Dido's Problem and Its Impact on Modern Mathematics

Catherine Bandle

ABSTRACT. We trace the isoperimetric problem from Queen Dido to some recent applications. Emphasis is put on the developments which are significant for the applications in analysis and in partial differential equations.

Ancient Time, Origin of the Problem

Dido Purchases Land for the Foundation of Carthage. Engraving by Matthäus Merian the Elder, in *Historische Chronica*, Frankfurt a.M., 1630. Dido's people cut the hide of an ox into thin strips and try to enclose a maximal domain.

Figure 1. Queen Dido enclosed maximal area with strips of the hide of a bull.

Dido's Problem

The Roman poet Publius Vergilius Maro (70–19 B.C.) tells in his epic *Aeneid* the story of queen Dido, the daughter of the Phoenician king of the 9th century B.C. After the assassination of her husband by her brother she fled to a haven near Tunis. There she asked the local leader, Yarb, for as much land as could be enclosed by the hide of a bull. Since the deal seemed very modest, he agreed. Dido cut the hide into narrow strips, tied them together and encircled a large tract of land which became the city of Carthage. Dido faced the following mathematical problem, which is also known as the *isoperimetric problem*:

Find among all curves of given length the one which encloses maximal area.

Dido found intuitively the right answer.

Catherine Bandle is professor emerita at the University of Basel. Her e-mail address is `c.bandle@gmx.ch`.

For permission to reprint this article, please contact: `reprint-permission@ams.org`.

DOI: http://dx.doi.org/10.1090/noti1576

The Isoperimetric Problem in Antiquity

In those days a formula for the area A of a circle of given length L was known. The Babylonians used around 1800 B.C. the formula $A = \frac{2}{25}L^2$ instead of $A = \frac{1}{4\pi}L^2$. This approximation of π by 3.125 is quite accurate. In his short treatise *On the Measurement of the Circle*, Archimedes (285–212 B.C.) circumscribed and inscribed a 96-gon around and inside the circle and determined that $3.1408 < \pi < 3.14285$.

The Greeks were interested in the isoperimetric problem also for practical reasons. It was useful to have an upper limit for the area in order to prevent the merchants from cheating when they stated the area of an island by its circumference. In those times it was commonly believed that the perimeter of a figure determines its area.

Around 150 B.C. Zenodorus proved rigorously by elementary geometrical arguments:

(i) *if there exists an n-gon having the largest area among all n-gons of given perimeter, it must be regular;*

(ii) *among all regular polygons of equal perimeter the one with more sides has a greater area; and*

(iii) *the circle encloses a greater area than any regular polygon of equal perimeter.*

Does a maximal polygon exist at all? The ancient geometers were not concerned with this question. It was settled many centuries later, for instance by Weierstrass.

No substantial mathematical progress was made for almost 1,900 years on the isoperimetric problem. During all this time it was taken for granted that the circle has the largest area among all plane domains of given perimeter. Similarly, the astronomers believed that an analogous property is also valid for domains in space namely, that the ball has the largest volume of all domains of given surface area. A detailed description of the isoperimetric problem in ancient time is given in [5].

Towards a Proof

The Eighteenth Century, Calculus of Variations

In 1744 Euler, motivated by the isoperimetric problem, which was suggested to him by the two brothers Johann and Jakob Bernoulli, laid down the foundations of the calculus of variations.

We shall explain his idea for a variant of Dido's problem where Dido wanted to secure access to the sea. Analytically it can be phrased as follows. Look for a graph $y(x), 0 < x < a$ with $y(0) = y(a) = 0$ and such that the area between $y(x)$ and the x-axis, $\int_0^a y(x)\,dx$, is maximal while the length, $L = \int_0^a \sqrt{1 + (y'(x))^2}\,dx$, is kept fixed.

