Canad. Math. Bull. Vol. **24** (2) 1981

THE ISOPERIMETRIC INEQUALITY FOR CURVES WITH SELF-INTERSECTIONS

BY
ANDREW VOGT

Banchoff and Pohl [3] have proved the following generalization of the isoperimetric inequality.

THEOREM. *If γ is a closed, not necessarily simple, planar curve of length L, and $w(p)$ is the winding number of a variable point p with respect to γ, then*

$$\int_{R^2} w^2\, dA \le \frac{L^2}{4\pi}, \tag{1}$$

with equality holding if and only if γ is a circle traversed a finite number of times in the same sense.

The object of this note is to give an elementary proof of this result based on the classical isoperimetric inequality, polygonal approximations, and some familiar theorems of real analysis. The identity

$$\int_{R^2} w\, dA = \frac{1}{2}\int_\gamma (x\, dy - y\, dx), \tag{2}$$

which relates two different formulas for signed area, is also established here, as a corollary.

In [3] Banchoff and Pohl actually proved an n-dimensional version of the theorem above, using methods of integral geometry and making smoothness assumptions not needed in the planar case. Federer and Fleming, in their theory of currents, also obtained an abstract isoperimetric inequality [5, p. 486, line 8] of which the inequality (1) seems to be a special case. The identity (2) was proved by Radó [9]. Radó later [10] proved an inequality similar to (1), with w^2 replaced by the smaller quantity $|w|$. These developments and others are reviewed in the survey article of Osserman [8].

We consider a closed rectifiable curve γ in the plane—that is, a continuous map taking real numbers t in some interval $[a, b]$ into points $p(t) = (x(t), y(t))$ in R^2 in such a way that $p(a) = p(b)$ and that the length

$$L(\gamma) = \sup\left\{\sum_{i=0}^{n-1} d(p(t_i), p(t_{i+1})) : n \ge 2,\ a = t_0 \le \cdots \le t_n = b\right\}$$

Received by the editors March 13, 1979

https://doi.org/10.4153/CMB-1981-026-8 Published online by Cambridge University Press

is finite. Here d is the Euclidean metric on R^2. To avoid triviality, we assume that $L(\gamma)>0$.

For a point p of R^2 not in range γ, the winding number $w(p)$ of p with respect to γ (see [1, Chap. 2] or [7, Chap. 7]) is equal to $(\theta(b)-\theta(a))/2\pi$, where $\theta(t)$ is the angle at p between $p(t)$ and a fixed point $q\neq p$, measured counterclockwise from q and continuously as t varies. The function w is integer-valued, the set $\{p: w(p)=n\}$ is open for each integer n, and w vanishes outside a compact set.

Since γ is rectifiable, range γ has Lebesgue measure zero. So w is defined a.e. and is measurable. Rectifiability of γ also implies that the component functions of γ, $t\mapsto x(t)$ and $t\mapsto y(t)$, are of bounded variation and define Riemann-Stieltjes measures on the interval $[a, b]$. In particular, the line integral in (2) is well-defined.

In case γ is polygonal, γ will be represented by a string $p_0p_1\cdots p_n$, where the points $p_0, p_1, \ldots, p_n$ are the vertices of γ, $p_i\neq p_{i+1}$ for $1\leq i+1\leq n$, and n is called the number of vertices of γ. Range γ is the union of the closed line segments $[p_i, p_{i+1}]$, with successive line segments being homeomorphic images of successive closed subintervals of $[a, b]$. Since the exact choice of homeomorphisms does not affect w or the line integral in (2), this choice is left unspecified. Because γ is closed, $p_n=p_0$ and $n\geq 2$.

LEMMA. *Let γ be polygonal. Then* (1) *and* (2) *are satisfied.*

Proof. Let $\gamma=p_0p_1\cdots p_n$. The proof is an induction on n. If $n=2$, range γ is collinear. Hence, $w=w^2=0$ a.e., w and w^2 are integrable, and the left sides of (1) and (2) vanish. Since the right side of (2) is invariant under Euclidean motions, we can assume $y\equiv 0\equiv dy$ along γ, and thus the right side of (2) also vanishes. Supposing that the lemma is valid for $n=2,\ldots,k$, we now consider the case when $n=k+1\geq 3$.

