THE ISOPERIMETRIC INEQUALITY

S. Kesavan
The Institute of Mathematical Sciences,
CIT Campus, Taramani,
Chennai - 600 113.

Abstract

A new proof (due to X. Cabre) of the classical isoperimetric theorem, based on Alexandrov's idea of moving planes, will be presented. Compared to the usual proofs, which use geometric measure theory, this proof will be based on elementary ideas from calculus and partial differential equations (Laplace equation).

The origin of the study of isoperimetric inequalities goes back to antiquity. Known as *Dido's Problem*, one of the first such inequalities arose when trying to determine the shape of a domain with maximum possible area, given its perimeter. Hence the name isoperimetric inequality (the prefix *iso* stands for 'same' in Greek). The answer to this question is that the circle, and the circle alone, maximizes the area for a given perimeter. Equivalently, given the area enclosed by a simple closed curve, the circle and it alone, minimizes the perimeter.

Nature too plays this game of shape optimization. Why are soap bubbles round? A bubble will attain a position of stable equilibrium if the potential energy due to surface tension is minimized. This, in turn, is directly proportional to the surface area of the air - soap film interface. Thus, for a given volume of air blown to form a bubble, the shape of the bubble will be that for which the surface area is minimized and this occurs only for the spherical shape.

In the case of the plane, the isoperimetric property of the circle was established by Steiner using very ingenious geometric arguments (see the book of Courant and Robbins for a very nice treatment of this). There are two aspects to a proof of this kind. First we *assume* that there is such an optimal shape and deduce that it must be the circle. Next we *prove* the existence of the optimal shape. Steiner's method does not work in three dimensions. Indeed, the proof of the isoperimetric property of the sphere in $\mathbb{R}^3$ was a far more daunting task and was proved in a rather difficult paper by H. A. Schwarz.

An analytic way of looking at this problem is to formulate an isoperimetric inequality. If L is the perimeter of a region in the plane and A is its area, then

$$L^2 \ \geq \ 4\pi A. \tag{1}$$

Thus, whatever be the plane domain of perimeter L, the greatest possible area it can have is $L^2/4\pi$ and this is attained for the circular region and for it alone. This settles the question of the existence and uniqueness of the optimal shape in a single stroke. In the case of three dimensions, if V is the volume of a region and S is the surface area, then the isoperimetric inequality reads as

$$S^3 \ \geq \ 36\pi V^2 \tag{2}$$

with equality only for the sphere. We can generalize this to N - dimensions. Let ω_N denote the volume of the unit sphere in $\mathbb{R}^N$ (Exercise: prove that

$\omega_N = \pi^{N/2}/\Gamma(N/2+1)$ where $\Gamma(s) = \int_0^\infty e^{-x}x^{s-1}dx$ is the usual gamma function). If $\Omega \subset I\!R^N$ is a bounded domain, and $\partial\Omega$ denotes its boundary, then

$$|\partial\Omega| \;\geq\; N\omega_N^{\frac{1}{N}}|\Omega|^{1-\frac{1}{N}} \tag{3}$$

where $|E|$ denotes the N - dimensional (Lebesgue) measure or the $(N-1)$ - dimensional surface measure of a subset E of $I\!R^N$ as the case maybe. Once again, equality is attained in (3) for the sphere and only for the sphere.

The inequality (1) can be proved very easily using Fourier series (cf. for example, Sitaram's article in *Resonance*, 1997, for a very readable exposition). However, for dimensions $N \geq 3$, the proof of (3) is not that immediate. In fact, even the notion of 'surface measure' of the boundary is not obvious. When $N = 2$, we clearly understand the notion of length of a rectifiable curve. In higher dimensions, $\partial\Omega$ will be a $(N-1)$- dimensional manifold and there are several ways to define $|\partial\Omega|$. There are, for instance, the induced $(N-1)$- dimensional surface measure (from $I\!R^N$), the Hausdorff measure, the Minkowski content, the de Giorgi perimeter *etc.* All these notions agree on smooth domains. The differences occur in the presence of singularities on the surface. However, whatever may be the definition chosen, (3) is always true. Indeed, the validity of the classical isoperimetric inequality (with equality only for the sphere) is a criterion for the acceptability of the notion of a surface measure.

