From the Triangle Inequality to the Isoperimetric Inequality

S Kesavan

Starting from the triangle inequality, we will discuss a series of shape optimization problems using elementary geometry and ultimately derive the classical isoperimetric inequality in the plane.

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.

We will now look at some simple consequences of this result in plane geometry.

DEFINITION 1.

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

For example, *Figure* 1 shows a polygonal path made up

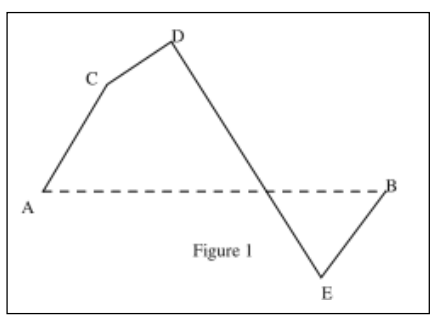


Figure 1

S Kesavan works at the Institute for Mathematical Sciences, Chennai. His area of interest is partial differential equations with specialization in elliptic problems connected to homogenization, control theory and isoperimetric inequalities. He has authored four books covering topics in functional analysis and its applications to partial differential equations.

Keywords
Calculus of variations, shape optimization, isoperimetric problems.

Figure 1.

of four line segments connecting the points A and B. 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 1. *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 $\gamma : [0,1] \to \mathbb{R}^2$. 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 2.

The curve γ is said to be rectifiable if

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

A continuous function which generates a rectifiable curve is known as a function of bounded variation.

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. □

A continuous function which generates a rectifiable curve is known in the literature as a *function of bounded variation* and the supremum obtained above is called the *total variation* of the function.

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

COROLLARY 1.

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

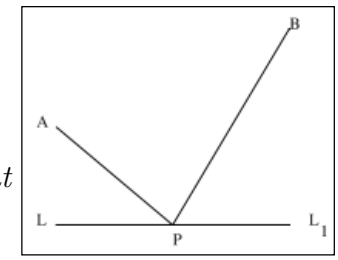


Figure 2.

The above result is one of the first obtained, using differential equations, when studying the *calculus of variations*, which deals with optimization problems in function spaces.

Heron's Theorem

Consider two points A and B in the plane and a line LL_1 lying below them (see *Figure* 2). Consider all possible polygonal paths from A to B consisting of two line segments AP and PB, where P lies on the line LL_1. What is the shortest possible such path?

To find the optimal path we proceed as follows. Let A_1 be the reflection of the point A with respect to the line LL_1 (see *Figure* 3). Since the triangles ΔAOP and Δ A_1OP are congruent, we have $AP = A_1P$. Thus, the length of the polygonal path APB is the same as that of A_1PB. But the shortest path from A_1 to B is the straight line A_1B. If it intersects the line LL_1 at Q, then the optimal path that we are looking for is AQB.

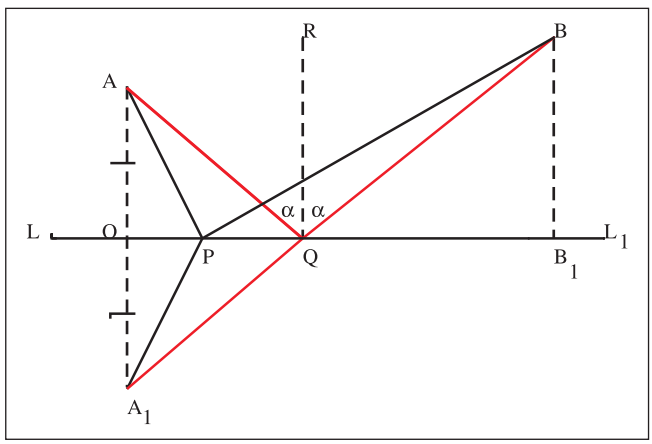
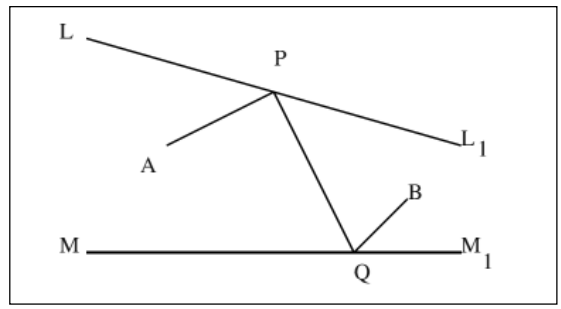


Figure 3.

