

From the Triangle Inequality to the Isoperimetric Inequality

S. Kesavan

Institute of Mathematical Sciences,
CIT Campus, Taramani, Chennai - 600113.

kesh@imsc.res.in

IMSc, Chennai
September 4, 2015

Triangle Inequality

One of the important results we learn in plane geometry at high school is that in any triangle, the sum of the lengths of any two sides is strictly greater than the length of the third side.

Triangle Inequality

One of the important results we learn in plane geometry at high school is that in any triangle, the sum of the lengths of any two sides is strictly greater than the length of the third side.

This has been generalized as the *triangle inequality* when defining a metric (which generalizes the notion of distance) in topology and plays a very key role in the study of metric spaces. A particular case of this is the inequality bearing the same name when defining a *norm* on vector spaces.

Polygonal Paths

Definition

A **polygonal path** joining two points in the plane is a path which is made up of line segments.

Polygonal Paths

Definition

A **polygonal path** joining two points in the plane is a path which is made up of line segments.

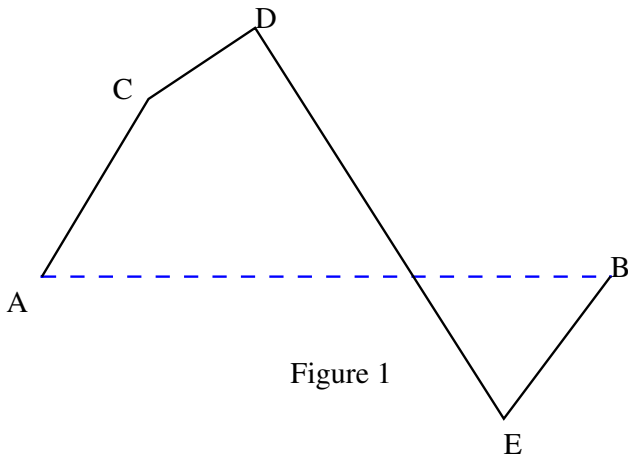


Figure 1

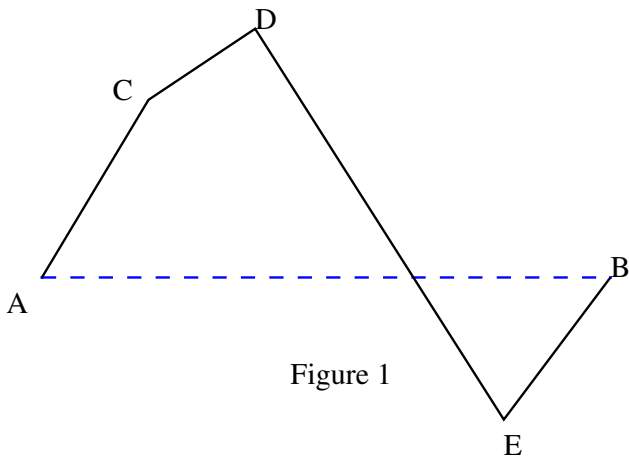


Figure 1

In this figure, we see that $AB < AC + CB$, $CB < CD + DB$ and $DB < DE + EB$ and combining these, we see that

$$AB < AC + CD + DE + EB.$$

We can generalize this, using mathematical induction, to any number of points and we deduce the following result.

Theorem

Of all polygonal paths joining two points in the plane, the straight line joining the points has the shortest length.

A curve in the plane can be considered as a continuous map $\gamma : [0, 1] \rightarrow \mathbb{R}^2$. The end points of the curve are $\gamma(0)$ and $\gamma(1)$. The curve is said to be a simple curve if γ is injective on $(0, 1)$ and it is a closed curve if $\gamma(0) = \gamma(1)$.

Rectifiable Arcs

A curve in the plane can be considered as a continuous map $\gamma : [0, 1] \rightarrow \mathbb{R}^2$. The end points of the curve are $\gamma(0)$ and $\gamma(1)$. The curve is said to be a simple curve if γ is injective on $(0, 1)$ and it is a closed curve if $\gamma(0) = \gamma(1)$.