If γ is a simple closed curve, the theory of Jordan curves [1, p. 64] shows that w is identically equal to 1 or -1 inside γ and identically equal to 0 outside γ. Then (1) is just the classical isoperimetric inequality (proved in Courant-Hilbert [4], for example), and this inequality shows incidentally that w^2 and w are integrable. The identity (2), on the other hand, is a special case of Green's Theorem [2, p. 289] or can be proved directly by decomposing the polygonal area bounded by γ into a sum of signed triangular areas.

If γ is not simple, let q be a point where γ intersects itself. Then γ can be decomposed into two closed polygonal subarcs γ_1 and γ_2, complementary portions of γ each having q as initial and terminal point. Apart from q itself, the vertices of γ_1 and γ_2 are consecutive members of the cyclically ordered list $p_1,\ldots,p_{k+1}$. If n_i is the number of vertices of γ_i for $i=1, 2$, then $2\leq n_i\leq k+1$: otherwise, one of the polygonal arcs γ_1 or γ_2 has q as its only vertex, and this is impossible.

Suppose γ_1 or γ_2—say γ_1—has exactly $n = k+1$ vertices. Then γ_2 has two vertices—q and an original vertex p_i, q is distinct from p_i, and q lies in $[p_{i-1}, p_i) \cap (p_i, p_{i+1}]$ (Here $1 \le i \le n$, and if $i = n$, "$i+1$" should be replaced by "1".) If two line segments have an endpoint p_i and one other point q in common, one segment must be a subset of the other. Then new subarcs γ_1 and γ_2 can be defined as follows: (i) if $[p_{i-1}, p_i)$ is a proper subset of $(p_i, p_{i+1}]$, let $\gamma_1 = p_{i-1}p_{i+1} \cdots p_n p_1 \cdots p_{i-1}$ and $\gamma_2 = p_{i-1}p_ip_{i-1}$; (ii) if $(p_i, p_{i+1}]$ is a proper subset of $[p_{i-1}, p_i)$, let $\gamma_1 = p_{i+1} \cdots p_n p_1 \cdots p_{i-1}p_{i+1}$ and $\gamma_2 = p_{i+1}p_ip_{i+1}$; and (iii) if $p_{i-1} = p_{i+1}$, let $\gamma_1 = p_{i+1} \cdots p_n p_1 \cdots p_{i-1}$ and $\gamma_2 = p_{i-1}p_ip_{i+1}$. The new subarcs are closed polygonal curves decomposing γ, and they satisfy $n_1 = n-1 = k$ and $n_2 = 2$.

The above demonstrates that if γ is not simple, γ can be divided into two closed polygonal curves γ_1 and γ_2 whose numbers of vertices n_1 and n_2 satisfy $2 \le n_i \le k$ for $i = 1, 2$. By the inductive hypothesis γ_1 and γ_2, and their winding number functions w_1 and w_2, satisfy (1) and (2). Since γ differs only trivially from the curve $\gamma_1\gamma_2$ consisting of γ_1 followed by γ_2, γ and $\gamma_1\gamma_2$ have the same winding number function and the same length and the line integrals of $x\,dy - y\,dx$ over γ and $\gamma_1\gamma_2$ are the same. It follows that $w = w_1 + w_2$ a.e., that w inherits integrability and square-integrability from w_1 and w_2, and that if $\| \, \|$ denotes the L_2 norm for functions defined on R^2, then

$$\|w\| = \|w_1 + w_2\| \le \|w_1\| + \|w_2\| \le \frac{L(\gamma_1)}{\sqrt{4\pi}} + \frac{L(\gamma_2)}{\sqrt{4\pi}} = \frac{L(\gamma)}{\sqrt{4\pi}}.$$

Squaring the first and last terms, we obtain (1). Likewise,

$$\int_{R^2} w\,dA = \int_{R^2} (w_1 + w_2)\,dA = \frac{1}{2}\left(\int_{\gamma_1} + \int_{\gamma_2}\right)(x\,dy - y\,dx) = \frac{1}{2}\int_{\gamma} (x\,dy - y\,dx),$$

and (2) is established. ∎

Proof of the theorem. For simplicity we reparametrize γ by arc length s so that s is mapped to $p(s)$ for $0 \le s \le L = L(\gamma)$. Such a reparametrization does not affect w or the line integral in (2).