In general, the proof uses difficult notions from geometric measure theory. Recently, Cabre (personal communication) has observed that it is possible to use an idea similar to that used by Alexandrov in proving certain estimates for solutions of elliptic partial differential equations to prove the classical isoperimetric theorem. We will present this proof.

While (1) or (3) is referred to as the classical isoperimetric inequality, by *an* isoperimetric problem, we mean today a problem of optimizing some domain dependent functional keeping some geometric parameter of the domain (like its measure) fixed.

Lower Contact Set

Let $f : [a,b] \to I\!R$ be a C^1 (*i.e.* continuously differentiable) function. Let $x_o \in (a,b)$ be a point in the interior such that the graph of the function f lies entirely above the tangent at x_o. Thus, for all $x \in [a,b]$,

$$f(x) \;\geq\; f(x_o) + f'(x_o)(x - x_o). \tag{4}$$

The set S of all points $x_o \in (a,b)$ such that (4) is true for all $x \in [a,b]$ is called the *lower contact set* of the function f. If, in addition, f is twice differentiable, then

$$f(x) \;=\; f(x_o) + f'(x_o)(x-x_o) + \frac{1}{2}f''(x_o)(x-x_o)^2 + o(|x-x_o|^2),$$

where $o(|x-x_o|^2)$ signifies an error term $\varepsilon(x-x_o)$ such that

$$\lim_{y\to x} \varepsilon(x-x_o)/|x-x_o|^2 = 0.$$

From this we deduce that

$$f''(x_o) \;\geq\; 0 \tag{5}$$

for all $x_o \in S$.

Let us now consider a straight line with slope m lying entirely below the graph of the function $f : [a,b] \to I\!\!R$ in the plane. Let us move this line parallel to itself. Eventually, the line must encounter the graph of f. The (abcissa of the) first point of contact could be a, b or in (a,b).

Let us assume that the (abcissa of the) first point of contact, x_o, lies in the interior (a,b). Then, if f is C^1,

$$g(x) \;=\; f(x) - f(x_o) - m(x-x_o) \;\geq\; 0$$

for all $x \in [a,b]$ and is equal to zero, *i.e.* it attains its minimum, at x_o. Thus $g'(x_o) = 0$, *i.e.* $f'(x_o) = m$ and $x_o \in S$.

Hence, any straightline moving parallel to itself from below (the graph of) f and first hitting f at an 'interior point' must do so as a tangent and so the slope of such a line must be in the set $f'(S)$.

If $\Omega \subset I\!\!R^N$ is a bounded domain, and if $f : \overline{\Omega} \to I\!\!R$ is a C^1 function, we can again define its lower contact set, S, analogously as follows:

$$S \;=\; \{x_o \in \Omega \mid f(x) \geq f(x_o) + \nabla f(x_o).(x-x_o) \text{ for all } x \in \overline{\Omega}\} \tag{6}$$

where the dot in the above inequality denotes the usual scalar product in $I\!\!R^N$. Again, if the function is twice differentiable, then

$$f(x) \;=\; f(x_o) + \nabla f(x_o).(x-x_o) + \frac{1}{2}(x-x_o)^T D^2 f(x_o)(x-x_o) + o(|x-x_o|^2)$$

where $D^2 f(x_o)$ denotes the Hessian matrix of second derivatives, *i.e.* the symmetric matrix whose entries are $\frac{\partial^2 f}{\partial x_i \partial x_j}(x_o)$, and $|x-x_o|$ denotes the Euclidean distance in $I\!\!R^N$. We can then easily see that, if $x_o \in S$, then $D^2 f(x_o)$

is a symmetric and positive semi-definite matrix, *i.e.* for all $\xi \in I\!R^N$, we have $\xi^T D^2 f(x_o)\xi \geq 0$.