Consider a partition \mathcal{P} of the interval $[0, 1]$:

$$\mathcal{P} : 0 = t_0 < t_1 < t_2 < \cdots < t_n = 1.$$

The points $\{\gamma(t_i)\}_{i=0}^n$ lie on the curve and if we connect pairs of consecutive points $\gamma(t_i)$ and $\gamma(t_{i+1})$, for $0 \leq i \leq n-1$, by line segments, we get a polygonal path from $\gamma(0)$ to $\gamma(1)$. Let $\ell(\mathcal{P})$ denote the length of this polygonal path.

Definition

The curve γ is said to be **rectifiable** if

$$\sup_{\mathcal{P}} \ell(\mathcal{P}) < +\infty$$

where the supremum is taken over all possible partitions of the interval $[0, 1]$. The finite supremum thus obtained is called the length of the curve.

As a consequence of the above definition and Theorem 1, we deduce the following result.

Corollary

Of all paths connecting two points in the plane, the straight line joining them has the shortest length.

Heron's Theorem

Consider two points A and B in the plane and a line ℓ lying below them. Consider all possible polygonal paths from A to B consisting of two line segments AP and PB , where P lies on the line ℓ . What is the shortest possible such path?

Heron's Theorem

Consider two points A and B in the plane and a line ℓ lying below them. Consider all possible polygonal paths from A to B consisting of two line segments AP and PB , where P lies on the line ℓ . What is the shortest possible such path?

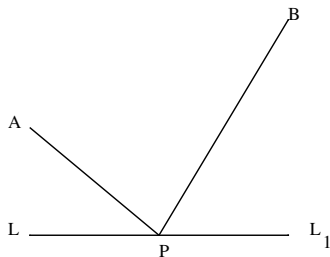


Figure 2

Proof of Heron's Theorem

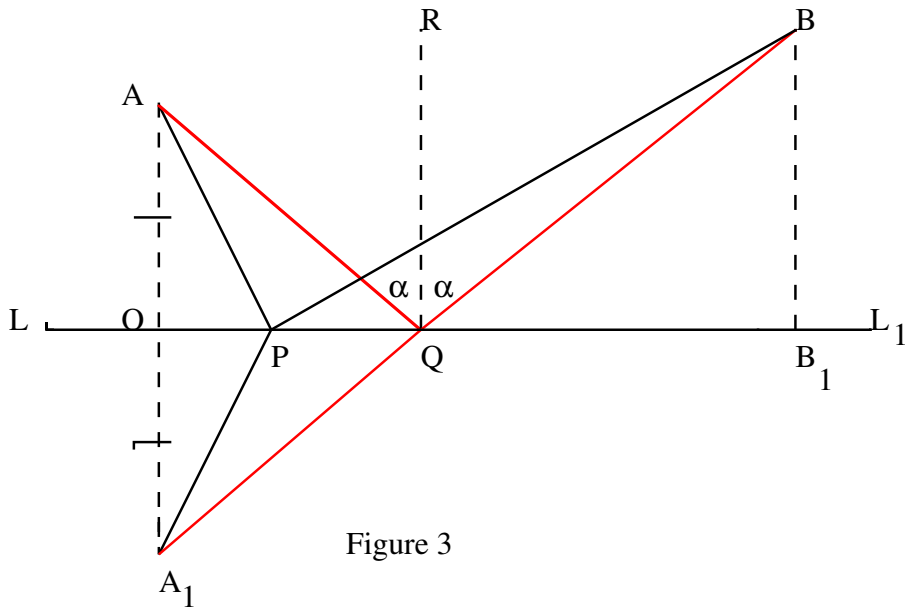


Figure 3

Thus the optimal point Q is such that the 'angle of incidence' is equal to the 'angle of reflection'. This result is called *Heron's theorem*.

