [3].

The ingeniousness with which Dido carried out the aforementioned feat is the basis for what is called Dido's problem. Its presentation varies but it is largely the same problem. Firstly, the original version is presented. Thereafter two other versions are shown. The solutions are due to G. Pólya [11].

8.1 Dido's First Problem

Let L be a line segment of infinite length. Of all the figures that can be encompassed by the line segment and a string of length *l'*, what figure encompasses the largest area?

Solution

Let the string form a semicircle, S, together with L. Now reflect the semicircle across L so that a full circle, C, is generated. We let $area(S) = A$ and thus $area(C) = 2A$. We also note that $perimeter(C) = 2l'$.

We will now create a new figure. Let the string form a region, R_1, together with L. This region is allowed to be any shape but a semicircle. We now proceed to reflect R_1 across L, thereby forming a new region, R_2. Now, let $area(R_1) = A'$ and so consequently $area(R_2) = 2A'$. We see that $perimeter(R_2) = 2l'$. See figure 18.

The regions C and R_2 have the same perimeter and therefore it follows from the isoperimetric theorem that $area(C) > area(R_2)$. So $2A > 2A' \Rightarrow A > A' \Rightarrow area(S) > area(R_1)$. We thus see that the figure yielding the greatest area is a semicircle.

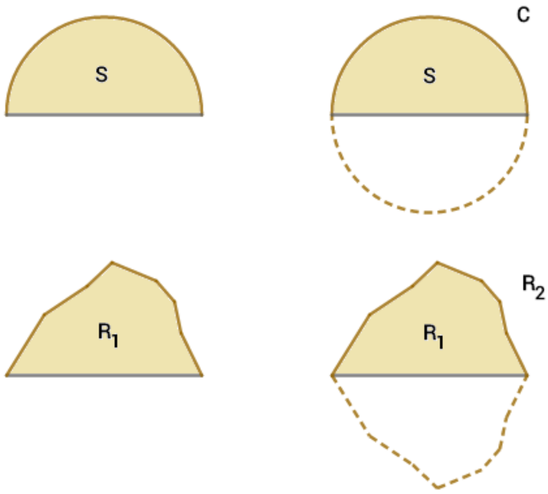


Figure 18

8.2 Dido's Second Problem

Let L be a finite line segment of length *l*. Of all the figures that can be encompassed by the line segment and a string of length *l'*, what figure encompasses the largest area?

Solution

Case 1 ($l' \leq l$): In this case the solution is exactly the same as the one offered above.

Case 2 ($l' > l$): Let the string form a circular arc, that together with L, encloses the region C_1. It thus holds that $perimeter(C_1) = l + l'$. Such an arc will always exist since the shortest distance between two points is a straight line. We now extend C_1 with a circular region C_1', creating a full circle, C_2. This circle is what C_1 would have been if we had continued drawing the perimeter all the way around. We thus see that $C_1 + C_1' = C_2$.

We will now create a new region. Let the string form a region together with L, called R_1, that is not a circular segment. It follows that $perimeter(R_1) = l + l'$. Extend R_1 with the exact same region C_1', creating a new region R_2. See figure 19. So $R_1 + C_1' = R_2$.

The regions C_2 and R_2 have equal perimeters. By the isoperimetric theorem it thus follows that $area(C_2) > area(R_2)$. Hence, we see that $area(C_1) + area(C_1') > area(R_1) + area(C_1') \Rightarrow area(C_1) > area(R_1)$ and accordingly we have proven that a circular segment generates the greatest area.

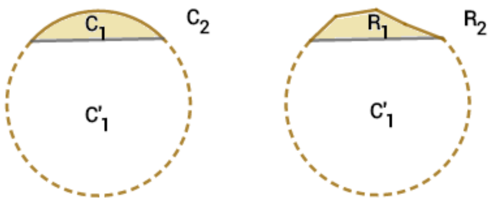


Figure 19

8.3 Dido's Third Problem

Let L be the angle $\theta < \pi$. In other words, L is comprised of two line segments of infinite length adjoined at the vertex, V, and θ is the angle between the two line segments. We have a string of fixed length, *l'*, which will be attached to two points (A and B) on L; both points being placed on separate line segments. The two attachment points are not fixated and one can therefore slide them up and down along their line segment. Of all the figures that can be encompassed by L and a string of length *l'*, what figure encompasses the largest area?