Let n be an integer ≥ 2, and let $s_j = jL/n$ for $0 \le j \le n$. Choose numbers $t_0, t_1, \ldots, t_{n+1}$ satisfying the following conditions: (i) $t_0 = s_0 = 0$; (ii) $t_{n+1} = s_n = L$; (iii) $s_{j-1} \le t_j \le s_j$ for $1 \le j \le n$; and (iv) $p(t_j) \ne p(t_{j+1})$ for $0 \le j \le n$. Condition (iv) can be satisfied since the set $\{p(t) : s_{j-1} \le t \le s_j\}$ consists of infinitely many points for each $j = 1, \ldots, n$.

Let γ_n be the closed polygonal curve $p(t_0)p(t_1) \cdots p(t_{n+1})$, parametrized in the following way. For $0 \le t \le L$, γ_n takes t to $q_n(t)$ so that: (i) $q_n(t) = p(t)$ for $t = t_0, \ldots, t_{n+1}$; and (ii) $t \mapsto q_n(t)$ is a homeomorphism of $[t_j, t_{j+1}]$ onto $[p(t_j), p(t_{j+1})]$ for $0 \le j \le n$.

Next we define a homotopy $h_n:[0, L]\times[0, 1]\to R^2$ by $h_n(t, \lambda)=(1-\lambda)p(t)+\lambda q_n(t)$. This homotopy deforms the curve γ into the polygonal curve γ_n. Since γ and γ_n are homotopic closed curves, their winding number functions w and w_n agree on points outside range h_n (see [1, p. 48] or [7, p. 192]).

For $j=0,\ldots,n$, let $B(n, j)=\{p: d(p, p(s_j))\le L/n\}$. For s in $[t_j, t_{j+1}]$, $d(p(s), p(s_j))\le|s-s_j|$ since s is an arc length parameter; and $|s-s_j|\le \max\{t_{j+1}-s_j, s_j-t_j\}\le L/n$ by definition of the t_j's. Thus $p(s)$ is in $B(n, j)$. However, $B(n, j)$ is a convex set. Consequently, for t in $[t_j, t_{j+1}]$, the point $q_n(t)$, which lies on the line segment $[p(t_j), p(t_{j+1})]$, is in $B(n, j)$. Moreover, for t in $[t_j, t_{j+1}]$ and λ in $[0, 1]$, the point $h_n(t, \lambda)$, which lies in $[p(t), q_n(t)]$, is also in $B(n, j)$. We conclude that range h_n is a subset of $V_n=\bigsqcup_{j=0}^n B(n, j)=\bigsqcup_{j=1}^n B(n, j)$.

The Lebesgue measure of V_n is $\le n\pi(L/n)^2=\pi L^2/n$ since each set $B(n, j)$ is a disc of radius L/n. Thus $w(p)=w_n(p)$ for all points p outside a set of measure $\le\pi L^2/n$. So the sequence $\{w_n\}$ converges in measure to w. By a theorem of F. Riesz [6, p. 156], there exists a subsequence $\{w_{n(k)}\}$ which converges pointwise a.e. to w. (For notational convenience, we will use the subscript k instead of $n(k)$ henceforth.) The polygonal curves $\{\gamma_n\}$ satisfy (1), and $L(\gamma_n)\le L(\gamma)$. So by Fatou's lemma [6, p. 172],

$$\int_{R^2} w^2\,dA=\int_{R^2} \liminf w_k^2\,dA\le\liminf\int_{R^2} w_k^2\,dA$$
$$\le\liminf L(\gamma_k)^2/4\pi\le L(\gamma)^2/4\pi.$$

Thus (1) is proved, along with square-integrability of w.

The inequality (1) becomes an equality when γ is a circle of radius r traversed n times in the same sense: for then both sides of (1) equal πn^2r^2. To see that equality can occur in no other way, consider the class Γ of all closed rectifiable curves, parametrized by arc length, which reduce (1) to an equality. Define a function $\gamma\mapsto n(\gamma)$ for γ in Γ by $n(\gamma)=\min\{|w(p)|: p$ is not in range γ and $w(p)\ne0\}$, where w is the winding number function of γ. Since (1) is an equality, positivity of $L(\gamma)$ implies that w takes nonzero values and so $n(\gamma)$ is well-defined.