Euler's idea was to reduce it to a finite-dimensional problem that can be handled with the standard method of calculus. He divided the interval $(0, a)$ into n intervals of equal length $x_{k+1} - x_k = h$ and searched for a polygon $\mathcal{P}(h)$ with the corners $C_k = (x_k, y_k)$ of given perimeter L and maximal area below its graph. He first solved this problem within the class of polygons with the corners C_k, $k \neq m, m+1$, $(x_m, y_m + \epsilon)$, and $(x_{m+1}, y_{m+1} + \delta)$, where ϵ, δ are free parameters. Letting $h \to 0$ he was able to show that the optimal graph $y(x)$ satisfies the *Euler-Lagrange differential equation* $\Phi_y - \frac{d}{dx}\Phi_{y'} = 0$, where $\Phi = y + \lambda\sqrt{1 + y'^2}$ and $\lambda \in \mathbb{R}$ is the Lagrange multiplier.

This so-called *Euler's polygon method* is still used for numerical treatment of ODEs.

Eleven years after the appearance of his treatise on the calculus of variations Euler received a letter from the nineteen-year-old Lagrange, who wrote that he had found a more general method. Instead of making pointwise changes, he perturbed the whole curve, the way it is still done today. Euler was very pleased and admitted that much deeper results could be obtained by Lagrange's approach.

Figure 2. Jacob Steiner applied symmetrization to Dido's problem.

The Nineteenth Century, Final Proofs of the Isoperimetric Problem

Many ingenious geometrical arguments were proposed by J. Steiner (1796–1863) to prove the isoperimetric property of the circle and the ball. He always took a figure that is not a circle or a sphere and showed that the area or volume can be increased by keeping the perimeter fixed. Edler, based on one of Steiner's arguments, showed that any domain in the plane with the same perimeter as the circle has a smaller area. This completed the mathematical proof of the isoperimetric problem in the plane.

Steiner invented the technique of *symmetrization*, which has become a key tool in geometric analysis. The aim is to transform a set $\Omega \subset \mathbb{R}^n$ such that the volume (Lebesgue measure) remains unchanged and the perimeter decreases.

Steiner Symmetrization. Let Ω be a domain in $\mathbb{R}^n$. Denote by x an arbitrary point in $\mathbb{R}^{n-1}$ and by e_n the unit vector perpendicular to $\mathbb{R}^{n-1}$. For each $x \in \mathbb{R}^{n-1}$ consider the line $\ell_x := \{x + x_n e_n : x_n \in \mathbb{R}\}$. The 1-dimensional Lebesgue measure of the slice $\ell_x \cap \Omega$ will be denoted by $\mathcal{L}_1(x)$. ($\mathcal{L}_1(x) = 0$ if $\ell_x \cap \Omega = \emptyset$). The Steiner symmetrization of Ω with respect to the hyperplane $\mathbb{R}^{n-1}$ is the domain $S\Omega$, symmetrically balanced around $\mathbb{R}^{n-1}$ and which has the property that each slice $\ell_x \cap S\Omega$ is an interval of length $\mathcal{L}_1(x)$.

By Cavalieri's principle the volume of Ω is equal to the volume of $S\Omega$. The fact that the perimeter decreases is more subtle.

It is intuitive that after infinitely many symmetrizations with respect to all possible hyperplanes a body is transformed into a ball. Lusternik (1935) showed that there exist countably many symmetrizations S_k such that $\Pi_{k=1}^{n} S_k(\Omega) \to \Omega^*$ as $n \to \infty$,[1] where Ω^* is the ball with the same volume as Ω. By performing infinitely many symmetrizations with respect to hyperplanes containing the x_n-axis one obtains the

Schwarz Symmetrization. In this case every horizontal slice of Ω at the height $x_n = c$ is replaced by an $(n-1)$-dimensional ball of the same $(n-1)$-dimensional volume, centered at the x_n-axis at height c. A domain $\Omega \subset \mathbb{R}^3$ is

[1] *Here the convergence is with respect to the Blaschke metric $d(A, C) := \inf h$ such that $A \subset C_h$, $C \subset A_h$ where A_h and C_h are the exterior parallel sets of A and C at distance h.*

transformed by Schwarz symmetrization into a surface of revolution with the x_3-axis as axis of revolution.

Again, as for the Steiner symmetrization, the volume does not change and the perimeter does not increase.

This symmetrization together with the calculus of variations enabled H. A. Schwarz to give in 1884 the first rigorous proof of the isoperimetric property of the ball in the class of domains with piecewise analytic boundaries. He was also the first who pointed out the lack of an existence theorem in Steiner's proofs. It turns out—as we will see later—that the symmetrizations play a crucial role in modern analysis and in mathematical physics.