Solution

The solution consists of solving two related problems, *i)* and *ii)*, and then finally using them to arrive at the solution to the original problem.

i) We begin by looking at the case in which A and B are fixated on L. The points are placed on separate line segments. We also let $\theta < \pi$. Recalling case 2 of Dido's second problem, we see that if we imagine a straight line from A to B and if the string with a given length forms a circular arc with endpoints at A and B, the greatest region is cut off from the angle. To see the veracity of this, recollect that the triangle ABV is like C_1' in the previous problem, only that it is a triangle this time. In *(I)* in figure 20 the string forms a circular arc and in *(II)* the string is in the shape of an arbitrary figure that is not a circular arc. In both cases the triangles are identical and thus cut off the same area. The area enclosed by the line segment AB and the string is the greatest when the string forms a circular arc by the solution to Dido's second problem. If we add this area and the area of the triangle we see that in *(I)* the greatest area is cut off in total. It should be noted that at this stage it does not matter whether or not the string ends up crossing one or both of the line segments. If $l' < |AB|$ we cannot form any figure.

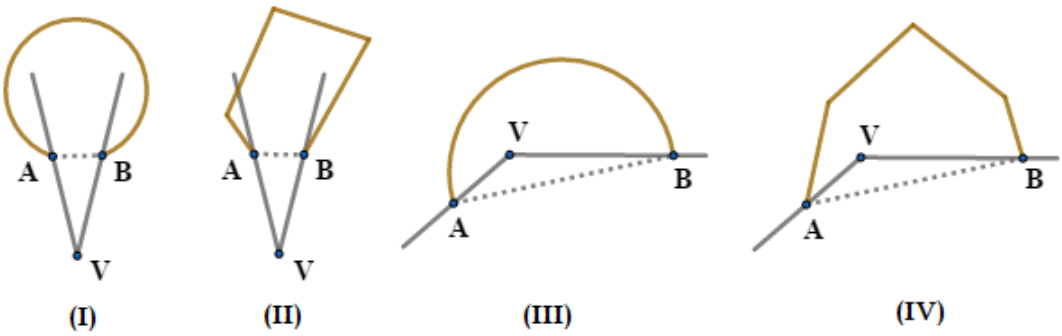


Figure 20

We now let $\theta > \pi$ and yet again fix A and B on separate line segments. A circular arc is formed with endpoints A and B in *(III)*. In *(IV)* an arbitrary shape is formed by the string. By adding the triangle ABV to the area cut off by the string we see through Dido's second problem that forming the string as in *(III)* yields the greatest area. In this case the string might also wind up crossing the line segments and should $l' < |AB|$ we will not be able to create any figure.

ii) We now turn to the situation when *only* A is fixed. Point B can be placed anywhere as long as it is located on the other line segment. The angle should be $< \pi$. Our goal is now, once again, to cut off as great an area as possible from the angle. The length of the string is fixed at l' and its endpoints should be placed at A and B. Seeing VB as a mirror we proceed to reflect VA over to the left of VB. Let the new reflected line segment be called VC.

With the help of the imaginary line AC we can create the triangle ACV.

If we seek to find the greatest area that can be cut off from the angle that is hemmed in by the line segment on which A is placed and its reflected line segment on which C is located, with a string that starts in A and ends in C, this problem is identical to *i)*. See figure 21 which shows two different scenarios, both of which can be solved in the same manner as *i)*. The solution is thus a circular arc which is perpendicular to the line segment BV at the point B. The location of B is chosen so that the string can pass through B on its way to C. Hence, we finally see that the solution to *ii)* is an arc of a circle perpendicular to BV at B. The string might end up crossing one or both of the line segments, which is fine at this stage too.