If γ is a member of Γ which is not a simple closed curve, γ can be decomposed into two successive closed subarcs γ_1 and γ_2 parametrized by arc length and having positive lengths. Since γ differs only trivially from the combined curve $\gamma_1\gamma_2$, $L(\gamma)=L(\gamma_1)+L(\gamma_2)$ and $w=w_1+w_2$ a.e., where w_1 and w_2 are winding number functions for γ_1 and γ_2. Applying the L_2 norm $\|\,\|$ and (1), we find that $\|w\|=\|w_1+w_2\|\le\|w_1\|+\|w_2\|\le(L(\gamma_1)+L(\gamma_2))/\sqrt{4\pi}=L(\gamma)/\sqrt{4\pi}$. the first and last terms above are equal by hypothesis. So $\|w_i\|=L(\gamma_i)/\sqrt{4\pi}$ for $i=1, 2$, and γ_1 and γ_2 are in Γ. Furthermore, equality in the triangle inequality implies that $w_1=\lambda uw$ and $w_2=(1-\lambda)uw$ where $u=\pm1$ and $0<\lambda<1$. Thus

$n(\gamma_1)=\lambda n(\gamma)$ and $n(\gamma_2)=(1-\lambda)n(\gamma)$, and both $n(\gamma_1)$ and $n(\gamma_2)$ are strictly smaller than $n(\gamma)$.

The decomposition described in the last paragraph can be repeated as long as one or more of the resulting subarcs fails to be simple. However, $\gamma\mapsto n(\gamma)$ is a function with positive integer values and decreases strictly at each decomposition. So only finitely many decompositions are possible before one arrives at simple closed curves. It follows that for a curve γ in Γ, there exist simple closed curves $\gamma_1,\ldots,\gamma_n$ in Γ such that $L(\gamma)=L(\gamma_1)+\cdots+L(\gamma_n)$, and such that $w=w_1+\cdots+w_n$ a.e., where $w_1,\ldots,w_n$, w are winding number functions of $\gamma_1,\ldots,\gamma_n$, γ.

Since $\gamma_1,\ldots,\gamma_n$ are simple closed curves, their winding number functions assume only two values—namely, 0 and ± 1. Hence, equality in (1) implies equality in the classical isoperimetric inequality; and so $\gamma_1,\ldots,\gamma_n$ must be circles. On the other hand, $\|w\|=\|w_1+\cdots+w_n\|\le\|w_1\|+\cdots+\|w_n\|=(L(\gamma_1)+\cdots+L(\gamma_n))/\sqrt{4\pi}=L(\gamma)/\sqrt{4\pi}=\|w\|$. Then by the Cauchy–Schwarz inequality, $w_j=\lambda_j uw$ a.e. for $u=\pm 1$ and for positive numbers $\lambda_1,\ldots,\lambda_n$ summing to 1. Since the winding number function of a circle is nonzero only inside the circle and since $w_1,\ldots,w_n$ vanish simultaneously, $\gamma_1,\ldots,\gamma_n$ must represent the same circle; since $w_1,\ldots,w_n$ have the same sign inside this circle, we further conclude that $\gamma_1,\ldots,\gamma_n$ represent the same circle traversed in the same sense.

To show that the initial curve γ is a circle traversed a finite number of times in the same sense, we reverse the decomposition process. It suffices to show that if γ_1 and γ_2 are obtained from γ by decomposition and if γ_1 and γ_2 are each circles traversed a finite number of times in the same sense—the circle and the sense being common to γ_1 and γ_2, then γ is likewise. But this is obvious. So the theorem is proved. ∎

COROLLARY 1. *Under the hypotheses of the theorem, let A_n be the Lebesgue measure of the set $\{p: w(p)$ exists and $|w(p)|\ge n\}$. Then*

$$\sum_{n=1}^{\infty} n^2(A_n-A_{n+1})\le L(\gamma)^2/4\pi$$

and

$$A_n\le L(\gamma)^2/4\pi n^2 \quad \text{for all } n\ge 1.$$

Proof. The first inequality is a restatement of (1). The second inequality also results from (1) when it is observed that the integral of w^2 over R^2 is greater than or equal to the integral of w^2 over $\{p:|w(p)|\ge n\}$, which in turn is greater than or equal to n^2A_n. ∎