Thus the optimal point Q is such that the 'angle of incidence' is equal to the 'angle of reflection'. This result is called *Heron's theorem*.

This is also the law governing the reflection of light on a plane mirror. It follows from Fermat's principle that light always follows the shortest possible path. This principle can also be used to derive the laws of refraction of light passing through different media.

Exercise:

Given two arbitrary lines ℓ and m in the plane and two points A and B between them, find the shortest polygonal path $APQB$ where P lies on ℓ and Q lies on m (see Figure 4). ■

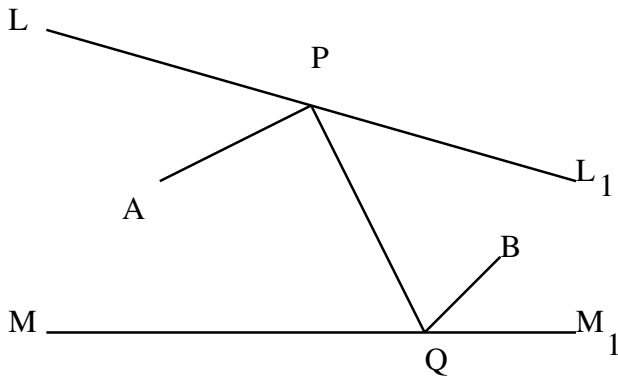


Figure 4

Optimal Triangles

Let $a, b \in \mathbb{R}$ be fixed positive constants. Amongst all triangles ΔABC with base length $BC = b$ and area equal to a , we look for the triangle such that $AB + AC$ is the least possible.

Optimal Triangles

Let $a, b \in \mathbb{R}$ be fixed positive constants. Amongst all triangles ΔABC with base length $BC = b$ and area equal to a , we look for the triangle such that $AB + AC$ is the least possible.

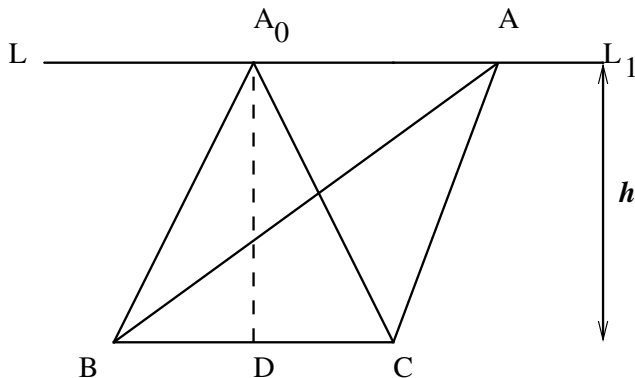


Figure 5

Theorem

Of all triangles with fixed base and fixed area, the isosceles triangle minimizes the sum of the lengths of the other two sides.

A Dual Problem

Of all triangles with fixed base length and fixed sum of the lengths of the other two sides, find the triangle with maximum area.

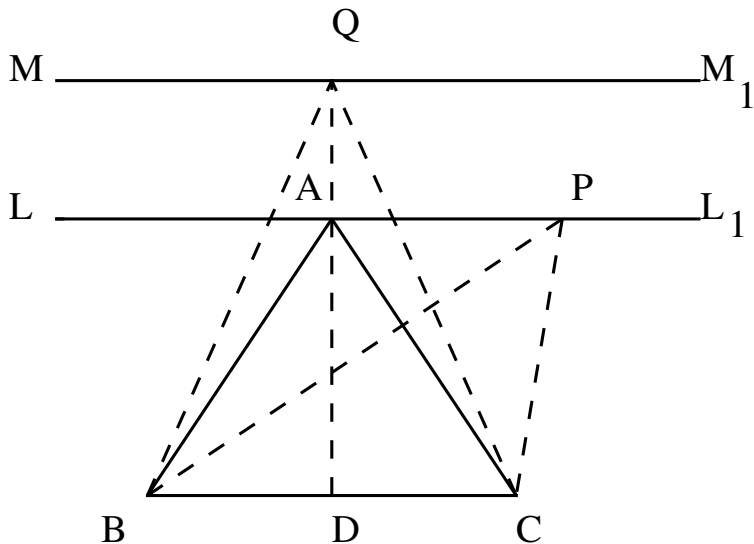


Figure 6

Theorem

Of all triangles with fixed base length and such that the sum of the lengths of the other two sides is a fixed constant, the isosceles triangle has the maximum area.