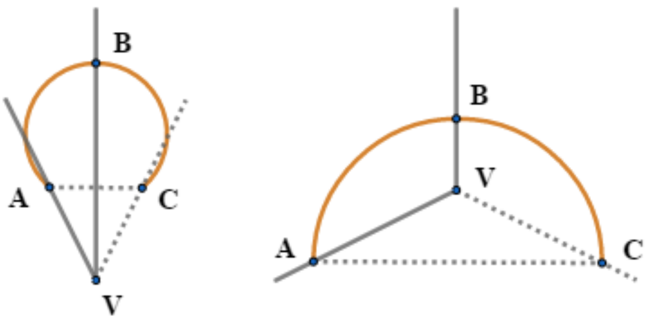


Figure 21

We are now ready to solve the original problem. So, to reiterate, we want to find the greatest area that can be cut off from an angle $\theta < \pi$. The points A and B will be placed on separate line segments and neither point will be fixed and can therefore be placed anywhere on their line segment. The string has the fixed length l'.

We saw in *ii)* that when fixing A, the solution was a circular arc perpendicular to BV at B. Likewise, if we fix B the solution will be a circular arc perpendicular to AV at A by *ii)*. Thus, in order to achieve the greatest area one has to place A and B so that a circular arc can be formed that is perpendicular at both points to their respective line segments, simultaneously. This is the arc of a circle with its center located at V.

9 Figures of Constant Width

The circle can be regarded as the quintessential shape of constant width. As it turns out there are an infinite number of these figures, but the only one that is truly ubiquitous is the circle. While the vast majority of people have heard of the circle, surely not as many would be able to name any other figure of constant width.

9.1 What are Figures of Constant Width?

We must not get ahead of ourselves and so we will first and foremost look at how figures of constant width can be defined. In order to do that we need to look at the concept of a supporting line.

A supporting line is a line segment that touches a curve in any number of points while still retaining the points of the curve either on the line segment itself or on only one side of the line segment. There are exactly two supporting lines running parallel to one another in every direction. To find a pair of supporting lines for a given direction, one can draw two parallel lines that do not touch the figure and then slide them toward one another until they make contact with the figure. The shortest distance between a pair of supporting lines is a straight line between the two that forms a right angle with both lines.

While a tangent line is a concept closely related to a supporting line, they are not interchangeable. For instance, a supporting line touching the vertex of a triangle is not a tangent line at the same vertex.

A figure of constant width is defined as a convex figure that has the same width, w, in all directions; more specifically the shortest distance between its two supporting lines in a given direction is the same, whichever the direction. Figure 22 shows four closed curves with two pairs of supporting lines for each curve; (I) and (II) being figures of constant width and (III) and (IV) not having constant width.