The isoperimetric problem in higher dimensions consists in finding among all domains of given perimeter the one with maximal volume. The solution is the ball. Its proof is much more delicate than in the plane. One reason is that the convex hull has not necessarily a smaller perimeter. It was solved in the most elegant way by means of an inequality derived by H. Brunn (1887) and H. Minkowski (1896) for convex sets and then generalized to nonconvex sets by L. A. Lyusternik (1935).

The Brunn-Minkowski Inequality. Let A and C be two domains in $\mathbb{R}^n$. Then the Lebesgue measure of the *Minkowski sum* $A + C$ is bounded from below by the Lebesgue measures of A and C as follows

$$|A + C|^{\frac{1}{n}} \geq |A|^{\frac{1}{n}} + |C|^{\frac{1}{n}}.$$

If $B(h)$ denotes the ball of radius h, centered at the origin, then $A + B(h) = A_h$ is the exterior parallel set at distance h. The limit inferior of the quotient $(|A_h| - |A|)/h$ as h tends to zero is called *Minkowski content* $\mathcal{M}(A)$ of A. For domains with smooth boundaries the Minkowski content coincides with the classical surface area. From $|A_h|^{\frac{1}{n}} - |A|^{\frac{1}{n}} \geq |B(h)|^{\frac{1}{n}}$ the isoperimetric inequality

$$n|B(1)|^{\frac{1}{n}}|A|^{\frac{n-1}{n}} \leq \mathcal{M}(A)$$

is immediate. Equality holds for the ball A^*.

Modern Times

Further Developments

The power of Steiner's work gave rise to a revival of the isoperimetric problem. A rich collection of geometrical and analytical proofs are now available, in particular for plane domains. It would be beyond the scope of this article to go into details. Interested readers are referred to the many reviews, books, and papers, for instance [2], [4], [7].

An interesting direction is the so-called *Bonnesen-type* inequalities for the *isoperimetric deficit* which measures how much a curve differs from a circle. Steiner made an important discovery that was exploited—at least in the plane and on two-dimensional surfaces—to tackle this problem. He observed that the volume and the perimeter of an exterior parallel set of a convex body at distance h can be expressed as a polynomial in h. For example, in two dimensions the area $A(h)$ and the perimeter $L(h)$ of Ω_h are expressed as follows:

$$A(h) = A(0) + L(0)h + \pi h^2, \text{ and } L(h) = L(0) + 2\pi h.$$

This implies that $L^2(h) - 4\pi A(h)$ is independent of h. If it were possible to show that for a parallel curve it is positive, the isoperimetric inequality would follow. Unfortunately, the exterior parallel curves don't seem suitable for such a conclusion.

A remedy was proposed by Bol. He studied the deficit of the interior parallel sets. Even for a smooth domain, $\Omega_{-h} := \{x \in \Omega : \text{distance}\{x, \partial\Omega\} > h\}$ has more and more corners and spikes as the distance increases. It is not clear at first if their boundary curves $\partial\Omega_{-h}$ are rectifiable. By a clever trick, Sz.-Nagy showed that for a simply connected domain the perimeter $L(-h)$ of Ω_h is usually rectifiable and satisfies for almost all h the inequalities

$$A(0) \leq A(-h) + L(0)h - \pi h^2 \text{ and } L(-h) \leq L(0) - 2\pi h.$$

This implies that $4\pi A(0) - L^2(0) \leq 4\pi A(-h) - L^2(-h)$, and the isoperimetric inequality follows.

The method of interior parallel curves was successfully generalized by Bol and later by Fiala and Hartmann to simply connected domains on two-dimensional surfaces, bounded by a Jordan curve of class C^2. A. D. Alexandrov extended these results by polyhedral approximation and A. Huber proved Fiala's inequality by methods of potential theory. A survey of this topic is found in Osserman's papers [6], [7].

The extension of the Bonnesen-type inequalities—which today are rather known as *quantitative isoperimetric inequalities*—has attracted many mathematicians, for instance Fusco and his collaborators, and has led to extensive and excellent research.

The isoperimetric property of the ball holds also in spaces of constant curvature such as the sphere and the hyperbolic space in general dimensions. The first lengthy proof for arbitrary dimensions is due to Schmidt (1943). Many shorter alternatives are now available.