For those who are familiar with conic sections in coordinate geometry, we can give another proof of Theorem 3. If B and C are fixed, then the locus of A which moves such that $AB + AC$ is a constant, L , is an ellipse. Then, the semi-major axis a is given by $2a = L$ and the eccentricity e of the ellipse is given by $BC = 2ae$. The semi-minor axis b is then defined by $b^2 = a^2(1 - e^2)$.

If the origin O is at the midpoint of BC and if θ is the angle the ray OA makes with the major axis (which is along BC), then the coordinates of A are given by $(a \cos \theta, b \sin \theta)$. The height of the triangle $\triangle ABC$ is therefore $b \sin \theta$ and its area is

$$\frac{1}{2} \cdot BC \cdot b \sin \theta$$

which is maximal when $\theta = \frac{\pi}{2}$, *i.e.* A lies on the minor axis, which implies that the triangle $\triangle ABC$ is isosceles.

Isoperimetric Problem for Polygons

Let $N \geq 3$ be a fixed positive integer. Consider the following problem:

Problem

Of all N -sided polygons with the same perimeter L , find that which encloses the maximum area.

Triangles

If $N = 3$, we are dealing with triangles. If L is the perimeter of a triangle of sides a, b and c , then $L = 2s = a + b + c$, where s is the semi-perimeter. Then by Hero's formula, we know that the area is given by

$$A = \sqrt{s(s-a)(s-b)(s-c)}.$$

Thus, to maximize the area, we need to maximize the product of three positive numbers

$$(s-a)(s-b)(s-c)$$

whose sum equals $3s - (a + b + c) = s$ which is a constant ($= L/2$). From the classical AM-GM inequality, we know that this is possible only when the three numbers are equal. Thus it follows that $a = b = c$, *i.e.* the triangle is *equilateral*.

Quadrilaterals

Let us now consider the case $N = 4$, *i.e.* the case of quadrilaterals. First of all, it is enough to consider convex quadrilaterals.