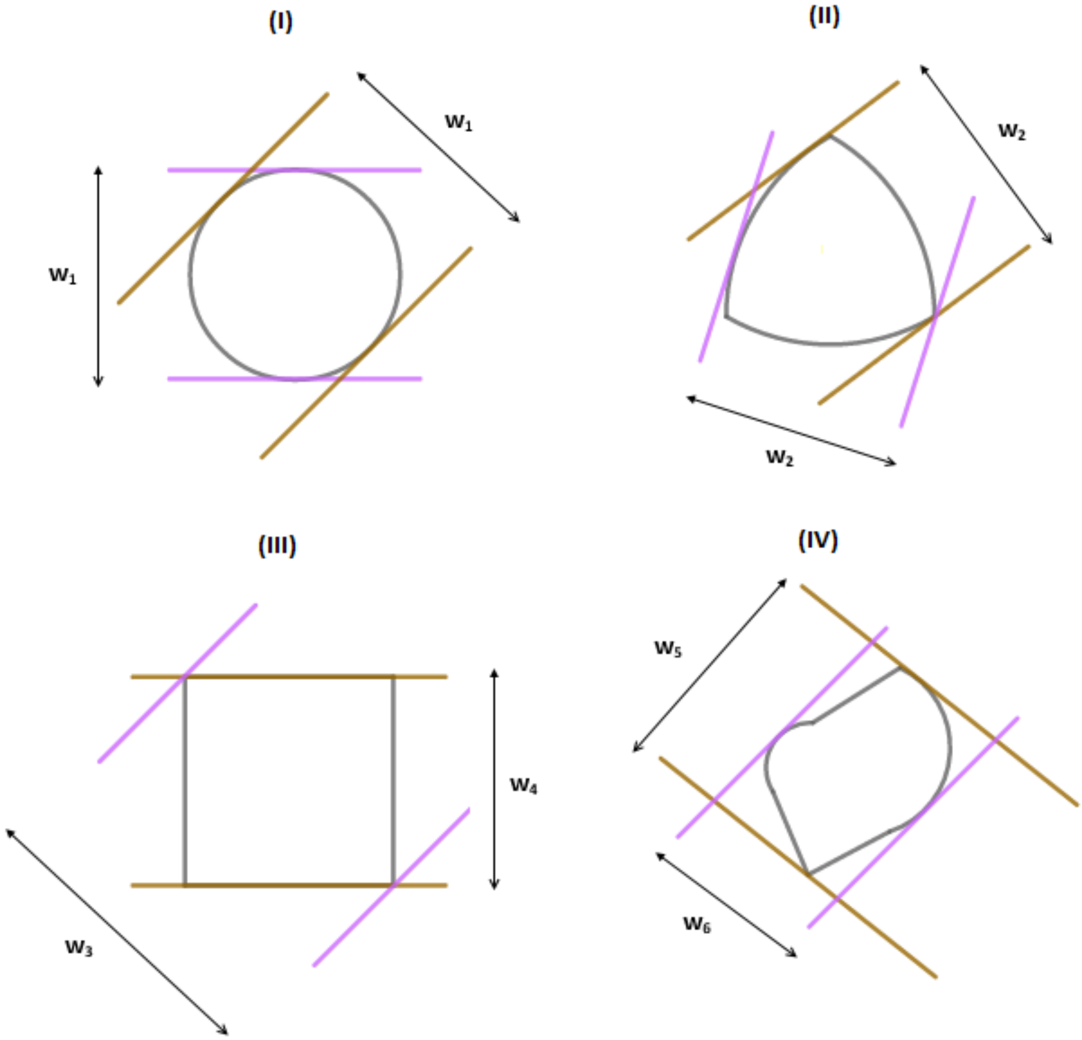


Figure 22

The diameter of a circle is also its width and therefore it truly is a figure of constant width. What other figures of constant width exist? For instance, the Reuleaux polygons all belong to the class of figures of constant width. A Reuleaux polygon can be constructed in the following way: Construct a regular n-gon, where n is an odd number. Next, let one of the vertices be the center of a circle that runs through the vertices that make up the endpoints of the opposite side. We repeat this by creating such a circle for all the vertices of the polygon. The perimeter of the Reuleaux polygon thus consists of n circular arcs. In figure 23 a Reuleaux pentagon is created.

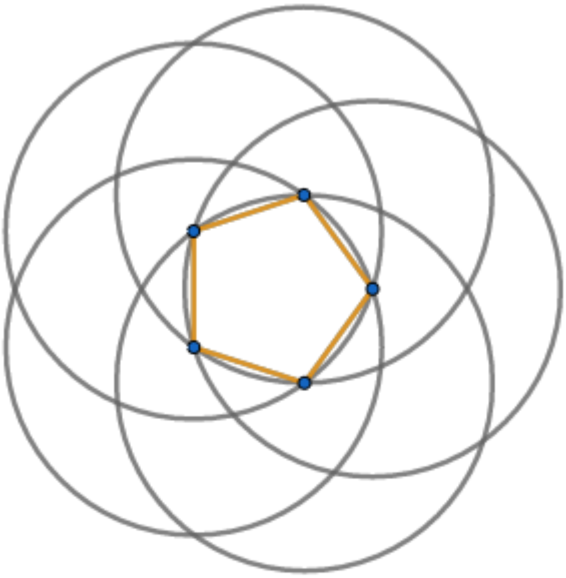


Figure 23

While not sharing the ubiquity of the circle, the Reuleaux polygons can still be found in various places. Examples include church windows in the shape of Reuleaux triangles and coins such as the British 20- and 50-pence coins which are in the form of Reuleaux heptagons. Figure 24 shows some Reuleaux polygons.