Figure 4.



Let QR be the perpendicular to the line LL_1 at Q. Then, again by the congruence of the corresponding triangles, we get that $\angle AQO = \angle A_1QO$ which in turn is equal to the vertically opposite angle $\angle BQB_1$ and hence we see that

$$\angle AQR = \angle BQR.$$

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 LL_1 and MM_1 in the plane and two points A and B between them, find the shortest polygonal path APQB where P lies on LL_1 and Q lies on MM_1 (see *Figure* 4). □

Heron's theorem also gives the law governing the reflection of light on a plane mirror which follows from Fermat's principle that light always follows the shortest possible path.

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.

The solution to this is an easy consequence of Heron's theorem. Indeed, fixing the base length and the area implies that the altitude, h, of the triangle is also fixed (see *Figure* 5). Let us draw a line LL_1 parallel to the

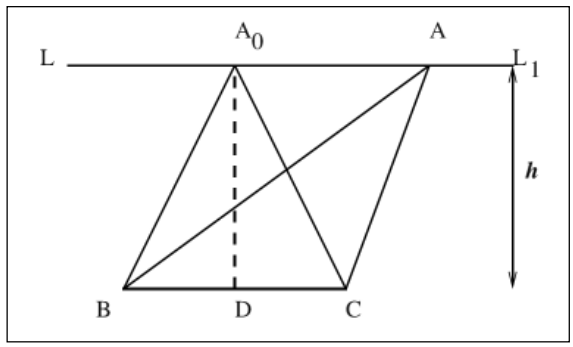


Figure 5.

base BC at a distance h. Then the vertex A can move only on this line.

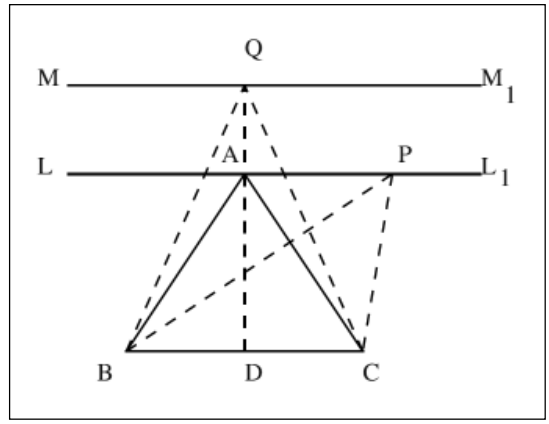
Our problem is to minimize the length of the polygonal path BAC and we know from Heron's theorem that this occurs for the point A_0 on LL_1 where the rays BA_0 and A_0C follow the laws of reflection, i.e., if A_0D is perpendicular to LL_1 (and hence to BC as well), then $\angle BA_0D = \angle CA_0D$. It is then obvious that $A_0B = A_0C$. Thus we have proved the following result.

Theorem 2. *Of all triangles with fixed base and fixed area, the isosceles triangle minimizes the sum of the lengths of the other two sides.*

When studying constrained optimization problems in multivariate calculus, we come across the notion of duality. For instance, when we try to maximize $f(x, y)$ such that $g(x, y) =$ constant, we have the dual problem of minimizing $g(x, y)$ such that $f(x, y) =$ constant. Both these problems, under suitable conditions, have the same optimal solution (x_0, y_0). In the same way, the problem posed above on triangles admits 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.* The answer again is that it is the isosceles traingle. To see this, construct the isosceles triangle with given base BC and such that the sum of the lengths of the other two (equal) sides is the given number, say, ℓ.

When studying constrained optimization problems in multivariate calculus, we come across the notion of duality.

Figure 6.



Draw a line LL_1 parallel to BC through A (see *Figure* 6). Any triangle ΔPBC with base BC and the same area will have its vertex P on the line LL_1. But then, by Theorem 2, we know that $PB + PC > AB + AC = \ell$ and so none of those triangles will qualify as candidates to our optimization problem. Any triangle with the same base but of greater area will have its vertex lying on a line MM_1 parallel to LL_1 (and to BC) at a greater distance. If DA meets MM_1 at Q, then, clearly, $QB+QC > AB+AC = \ell$. By Theorem 2, any triangle with base BC and vertex lying on MM_1 will have the sum of its other two sides greater than $QB+QC > \ell$ and so it will not qualify either. Thus the only other triangles which satisfy our constraint must have area smaller than that of ΔABC. So we now have the following result, which we will use repeatedly in the sequel.