If we now consider a hyperplane moving parallel to itself, it is easy to see that, analogously, if the first point of contact is an interior point, then the plane becomes the tangent at that point. The direction cosines of the normal to the plane will then belong to the set $\nabla f(S)$, where S is the lower contact set.

These ideas justify the terminology we have used for the set S.

The Neumann Problem

Let $\Omega \subset I\!R^N$ be a bounded and smooth domain. Let Δ denote the Laplace operator, *i.e.*

$$\Delta = \sum_{i=1}^{N} \frac{\partial^2}{\partial x_i^2}.$$

If $\nu(x)$ denotes the unit outward normal to the boundary $\partial\Omega$ at the point $x \in \partial\Omega$, then the outer normal derivative of a differentiable function v is given by

$$\frac{\partial v}{\partial \nu}(x) = \nabla v(x).\nu(x).$$

Let us examine the lower contact set of a solution of the Neumann problem:

$$\left.\begin{array}{rcll} \Delta u & = & f & \text{in } \Omega \\ \frac{\partial u}{\partial \nu} & = & g & \text{on } \partial\Omega. \end{array}\right\} \tag{7}$$

This problem will have (an infinite number of) solutions if, and only if, f and g satisfy the compatibility condition (see Box 1)

$$\int_\Omega f = \int_{\partial\Omega} g. \tag{8}$$

Let us now take $g \equiv 1$ on $\partial\Omega$. Then, if we take f to be constant on Ω, it follows from (8) that

$$f \equiv \frac{|\partial\Omega|}{|\Omega|}. \tag{9}$$

Let m be an arbitrary vector in $I\!R^N$. There do exist planes, of the form $z = m.x + c$, lying below the graph of u. Moving the plane parallel to itself, we will eventually meet the graph of u. If the (abcissa of the) first point of contact lies in Ω, we already observed that the plane becomes a tangent to u and that $m \in \nabla u(S)$, where S is the lower contact set of u.

On the other hand, if the (abcissa of the) first point of contact $x_o \in \partial\Omega$, then, for all $x \in \overline{\Omega}$,

$$g(x) \;\equiv\; u(x) - u(x_o) - m.(x - x_o) \;\geq\; 0$$

and $g(x_o) = 0$ is the minimum and is attained on the boundary. It follows that if $\nu(x_o)$ is the unit outward normal of $\partial\Omega$ at x_o, then,

$$\nabla g(x_o).\nu(x_o) \;\leq\; 0$$

since g is decreasing in that direction at x_o. Thus,

$$m.\nu(x_o) \;\geq \nabla u(x_o).\nu(x_o) \;=\; \frac{\partial u}{\partial \nu}(x_o) \;=\; 1.$$

Hence $|m| \geq 1$. Therefore, if $|m| < 1$, the moving plane of the form $z = m.x + c$ can meet the graph of u only as a tangent at an 'interior point' and so $m \in \nabla u(S)$. We have thus established the following result.

Lemma 1 *If $B_1(0)$ denotes the ball of unit radius in $I\!R^N$ having its centre at the origin, then*

$$B_1(0) \;\subset\; \nabla u(S).\blacksquare \tag{10}$$

The Isoperimetric Inequality

Let $\Omega \subset I\!R^N$ be a smooth domain and let u be a solution of (7) when $g \equiv 1$ and f given by (9). Then u will be a smooth function. Then, by Lemma 1,

$$\omega_{_N} = |B_1(0)| \leq |\nabla u(S)| = \int_{\nabla u(S)} dx \leq \int_S |\det(D^2u)| dx,$$

by the change of variable formula applied to the mapping $\nabla u : \Omega \to I\!R^N$. The reason we have an inequality for the last term is that this mapping may not be a diffeomorphism (see Box 2). Now, recall that, on S, D^2u is symmetric and non-negative definite. Hence its eigenvalues and, therefore, its determinant will be non-negative. Thus,

$$\omega_{_N} \leq \int_S \det(D^2u) dx \leq \int_S \left(\frac{\mathrm{tr}(D^2u)}{N}\right)^N dx$$

by the AM-GM inequality. But $\mathrm{tr}(D^2u) = \Delta u$ and by (7) and (9), we get

$$\omega_N \leq \int_S \left(\frac{|\partial\Omega|}{N|\Omega|}\right)^N dx \leq \int_\Omega \left(\frac{|\partial\Omega|}{N|\Omega|}\right)^N dx = \frac{|\partial\Omega|^N}{N^N|\Omega|^{N-1}}$$

from which we easily deduce (3).