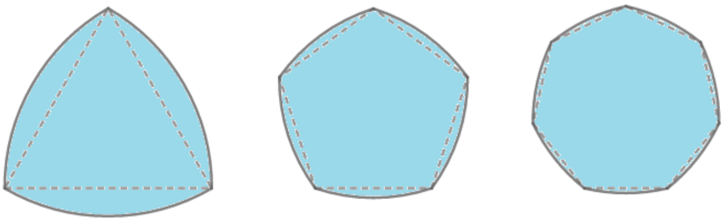


Figure 24

9.2 Barbier's Theorem

An interesting result regarding figures of constant width is Barbier's theorem. Hereunder we will prove its veracity. First however, we will give the solution to the rather famous problem called *Buffon's Needle* (the version with a shorter needle). It was stated by George-Louis Leclerc (Comte de

Buffon) in the 18th century and we will solve it using calculus [1]. Having done that, the proof of Barbier's theorem follows by the use of a little mathematical statistics and the definition of supporting lines.

Theorem 9.1 [Buffon's Needle]
Suppose that a needle of length l is dropped on a piece of paper. Furthermore, we assume that on the piece of paper there are parallel straight horizontal lines whose distance from one another is w and that $l \leq w$. Then the probability of the needle crossing one of the lines is equal to $\frac{2l}{\pi w}$.

Proof. Suppose that a needle has just been dropped on the previously mentioned sheet of paper. Let d be the distance between the lowest point of the needle and the line immediately above it. It thus holds that $0 \leq d \leq w$. We also imagine a vertical line spanning the entire distance d. Now let θ be the angle subtended by the part of the horizontal line between the imaginary line and the needle. Therefore $0 \leq \theta \leq \pi/2$. Once the needle has landed on the paper, the needle will be crossing a line if and only if $d \leq l cos\theta$. See figure 25.

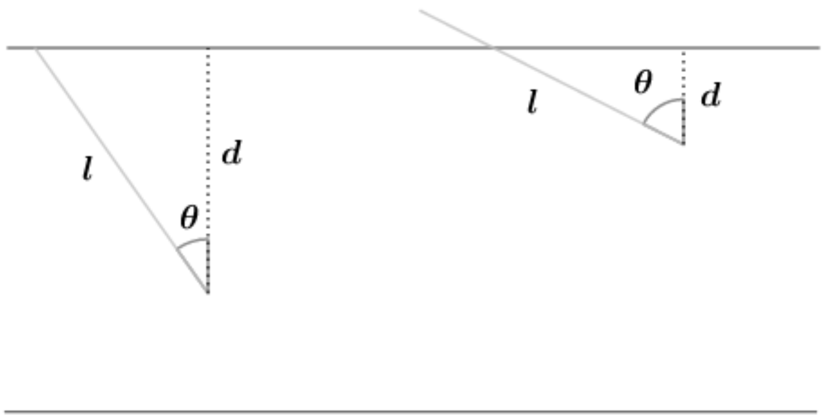


Figure 25

The points between the graph of the function $l\, cos\, \theta$, $0 \leq \theta \leq \pi/2$ and the θ-axis represent all the possible combinations of d and θ for which the needle lands on a line. This area has been marked in blue in figure 26. Similarly, the points lying between the constant function w, $0 \leq \theta \leq \pi/2$ and the θ-axis show all the possible outcomes of d and θ.

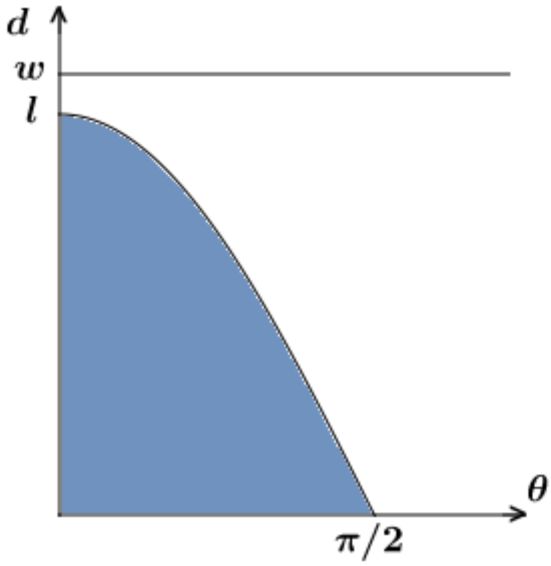


Figure 26

Therefore the probability of the needle landing on a line is

$$P = \frac{\int_0^{\pi/2} l\, cos\,\theta\, d\theta}{\int_0^{\pi/2} w\, d\theta} = \frac{2l}{\pi w}. \tag{9.2.1}$$

□

Thus theoretically, one could approximate π by randomly dropping needles on a piece of paper. To get close to the true value one would have to drop a vast amount of needles. Today however, by simulating on a computer, this is certainly feasible.