Theorem 3. *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 another proof of this result, see *Box* 1.

Isoperimetric Problem for Polygons

Let $N \geq 3$ be a fixed positive integer. Consider the following problem: *of all N*-sided polygons with the same

Box 1.

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 $2a = \ell$ and the eccentricity e of the ellipse is given by $BC = 2a\mathsf{e}$. The semi-minor axis b is then defined by $b^2 = a^2(1 - \mathsf{e}^2)$. 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 Δ ABC is therefore $b\sin\theta$ and its area is

$$\frac{1}{2}.BC.b\sin\theta\ ,$$

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

perimeter L, find that which encloses the maximum area. 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 (which compares the arithmetic and geometric means of a finite set of positive integers), 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*.

Let us now consider the case $N = 4$, i.e., the case of quadrilaterals. Given a quadrilateral ABCD as in *Figure* 7, by reflecting the point D with respect to the line AC, we produce a quadrilateral ABCE which has the same perimeter but is of larger area. So it is enough to look at convex quadrilaterals.

Figure 7.

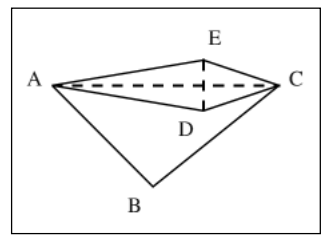
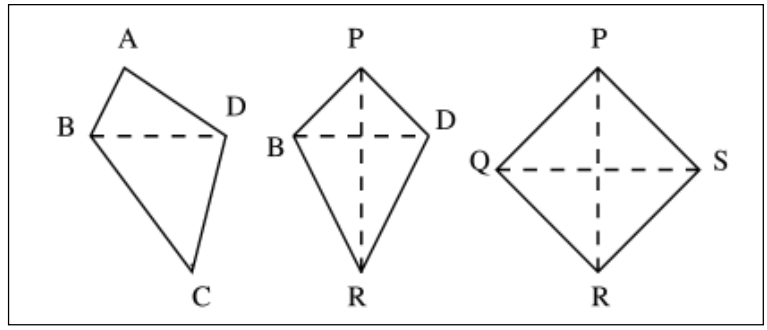


Figure 8.



Given a convex quadrilateral ABCD with perimeter L, construct isosceles triangles ΔPBD and ΔRBD on the diagonal BD such that $PB + PD = AB + AD$ and $RB + RD = CB + CD$ (see *Figure* 8). Then the perimeter of the new quadrilateral PBRD continues to be L. However, by Theorem 3, the area of ΔPBD is greater than that of ΔABD and the area of ΔRBD is greater than that of ΔCBD. Thus, the quadrilateral PBRD has greater area than the quadrilateral ABCD, but is of the same perimeter.

Now construct isosceles triangles ΔPQR and ΔPSR on the diagonal PR of this new quadrilateral such that $QP + QR = BP + BR$ and $SP + SR = DP + DR$. The quadrilateral PQRS still has perimeter L and an appeal to Theorem 3 shows that its area is larger than that of the quadrilateral PBRD, which we saw exceeds that of the original quadrilateral ABCD. By construction, it is clear that all the four sides of the quadrilateral PQRS are equal to each other and so this quadrilateral is indeed a rhombus.

Given any quadrilateral, we can construct a rhombus of the same perimeter but of larger 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 $\theta = \frac{\pi}{2}$.

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

In general, if $N \geq 3$ is a positive integer, we have the following result.

Theorem 4. *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 whose sides are all of equal length and all of whose vertices lie on a circle. In this case, the sides subtend equal angles at the centre of this circle.

Unlike the cases of triangles and quadrilaterals, the proof in the general case is more involved. We will present here an ingeneous 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

Unlike the cases of triangles and quadrilaterals, in the general case we first prove the existence of an optimal polygon and then deduce its properties.

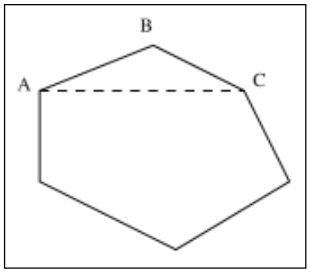


Figure 9.

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 (see *Figure* 9). Let us freeze all the vertices except B.

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 ΔBAC. 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 the optimal polygon must be *equilateral*, i.e., all its sides are equal in length.