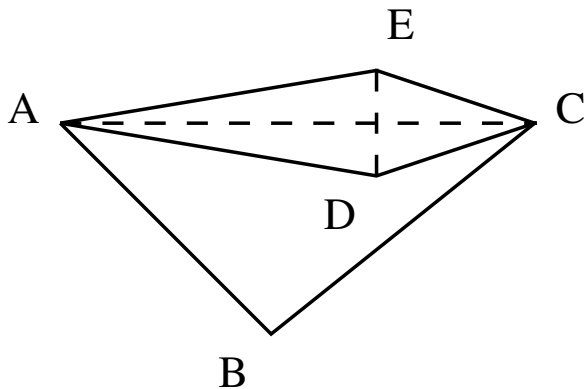


Figure 7

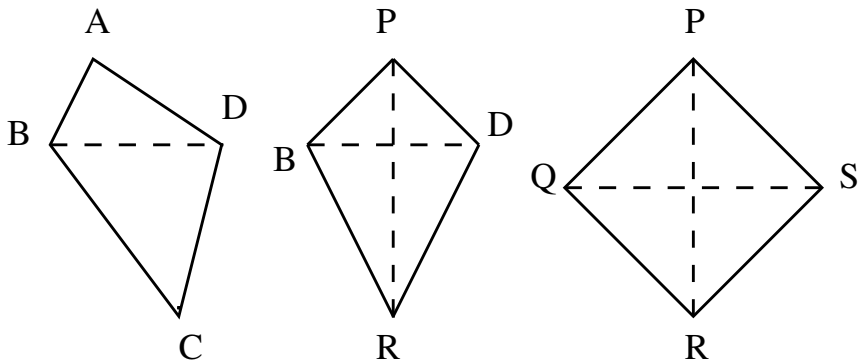


Figure 8

Thus, given any convex quadrilateral, we can construct a rhombus of equal perimeter but of larger area. If $\theta \leq \frac{\pi}{2}$ is an internal angle of the rhombus (whose side is $L/4$, where L is the perimeter), its area is

$$\left(\frac{L}{4}\right)^2 \sin \theta$$

and this is maximal when $\theta = \frac{\pi}{2}$.

Of all quadrilaterals of given perimeter L , the square has the maximum area.

Thus, given any convex quadrilateral, we can construct a rhombus of equal perimeter but of larger area. If $\theta \leq \frac{\pi}{2}$ is an internal angle of the rhombus (whose side is $L/4$, where L is the perimeter), its area is

$$\left(\frac{L}{4}\right)^2 \sin \theta$$

and this is maximal when $\theta = \frac{\pi}{2}$.

Of all quadrilaterals of given perimeter L , the square has the maximum area.

Theorem

Of all N -sided ($N \geq 3$) polygons of fixed perimeter, the regular polygon encloses the maximum area.

By a regular polygon, we mean one which has all sides of equal length and all of whose vertices all lie on a circle.

Unlike the cases of triangles and quadrilaterals, the proof in the general case is more involved. We will present here an ingenious argument due to Steiner for polygons with an *even* number of sides.

Up to now ($N = 3, 4$), we have actually verified that the regular polygon indeed maximizes the area. We will now change our mode of proof. We will first prove that there exists an optimal polygon and based on this assumption of existence, deduce that it must be the regular polygon.

For the existence of an optimal polygon, we argue as follows. The N vertices of an N -sided polygon are fixed by $2N$ coordinates in the plane. Since the perimeter is fixed, say, L , the diameter of the polygon cannot exceed L . Hence all such polygons can be considered to lie inside a sufficiently large box. In other words, these $2N$ coordinates vary in some fixed bounded interval. The area and perimeter are continuous functions of the coordinates.

Thus the set of all N -sided polygons with perimeter L is represented by a closed and bounded set in $2N$ -dimensional space and such a set is compact. Since the area is a continuous function it attains its maximum at some point in the compact set which corresponds to the optimal polygon. This proves the existence of the optimal polygon.

Henceforth we will assume the existence of an optimal polygon of $2N$ sides, $N \geq 2$, and deduce its properties. As in the case of quadrilaterals, it is clear that such a polygon has to be convex.

Now consider an arbitrary pair of adjacent sides of this polygon. Let us name the corresponding vertices A , B and C . Let us freeze all the vertices except B .

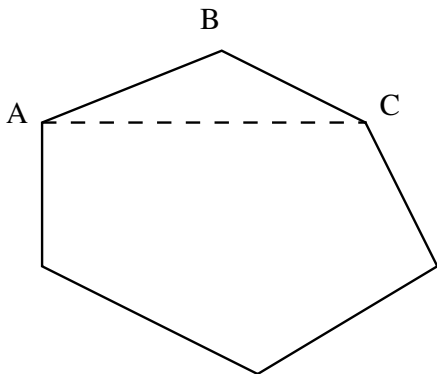


Figure 9

We vary the polygon by just moving the vertex B in a manner that $BA + BC$ is fixed so that the perimeter is not altered. The area of the polygon can be altered only by altering the area of the triangle $\triangle BAC$. Since AC is also frozen and since $BA + BC$ is constant, it follows from Theorem 3 that for the optimal polygon, we must have $BA = BC$. Thus any pair of adjacent sides are equal and so it must be an equilateral polygon (*i.e.* all its sides are equal in length).