Now, in order to prove Barbier's theorem we will need some basic results from mathematical statistics. For a discrete random variable X, the expected value (expectation) is defined as $E[X] = \sum_j jP(X = j)$. Another useful result is that $E[kX] = kE[X]$, $k \in \mathbb{R}$. Also of use is the fact that $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$ for any random variables (even if they are dependent) $X_1, ..., X_n$. We are now ready to prove the theorem of Barbier. The idea comes from [15].

Theorem 9.2 [Barbier's Theorem]
All figures of constant width with the width w, have perimeter πw.

Proof. Let X_1 be a random variable denoting how many lines the needle crosses when dropped on the aforementioned sheet of paper. This stochastic variable will take the value 0 or 1 since $l \leq w$. So X_1 has what is called a Bernoulli distribution. Using the result from (9.2.1) we see that $E[X_1] = 0P(X_1 = 0) + 1P(X_1 = 1) = \frac{2l}{\pi w}$.

Now let us imagine that we have n needles and let X_i, $i = 1, ..., n$ be n random variables that denote how many times needle $1, ..., n$ crosses a line if we were to drop them on the piece of paper. It holds that X_i, $i = 1, ..., n$ are identically distributed. Since $E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i]$ it does not matter whether the n pieces are joined together and then dropped or if they are dropped individually one at a time. Furthermore, due to the identical distribution of the stochastic variables we can write $\sum_{i=1}^{n} X_i = nX$. Here $X = X_i$, $i = 1, ..., n$.

A circle consists of an immense amount of infinitesimally small line segments of the same length joined together at equal angles to one another. In other words, a regular n-gon turns into a circle as $n \to \infty$. A circle of diameter (width) w has a perimeter equal to πw. If one were to drop a "needle" in the shape of the aforesaid circle on the piece of paper, the circle would clearly always end up intersecting two lines. We let $Y = nX$ be the random variable denoting the number of times n line segments intersect one of the lines. If these line segments are linked together in the shape of a regular n-gon and n is a very large number we are essentially dealing with a circle. Thus $E[Y]$ is the expected value of the number of times the circle will intersect a line. We see that

$$E[Y] = E[nX] = nE[X] = n\frac{2\frac{\pi w}{n}}{\pi w} = 2. \tag{9.2.2}$$

So when the length of the needle is πw, the expected value of the number of intersections is exactly equal to two. From the definition of figures of constant width and supporting lines, we see that if we were to drop any figure of constant width (its width equal to w) on the sheet of paper, it would always generate *two* intersections. Particularly, the expected value of the number of intersections has to be equal to two. In order for that to hold, the "length of the needle" shaped like a figure of constant width has to be πw. Hence the width of every figure of constant width is equal to πw. $\square$

Some vending machines and other machinery that handle coins require them to be more or less of constant width. From the isoperimetric theorem (theorem 1.1) it follows that out of all shapes of constant width, the circle calls for the use of the most amount of metal. In order to reduce the amount of metal used, any other shape of constant width would actually fare better.

10 The Isoperimetric Theorem in Higher Dimensions

So far this thesis has only dealt with the isoperimetric theorem in two dimensions. Naturally several questions arise. Is there a corresponding theorem for $\mathbb{R}^3$? If so, how can it be proven? Moreover, would it be possible or even meaningful to venture even further to the dizzying heights of $\mathbb{R}^n, n \geq 4$? To the joy of mathematicians everywhere, the answer to all of these questions is yes.

To delve deeply into this is unfortunately outside the scope of this thesis. We can, however, make a brief foray into the broad topic of the isoperimetric theorem for higher dimensions. For three dimensions, the isoperimetric inequality can be stated as follows:

Theorem 10.1 [The Isoperimetric Inequality in $\mathbb{R}^3$]
For any three-dimensional region with surface area S and volume V, the following inequality holds:

$$36\pi V^2 \leq S^3.$$

Equality holds for the sphere only.