This proves the inequality for smooth domains. For general domains, depending on the definition of the surface measure, the inequality usually follows by approximation of the domain by smooth domains.

The Equality Case

Let us now assume that $\Omega \subset I\!R^N$ is a smooth domain such that equality is attained in (3). We will show that Ω must be a ball. (It is obvious that, conversely, if Ω is a ball, then we do have equality in (3); for, $|\Omega| = \omega_N r^N$ and $|\partial\Omega| = N\omega_N r^{N-1}$, where r is the radius of the ball.)

If we have equality in the isoperimetric inequality, then, retracing the proof of the theorem presented in the previous section, we see that all the inequalities become equalities. In particular, we get that $|S| = |\Omega|$, *i.e.* $\Omega\backslash S$ has measure zero, and so S is dense in Ω. But it is immediate to see from (6) that, since u is smooth, $S = \Omega$. Further, on $S(= \Omega)$, we have equality in the AM-GM inequality for the eigenvalues of D^2u and so the eigenvalues are all equal, *i.e.* D^2u is a scalar matrix. Thus

$$\frac{\partial^2 u}{\partial x_i \partial x_j}(x) \equiv \lambda(x)\delta_{ij} \tag{11}$$

where the delta on the right-hand side is the usual Kronecker delta and $\lambda(x)$ is easily seen to be a constant given by

$$\lambda = \left(\frac{\omega_N}{|\Omega|}\right)^{1/N}.$$

Next, since $B_1(0) \subset \nabla u(S)$ and, in the equality case, both have the same measure, we have that $B_1(0)$ is dense in $\nabla u(S) = \nabla u(\Omega)$. By the smoothness of u, it thus follows that $|\nabla u| \leq 1$ in $\overline{\Omega}$. But, on the boundary, $|\nabla u| \geq |\frac{\partial u}{\partial \nu}| = 1$. Thus, $|\nabla u| \equiv 1$ on the boundary and this implies that the tangential component of the gradient is zero on the boundary. Thus u is constant on the boundary. Since any two solutions of (7) differ by an additive constant, we may henceforth assume that

$$\left.\begin{array}{rcll} \Delta u & = & \frac{|\partial\Omega|}{|\Omega|} & \text{in } \Omega \\ u & = & 0 & \text{on } \partial\Omega \\ \frac{\partial u}{\partial\nu} & = & 1 & \text{on } \partial\Omega. \end{array}\right\} \quad (12)$$

By the maximum principle (see Box 3), $u < 0$ in Ω and so u must attain a minimum at a point x_o in Ω. Clearly $\nabla u(x_o) = 0$.

Let B be the largest possible ball in Ω with centre at x_o. Now, if $x \in B$, then for some ξ in the line segment joining x and x_o, we have, by the mean value theorem,

$$\begin{array}{rcl} u(x) & = & u(x_o) + \nabla u(x_o).(x - x_o) + \frac{1}{2}(x - x_o)^T D^2 u(\xi)(x - x_o) \\ & = & u(x_o) + \frac{\lambda}{2}|x - x_o|^2. \end{array}$$

By the nature of B, there must be a point on ∂B which also lies on $\partial\Omega$ and so $u = 0$ at that point. But by the above formula, it then follows that $u = 0$ on all of ∂B. Since $u < 0$ in Ω, this will be possible only if B coincides with Ω, *i.e.* Ω is a ball. In fact, if $-M = u(x_o) < 0$ is the minimum of u, then

$$\partial\Omega \;=\; \{x \in \overline{\Omega} \mid u(x) = 0\} \;=\; \{x \mid |x - x_o|^2 = \frac{2M}{\lambda}\}.$$