Now let us draw a diagonal of this polygon with N sides on either side of it, *i.e.* it bisects the perimeter. We claim that it simultaneously bisects the area as well. If not, if one side had larger area than the other, we can reflect this side with respect to the diagonal to produce a polygon of the same perimeter but of larger area, contradicting the optimality of the polygon.

From now on, we will work with half the polygon defined by this diagonal (this is the reason why the argument works only for an even number of sides).

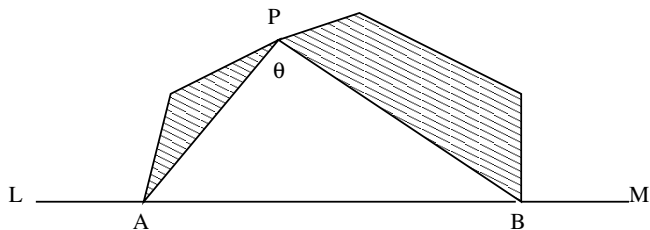


Figure 10

Let AB be the diagonal and consider the upper half of the polygon. Let P be a fixed vertex (other than A and B). The points A and B will be allowed to slide on the line ℓ defining the diagonal. The vertices will move so that the distances PA and PB are fixed and the areas shaded in the figure are fixed.

The new polygon got by reflecting this figure across the line ℓ will still be a regular polygon with the same perimeter. The change in the area is got by changing the area of the triangle ΔPAB which is given by

$$\frac{1}{2}PA.PB.\sin\theta$$

where θ is the angle subtended at P by AB , and this area is maximal when θ is a right angle. Thus, in the optimal polygon, the diagonal bisecting the perimeter and the area subtends a right angle at all the other vertices. This shows that the optimal polygon is regular with the vertices lying on the circle with the bisecting diagonal as diameter.

Now, let L be the perimeter and A the area of a $2N$ -sided polygon. Consider the corresponding regular cyclic polygon with the same perimeter. Then each side has length $L/2N$. The angle subtended by each side at the centre O of the circumscribing circle is $2\pi/2N = \pi/N$.

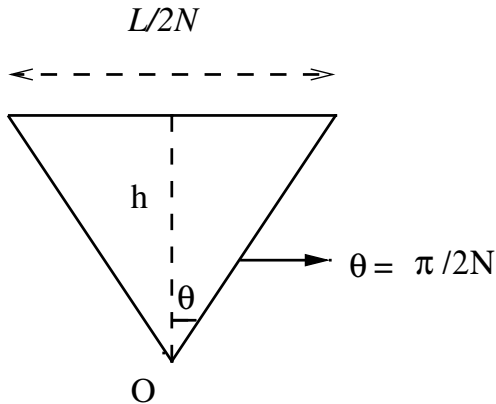


Figure 11

The height of the triangle with one side as base and the centre of the circumscribing circle as the vertex is then given by

$$h = \frac{L}{4N} \frac{1}{\tan(\pi/2N)}$$

and so its area is then seen to be

$$\frac{L^2}{16N^2} \frac{1}{\tan(\pi/2N)}.$$

The height of the triangle with one side as base and the centre of the circumscribing circle as the vertex is then given by

$$h = \frac{L}{4N} \frac{1}{\tan(\pi/2N)}$$

and so its area is then seen to be

$$\frac{L^2}{16N^2} \frac{1}{\tan(\pi/2N)}.$$

The area of the regular cyclic polygon is then $2N$ times this quantity, since there are in all $2N$ identical triangles, which yields

$$\frac{L^2}{8N} \frac{1}{\tan(\pi/2N)}.$$

Since this is the optimal area, we deduce the following result.