To actually prove that this is the case is far from simple and significantly harder than the planar case. Just like in $\mathbb{R}^2$, one also has to prove the existence of such a figure. The first proof of the isoperimetric inequality in $\mathbb{R}^3$

is due to H.A. Schwarz [12]. The paper is in German and according to [5], this proof is given "in a rather difficult paper."

We will, however, prove equality. For a sphere in $\mathbb{R}^3$, the volume can be expressed as $V = \frac{4\pi r^3}{3}$. Also, its surface area can be written as $S = 4\pi r^2$. Using the latter formula we see that $r = \left(\frac{S}{4\pi}\right)^{1/2}$. Hence the volume can be expressed in terms of the surface area: $V = \frac{4\pi}{3}\left(\frac{S}{4\pi}\right)^{3/2} = \frac{S^{3/2}}{6\sqrt{\pi}}$. Squaring both sides and rearranging we acquire $36\pi V^2 = S^3$.

One can formulate the isoperimetric inequality in $\mathbb{R}^n, n \geq 2$ in the following way [8]. It should be noted that a formal definition of the theorem requires us to define what type of domain we are looking at since not all domains have a volume in n dimensions that is well-defined. Moreover, the boundary also needs to have a well-defined "area" in dimension $n-1$.

Theorem 10.2 [The Isoperimetric Inequality in $\mathbb{R}^n, n \geq 2$]
Let $\Omega \subset \mathbb{R}^n$ be a bounded domain and let $\partial\Omega$ be its boundary. Furthermore, let V_n be the volume of the unit sphere in $\mathbb{R}^n$. Then the following inequality holds:

$$nV_n^{1/n}|\Omega|^{1-1/n} \leq |\partial\Omega|.$$

Here $|E|$ stands for the Lebesgue volume measure in n dimensions or the surface measure in dimension $n-1$ of $E \subset \mathbb{R}^n$. Equality holds for the sphere only.

Kesavan [8], states that for $n \geq 3$ "even the notion of 'surface measure' of the boundary is not obvious." Furthermore, he says that "when $N = 2$, we clearly understand the notion of length of a rectifiable curve (a curve whose length we can define). In higher dimensions, $\partial\Omega$ will be a (N-1)-dimensional manifold and there are several ways to define $|\partial\Omega|$. (...) In general, the proof uses difficult notions from geometric measure theory. Recently, Cabre (...) has observed that it is possible to use an idea similar to that by Alexandrov in proving certain estimates for solutions of elliptic partial differential equations to prove the classical isoperimetric theorem." The latter approach is then used by Kesavan to prove the isoperimetric theorem in $\mathbb{R}^n, n \geq 2$.

We will now prove that theorem 10.2 holds for $n = 2$ and $n = 3$. In $\mathbb{R}^2$, the surface area is equivalent to the perimeter and so $|\partial\Omega| = P$. The area of the unit circle is $V_2 = \pi$. The inequality thus becomes $2\pi^{1/2}A^{1-1/2} \leq P$ which can be written as $4\pi A \leq P^2$.

In $\mathbb{R}^3$ the volume of the unit sphere is $V_3 = \frac{4\pi}{3}$. Just like before, we denote the surface area with S. So by theorem 10.2 we obtain $3\left(\frac{4\pi}{3}\right)^{1/3}V^{1-1/3} \leq S$. This can be rewritten as $36\pi V^2 \leq S^3$.

In conclusion, there does not seem to be a way of proving the isoperimetric theorem for $\mathbb{R}^n, n \geq 3$ by only using elementary geometry. One has to make use of advanced mathematics that lie beyond the scope of this thesis. Even so, only just knowing that it is possible to prove for n dimensions is interesting indeed!

11 Summary

The isoperimetric problem is truly an ancient problem. As we have previously seen, Zenodorus allegedly proved that the circle was the solution in the 2nd century B.C. Nevertheless, by modern standards the proof would not suffice. The search for a proof that would be accepted in modern times was commenced in the 19th century by Jakob Steiner. He constructed several proofs, the first of which was published in 1841. For all their beauty, they were unfortunately not complete. He had failed to address whether or not a figure of maximum area actually exists.