Remark. This argument works because we already assumed that the polygon is optimal. Given an arbitrary polygon, we can create another one with the same perimeter and larger area with any *fixed* pair of adjacent sides equal. However, a moment's reflection will show that we cannot repeatedly iterate the method to produce an equilateral polygon of equal perimeter and larger area. □

In the optimal polygon, the diagonal which bisects the perimeter also simultaneously bisects the area.

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

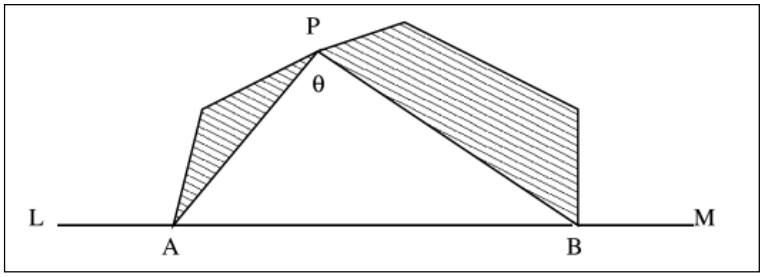


Figure 10.

works only for an even number of sides).

Let AB be the diagonal and consider the upper half of the polygon (see *Figure* 10). Let P be a fixed vertex (other than A and B). We will now vary this figure as follows.The points A and B will be allowed to slide on the line LM defining the diagonal. The vertices will move so that the distances PA and PB are fixed and the areas shaded in the figure (i.e., all parts of the enclosed area, except the triangle ΔPAB) are fixed. The new polygon got by reflecting this figure across the line LM will still be an equilateral polygon with the same perimeter. The change in the area is got by changing the area of the triangle ΔPAB which is given by

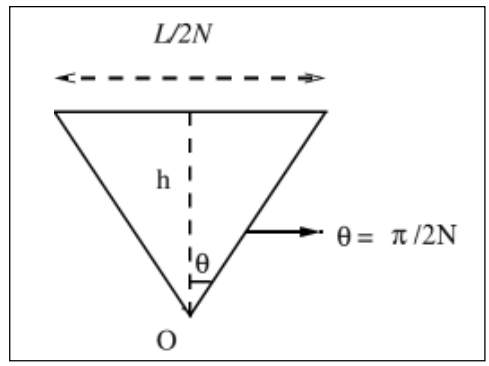
$$\frac{1}{2}PA.PB.\sin\theta\,,$$

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 cyclic with the vertices lying on the circle with the bisecting diagonal as diameter and hence it is a regular polygon.

Remark. Once again, this proof depends on the fact that we are working with the optimal polygon. Given an equilateral polygon bisected by a diagonal, we can create another polygon of the same perimeter and of larger area, with the diagonal subtending a right angle at any specified vertex. But we cannot iterate the procedure to get right angles at all the vertices. □

In the optimal polygon, the diagonal bisecting the perimeter and the area subtends a right angle at all the other vertices.

Figure 11.



This proves Theorem 4 for even-sided polygons. Now, let L be the perimeter and A the area of a $2N$-sided polygon. Consider the corresponding regular polygon with the same perimeter (see *Figure* 11). Then each side has length $L/2N$. The angle subtended by each side at the centre O of the circumscribing circle is $2\pi/2N = \pi/N$. 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 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 5. *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 rectifiable 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 perimenter of this polygon and let A_N be its area.We can choose these points in such a way that, as $N \to \infty$, we have $L_N \to L$ and $A_N \to A$. 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 $N \to \infty$,

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

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.

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.*

Of all simple closed curves of the same perimeter, the circle alone encloses the maximal area.

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 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 closed 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 the 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.

In the case of smooth simple closed curves in the plane, a very nice proof of the isoperimetric inequality can be given using Fourier series [1]. The isoperimetric inequality can be stated in all space dimensions. For a proof in case of smooth domains, see [2].

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] A Sitaram, The isoperimetric problem, *Resonance*, Vol.2, No.9, pp.65–68, 1997.

[2] S Kesavan, The isoperimetric inequality, *Resonance*, Vol.7, No.9, pp.8–18, 2002.

[3] R Courant and H Robbins, *What is Mathematics*?, Second Edition (revised by Ian Stewart), Oxford University Press, 1996.

[4] S Hildebrandt and A Tromba, *The Parsimonious Universe*: *Shape and Form in the Natural World*, Copernicus Series, Springer, 1996.

Address for Correspondence
S Kesavan
The Institute of Mathematical Sciences
CIT Campus, Taramani
Chennai 600 113
Email: kesh@imsc.res.in