Theorem

If L is the perimeter and A is the area of any $2N$ -sided polygon in the plane, then

$$L^2 \geq 8NA \tan\left(\frac{\pi}{2N}\right).$$

The Isoperimetric Inequality

Consider a simple closed curve of length L enclosing a region Ω in the plane, whose area is A . Pick $2N$ points on the curve which will form a polygon of $2N$ sides. Let L_N be the perimeter of this polygon and let A_N be its area. We can choose these points in such a way that, as $N \rightarrow \infty$, we have $L_N \rightarrow L$ and $A_N \rightarrow A$. Now, we have seen that

$$L_N^2 \geq 8NA_N \tan\left(\frac{\pi}{2N}\right) = 4\pi A_N \frac{\tan\left(\frac{\pi}{2N}\right)}{\frac{\pi}{2N}}$$

which yields, as $N \rightarrow \infty$,

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

This is called the *classical isoperimetric inequality* in the plane. It can be shown that equality occurs if, and only if, the curve is a circle. Indeed, for a circle of radius r , we have $L = 2\pi r$ and $A = \pi r^2$, which shows that $L^2 = 4\pi A$. The converse is also true.

This is called the *classical isoperimetric inequality* in the plane. It can be shown that equality occurs if, and only if, the curve is a circle. Indeed, for a circle of radius r , we have $L = 2\pi r$ and $A = \pi r^2$, which shows that $L^2 = 4\pi A$. The converse is also true.

Rephrasing this, if L is the perimeter of a simple closed curve, the maximum area it can enclose is $L^2/4\pi$ and, if A is the area enclosed by a simple closed curve, the perimeter should at least be $\sqrt{4\pi A}$. The circle alone achieves these optimal values.

In other words, of all simple closed curves with fixed perimeter, the circle alone encloses the maximal area, and of all simple closed curves enclosing a fixed area, the circle alone has the least perimeter.

In other words, *of all simple closed curves with fixed perimeter, the circle alone encloses the maximal area, and of all simple closed curves enclosing a fixed area, the circle alone has the least perimeter.*

In ancient literature, Virgil's *Aeneid* mentions *Dido's problem*. Dido was a queen who founded Carthage (modern Tunisia) and she was told that she could have as much land as she could enclose with a piece of oxhide.

Interpreting the word 'enclose' broadly, she cut the oxhide into very thin strips which she then knotted together to form a closed rope. Thus she had a rope of fixed length and she had to place it on the earth so as to enclose the maximum possible area.

In three dimensions, the isoperimetric inequality reads as follows:

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

Here S denotes the surface area of a bounded domain in \mathbb{R}^3 and V is its volume. Again equality holds only for the ball ($S = 4\pi r^2$, $V = \frac{4}{3}\pi r^3$). Thus *of all possible surfaces of fixed surface area, the sphere alone encloses the maximum volume and of all domains of fixed volume, the ball has the least surface area.*

This has a nice application in case of soap bubbles. A soap bubble involves an interface of a liquid and air. The bubble will be stable only if the potential energy due to the surface tension is minimal. This quantity is proportional to the surface area of the liquid-air interface. Thus, when we blow a bubble enclosing a fixed volume of air, nature adjusts the shape of the bubble so that the surface area is minimal and this occurs for the spherical shape and so soap bubbles are round in shape.

The isoperimetric inequality is the starting point of the subject of shape optimization problems, where we look for shapes which optimize some functional subject to some geometric constraints. This is a very active area of research today and lies in the confluence of several areas of mathematics like geometry, partial differential equations, functional analysis and so on.

The isoperimetric inequality is the starting point of the subject of shape optimization problems, where we look for shapes which optimize some functional subject to some geometric constraints. This is a very active area of research today and lies in the confluence of several areas of mathematics like geometry, partial differential equations, functional analysis and so on.

Suggested Reading

1. R. Courant and H. Robbins, **What is Mathematics?**, Second Edition (revised by Ian Stewart), Oxford University Press, 1996.
2. S. Hildebrandt and A. Tromba, **The Parsimonious Universe: Shape and Form in the Natural World**, Copernicus Series, Springer, 1996.

Thank You!