In 1879, Karl Weierstrass constructed the first complete proof of the isoperimetric theorem. Utilizing calculus of variations, he had managed to produce a proof that left no doubt as to the existence of a figure of maximum area. Since then several more proofs of the isoperimetric theorem have surfaced. The methods used can vary greatly as we have seen in section 6. Not only is this fascinating in and of itself, but it also truly displays the beauty and versatility of mathematics!

We have also seen that the isoperimetric problem can be defined for polygons as well. The proof for n-gons makes it clear that the regular polygon solves the isoperimetric problem for all $n \geq 3$. Still, curiously enough, no proofs can be found that do not employ the original isoperimetric theorem.

The isoperimetric theorem also lends itself to several intriguing problems and related topics. There are, for instance, various renditions of Dido's problem. Figures of constant width is also a subject that is isoperimetric in nature.

We had a brief look at the isoperimetric problem in higher dimensions. As we venture beyond the planar realm the complexity increases. Nevertheless, it has been proven that the sphere solves the isoperimetric theorem in $\mathbb{R}^3$. Not only that but it has also been shown that the n-dimensional sphere is the solution to the isoperimetric problem in $\mathbb{R}^n$. When making a formal definition of the isoperimetric theorem in $\mathbb{R}^n$ we need to specify what domain we are looking at. The reason for this is that the volume in dimension n has to be well-defined. In addition, the boundary should have an "area" in dimension $n-1$ that is just as well-defined.

Surely, problems of an isoperimetric nature appear to offer an endless source of possibilities and in that regard we seem to have just barely scratched the surface. In other words more discoveries remain to be made and that is undoubtedly a satisfying thing to know.

12 References

[1] Aigner, M. and Ziegler, G.M. (2010). *Proofs from THE BOOK.* 4th ed. Berlin: Springer-Verlag, pp. 155-157.
[2] Blåsjö, V. (2005). THE EVOLUTION OF...The Isoperimetric Problem. *American Mathematical Monthly June-July 2005*, **112**, pp. 533-534.
[3] Cartwright, M. *Dido.* (Retrieved 2017-09-05 from *https://www.ancient.eu/Dido/*)
[4] Claydon, D. *The AM-GM Inequality.* (Retrieved 2018-04-24 from *https://dansmaths.quora.com/The-AM-GM-Inequality*)
[5] Courant, R. and Robbins, H. (1996). *What is Mathematics?* 2nd ed. New York: Oxford University Press, p. 376.
[6] Jaglom, I.M. and Boltyanskii, V.G. (1961). *Convex Figures.* New York: Holt, Rinehart and Winston, pp. 207-211.
[7] Kazarinoff, N.D. (1961). *Geometric Inequalities.* Washington, D.C.: The Mathematical Association of America, pp. 35-36, 38, 48-51, 53.
[8] Kesavan, S. (2002). The Isoperimetric Inequality. *Resonance, Volume 7, Issue 9, September 2002*, pp. 2-3.

[9] Lawlor, G. (1998). A New Area-Maximization Proof for the Circle. *The Mathematical Intelligencer, Number 1 1998*, **20**, pp. 29-31.

[10] Lax, P.D. (1995). A Short Path to the Shortest Path. *The American Mathematical Monthly, Number 2 Feb. 1995*, **102**, pp. 158-159.

[11] Pólya, G. (1954). *Induction and Analogy in Mathematics.* New Jersey: Princeton University Press, pp. 183-185, 269.

[12] Schwarz, H.A. (1890). *Gesammelte Mathematische Abhandlungen, Zweiter Band.* Berlin: Verlag von Julius Springer.

[13] Siegel, A. *A Historical Review of the Isoperimetric Theorem in 2-D, and its place in Elementary Plane Geometry.* (Retrieved 2017-09-11 from *https://cs.nyu.edu/faculty/siegel/SCIAM.pdf*)

[14] Weierstrass, K. (1927). *Mathematische Werke. Bd 7, Vorlesungen über Variationsrechnung.* Berlin: Mayer & Müller.

[15] *Buffon's noodle.* (Retrieved 2018-03-10 from *https://en.wikipedia.org/wiki/Buffon's_noodle*)