For application in crystallography anisotropic inequalities are of interest. The anisotropic surface energy is a generalization of the perimeter. The problem is to find the set of given measure for which this energy is minimal. The best known result in this direction is *Wulff's inequality*.

Notions of Perimeter[2]

In the plane the definition of the perimeter is straightforward. If the boundary curve is *rectifiable*, it is the supremum of the perimeters of the polygon approximations. If it is not rectifiable we set $L = \infty$. In higher dimensions the definition for nonsmooth domains causes some problems. Several notions have been proposed that apply to boundaries with all kinds of wiggles. Especially for applications in the calculus of variations it is desirable to have a perimeter that is lower semicontinuous with respect to domain convergence. This is not the case for the $(n-1)$-dimensional Hausdorff measure $\mathcal{H}^{n-1}$ nor for the Minkowski content $\mathcal{M}$.

The definition that is now standard in analysis is the perimeter of Caccioppoli and De Giorgi. In geometric terms it is described as follows. If Ω is measurable and

[2] *See "WHAT IS… Perimeter" in this issue of Notices, page 1009.*

$\mathcal{P} = \{P_i\}_1^\infty$ is a series of polyhedra converging to Ω, then the *perimeter of Caccioppoli and De Giorgi* is defined by

$$P(\Omega) = \inf_{\mathcal{P}}(\liminf \mathcal{H}^{n-1}(\partial P_i)).$$

Analytically it is given by

$$P(\Omega) = \sup\left\{\int_\Omega \operatorname{div}\phi dx : \phi \in C_0^\infty(\Omega, \mathbb{R}^n), \max \sum_{k=1}^n \phi_k^2(x) \le 1\right\}.$$

De Giorgi has shown that it is lower semicontinuous and satisfies the isoperimetric inequality. Moreover $P(\Omega) \le \mathcal{H}^{n-1}(\Omega)$.

Analytic Tools

In analysis the method of symmetrization leads to a useful transformation of measurable functions $u : \Omega \to \mathbb{R}^+$. Denote by $\mu(t)$ its distribution function $|\{x \in \Omega : u(x) > t\}|$. As before we write A^* for the ball of the same volume as $A \subset \mathbb{R}^n$, centered at the origin. The *Schwarz symmetrized (rearranged)* function $u^* : \Omega^* \to \mathbb{R}^+$ has the same distribution function as u and its level surfaces are balls. Its main properties are:

1. $\int_\Omega g(u)\,dx = \int_{\Omega^*} g(u^*)\,dx$ for any continuous function $g : \mathbb{R} \to \mathbb{R}$;
2. if u vanishes on the boundary and if $p > 1$, then $\int_\Omega |\nabla u|^p\,dx \ge \int_{\Omega^*} |\nabla u^*|^p\,dx$.

These properties are valid for large classes of functions. Talenti (1976) was the first who extended them to functions belonging to Sobolev spaces. For additional properties of the symmetrization of functions and further references see Brock's survey article [3].

An opposite transformation of the symmetrization is the *harmonic transplantation.* It arose from the conformal transplantation in complex function theory and was introduced by Hersch (1969). In this case a radial function $u : B(R) \to \mathbb{R}^+$ is transformed into a function $U : \Omega \to \mathbb{R}^+$ by means of the Green's function. If Ω has the *harmonic (conformal) radius* R, then $\int_\Omega |\nabla U|^p\,dx = \int_{B(R)} |\nabla u|^p\,dx$ and for any continuous, positive function g there holds $\int_\Omega g(U)\,dx \ge \int_{B(R)} g(u)\,dx$.

A further link between analysis and geometry is the *co-area formula* relating an integral to integrals over slices by level sets of a function u. Let $u : \Omega \to \mathbb{R}$ be Lipschitz, hence almost everywhere differentiable. Then for any measurable $g : \Omega \to \mathbb{R}$

$$\int_\Omega g(x)|\nabla u(x)|\,dx = \int_{\mathbb{R}}\left(\int_{u^{-1}(t)} g(x)\,d\mathcal{H}^{n-1}(x)\right)dt.$$

Fleming and Rishel have extended this formula to functions whose distributional gradient is of bounded variation. One has

$$\int_\Omega |\nabla f|\,dx = \int_{\mathbb{R}} P_\Omega\{x \in \Omega : f(x) > t\}\,dt,$$
$$(P_\Omega(A) = P(A \cap \Omega)).$$

This relation applies also to functions in Sobolev spaces.