Remark: Problem (12) is an overdetermined boundary value problem. Serrin formulated a method which was further developed as the *method of moving planes* by Gidas, Ni and Nirenberg to study symmetry properties of positive solutions of semilinear elliptic equations. This has been further refined by Berestycki and Nirenberg. Their method uses, in an essential way, maximum principles. In particular, a maximum principle in 'small domains' is very useful and it was proved by Varadan using an estimate for solutions of second order elliptic equations due to Alexandrov, Bakelman and Pucci. This last estimate was proved using the idea of the lower contact set and an inclusion analogous to that stated in Lemma 1, and inspired Cabre to imitate it to suggest the proof of the isoperimetric inequality presented here.■

Suggested Reading

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

- Sitaram, A. (1997) The isoperimetric problem, *Resonance*, **2**, No. 9, pp. 65 - 68.

Box 1

If A is a square matrix of order n which is singular, *i.e.* there exists a non-zero vector x_o such that $Ax_o = 0$, then the system of linear equations $Ax = b$ either has no solution or an infinity of solutions according as b does not or does satisfy a compatibility condition. If the matrix is symmetric (or self-adjoint, in the complex case) the condition is $b.x_o = 0$ for any x_o in the null space of A. For, if $Ay = b$, then

$$b.x_o \;=\; Ay.x_o \;=\; y.A^T x_o \;=\; y.Ax_o \;=\; 0.$$

By dimension arguments, it can also be shown that this condition is sufficient. The situation in the case of the Neumann problem is very similar. The problem (7) can be put in the form of a linear equation in an infinite dimensional Hilbert space with the linear operator being what is known as a self-adjoint compact operator. Such operators have properties very similar to those of linear operators in finite dimensional spaces. In particular, if $f = 0$ and $g = 0$ in (7) we have non-trivial solutions *viz.* constant functions, as solutions to (7). Thus, in the general case, (7) either has no solution or an infinite number of solutions according as (8) is not or is satisfied. This property is known is functional analysis as the *Fredhölm Alternative.*

The necessity of (8) is easy to see by just integrating both sides of the differential equation in (7) and applying Green's (*i.e.* integration by parts) formula to the term on the left-hand side.

Box 2

If $T = (T_1, T_2, ..., T_N) : \Omega \to T(\Omega)$, where Ω is an open set in $\mathbb{R}^N$, is a C^1-diffeomorphism (*i.e.* T is invertible and both T and T^{-1} are C^1- mappings), then for a subset S of Ω, by the change of variable formula

$$\int_{T(S)} dx \;=\; \int_S |\det(T'(x))| dx$$

where $T'(x)$ is the Jacobian matrix whose entries are $\frac{\partial T_i}{\partial x_j}(x)$. However, if T is not a diffeomorphism, we have that the equality in the above relation is replaced by the inequality "$\leq$". For example consider $T : \mathbb{R} \to [0, +\infty)$ given by $T(x) = x^2$ and $S = [-1, 1]$.

In our case $T = \nabla u$ and so $T'(x) = D^2 u(x)$.

Box 3

Many of you would have come across something called the maximum modulus principle when studying analytic functions in the complex plane. The real and imaginary parts of an analytic function are harmonic functions, *i.e.* they satisfy the equation $\Delta u = 0$. Solutions of the Laplace equation (and, more generally, those of a class of partial differential equations known as elliptic second order equations, of which the Laplace equation is the prototype) enjoy special properties which go under the name of *maximum principles*. For instance, the weak maximum principle states that if $\Delta u \geq 0$ in a domain and if $u \leq 0$ on the boundary, then $u \leq 0$ in the domain as well. The strong maximum principle then asserts that either u is identically equal to a constant in the closure of the domain or u attains its maximum *only* on the boundary. In particular, if $u = 0$ on the boundary and if u were non-constant (as is the case in the Neumann problem (12) considered above, since $\Delta u \neq 0$), we deduce that $u < 0$ in the interior of the domain.