Applications

Mathematical Physics

There are several physical quantities that depend on the shape of the domain and for which the ball is optimal. Pólya and Szegö have made extensive use of symmetrizations to treat such problems.

It is now customary to call an inequality *isoperimetric* if it relates quantities associated with the same domain and if the equality sign is attained for some domain. The history of isoperimetric inequalities in mathematical physics began with the conjectures of St. Venant (1856) for the torsional rigidity and of Rayleigh (1877) for the principal eigenvalue of a membrane. For more information on this topic see for instance [8], [1], [3]. The strength of symmetrization shall be illustrated by

The Rayleigh-Faber-Krahn Inequality. Let Δ be the Laplace operator and consider the eigenvalue problem $\Delta\phi + \lambda\phi = 0$ in the domain Ω and $\phi = 0$ on the boundary $\partial\Omega$. The lowest eigenvalue is characterized by the Rayleigh principle

$$\lambda_1(\Omega) = \inf_{C_0^\infty(\Omega)} \frac{\int_\Omega |\nabla v|^2\,dx}{\int_\Omega v^2\,dx}.$$

By symmetrizing the trial functions and using the properties of the symmetrized functions mentioned above we conclude that $\lambda_1(\Omega) \ge \lambda_1(\Omega^*)$. In the same way Talenti (1976) computed the best Sobolev constants.

Symmetrizations and PDEs

Symmetrizations can be used not only to estimate energies related to elliptic and parabolic problems, but also to derive information on the distribution function of the solutions [1], [3]. This topic has attracted many mathematicians and has been exploited in all possible directions.

We describe this idea with the simple problem $\Delta u + 1 = 0$ in $\Omega \subset \mathbb{R}^2$ with $u = 1$ on $\partial\Omega$. Denote by $\mu(t)$ the area of $\Omega_t := \{x : u(x) > t\}$ and by $L(t)$ its perimeter. Then integration over Ω_t implies that $\oint_{\partial\Omega_t} |\nabla u|\,ds = \mu(t)$, where s denotes the arclength. The co-area formula together with Sard's lemma yields for almost every $t > 0$

$$-\mu'(t) = \oint_{\partial\Omega_t} \frac{ds}{|\nabla u|}.$$

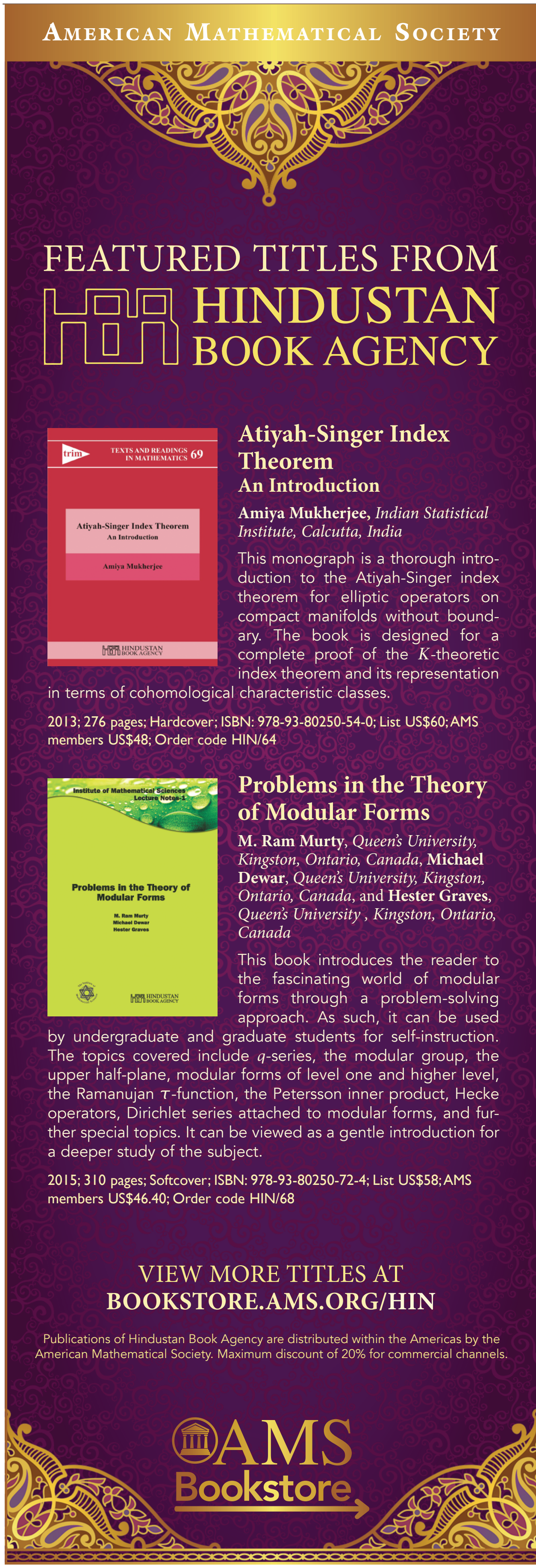
By the isoperimetric and the Schwarz inequalities we get for almost every t

$$4\pi\mu(t) \le L^2(t) \le \oint_{\partial\Omega_t} |\nabla u|\,ds \oint_{\partial\Omega_t} \frac{ds}{|\nabla u|}\,d\mathcal{H}^1 = \mu(t)(-\mu'(t)).$$

Consequently $4\pi t \le -\mu(t) + |\Omega|$ and thus $u^* \le U$, where U is the solution in Ω^*.

Shape Derivatives and Optimization

A direct approach to characterize the optimal domain is the variation of domains. In the spirit of calculus this can be done by studying the volume and the perimeter of infinitesimal changes of the domain. Suppose that $\Omega \subset \mathbb{R}^{n-1}$ is a smooth domain and let ν be its outer normal. If we displace each point on $\partial\Omega$ by the vector $t\eta\nu$,



where η is a smooth function on $\partial\Omega$, we obtain for small $t \in (-\epsilon, \epsilon)$ a family of domains Ω^t. Their perimeter is $P(t)$ and their volume $V(t)$. The first variations (shape derivatives) at $t = 0$ are

$$\frac{dV}{dt}(0) = \oint_{\partial\Omega} \eta \, d\mathcal{H}^{n-1} \text{ and } \frac{dP}{dt}(0) = -\oint_{\partial\Omega} H\eta \, d\mathcal{H}^{n-1},$$

where H is mean curvature of $\partial\Omega$. From here it is not difficult to infer that the solution of the isoperimetric problem must be a surface of constant mean curvature. This type of argument is now widely used to treat problems in shape optimization.

References

[1] C. Bandle, *Isoperimetric inequalities and applications*, Pitman, 1980. MR 572958

[2] V. Blåsjö, The isoperimetric problem, *Amer. Math. Monthly* **112** (2005), 526–566. MR 2142606

[3] F. Brock, Rearrangements and applications to symmetry problems in PDE, Vol. 4, M. Chipot editor, *Handbook of Differential Equations*, Elsevier, 2007, 1–60. MR 2569330

[4] Yu. D. Burago and V. A. Zalgaller, *Geometric Inequalities*, Springer-Verlag, New York, 1988; Original Russian edition: Geometricheskie neravenstva, Leningrad. MR 936419

[5] H. Gericke, Zur Geschichte des isoperimetrischen Problems, *Math. Sem. Ber.* **29** (1982), 160–187. MR 681046

[6] R. Osserman, Bonnesen-style isoperimetric inequalities, *Amer. Math. Monthly* **86** (1979), 1–29. MR 519520

[7] R. Osserman, The isoperimetric inequality, *Bull. Amer. Math. Soc.* **84** (1978), 1182–1238. MR 0500557

[8] L. E. Payne, Isoperimetric inequalities and their applications, *SIAM Review* **9** (1967), 453–488. MR 0218975

Image Credits

Figures 1 and 2 are in the public domain.
Author headshot courtesy of Catherine Bandle.

Catherine Bandle

ABOUT THE AUTHOR

Catherine Bandle's areas of expertise include elliptic and parabolic partial differential equations, the calculus of variations, and shape optimization.