COROLLARY 2. *Under the hypotheses of the theorem, the signed area identity* (2) *is satisfied.*

https://doi.org/10.4153/CMB-1981-026-8 Published online by Cambridge University Press

Proof. Let $\{\gamma_n\}$ be the sequence of closed polygonal curves described in the proof of the theorem. Then a straightforward computation shows that $(\frac{1}{2})\int_{\gamma_n}(x\,dy - y\,dx)$ is equal to

$$\left(\frac{1}{2}\right)\sum_{j=0}^{n}[x(t_j)(y(t_{j+1})-y(t_j))-y(t_j)(x(t_{j+1})-x(t_j))].$$

The last expression can be interpreted as a Riemann–Stieltjes sum for the line integral of $(\frac{1}{2})(x\,dy - y\,dx)$ along γ because the points $p(t_j)=(x(t_j), y(t_j))$ for $0 \le j \le n$ are successive points on both γ_n and γ. As n approaches ∞, this sum converges to the Riemann–Stieltjes integral $\frac{1}{2}\int_\gamma (x\,dy - y\,dx)$. Since (2) is valid for polygonal curves by the lemma, it follows that the winding number functions w_n of γ_n satisfy:

$$\lim_{n\to\infty}\int_{R^2} w_n\,dA = \frac{1}{2}\int_\gamma (x\,dy - y\,dx). \tag{2.1}$$

Recall from the proof of the theorem the subsequence $\{w_k\}$ which converges a.e. to w. Let M be a fixed positive integer, and let $g_{M,k}(p)=1$ if $|w_k(p)|<M$ and $|w(p)|<M$ and $g_{M,k}(p)=0$ otherwise. Then

$$\begin{aligned}\int_{\{p:|w_k(p)|\ge M\}} w_k\,dA &= \int_{R^2} w_k\,dA + \int_{R^2} g_{M,k}(w-w_k)\,dA - \int_{\{p:|w(p)|<M\}} w\,dA \\ &\quad - \int_{\{p:|w_k(p)|<M, |w(p)|\ge M\}} w_k\,dA + \int_{\{p:|w_k(p)|\ge M, |w(p)|<M\}} w\,dA.\end{aligned} \tag{2.2}$$

We analyze what happens to the terms on the right side of (2.2) as k approaches ∞. By (2.1) the first term converges to $(\frac{1}{2})\int_\gamma (x\,dy - y\,dx)$. Let C be a compact set outside of which all the winding number functions $\{w_k\}$ and w vanish (the manner of construction of the curves $\{\gamma_n\}$ shows that we may take C to be any closed disc containing range γ), and let $\chi_C(p)$ be the characteristic function of C. Then Lebesgue's Dominated Convergence Theorem [6, p. 172] can be applied to $g_{M,k}(w-w_k)$ because $|g_{M,k}(w-w_k)| \le 2M\chi_C$ and the latter is integrable. Thus, since $\lim_{k\to\infty} g_{M,k}(w-w_k)=0$ a.e., the second term on the right side of (2.2) converges to 0. The fourth term on the right side of (2.2) has absolute value $\le M$ times the measure of $\{p: w_k(p) \ne w(p)\} \le M\pi L^2/n(k)$ (see the proof of the theorem), and since $n(k)$ approaches ∞ as k does, this term approaches 0. In like manner, the fifth term on the right side of (2.2) approaches 0.

Thus

$$\lim_{k\to\infty}\int_{\{p:|w_k(p)|\ge M\}} w_k\,dA = \frac{1}{2}\int_\gamma (x\,dy - y\,dx) - \int_{\{p:|w(p)|<M\}} w\,dA,$$

or equivalently

$$(2.3)\quad \frac{1}{2}\int_\gamma (x\,dy - y\,dx)\int_{R^2} w\,dA = \lim_{k\to\infty} \int_{\{p:|w_k(p)|\geq M\}} w_k\,dA - \int_{\{p:|w(p)|\geq M\}} w\,dA.$$

But the integral of w_k over $\{p:|w_k(p)|\geq M\}$ has absolute value $\leq$ the integral of w_k^2/M over $R^2 \leq L(\gamma_k)^2/4\pi M \leq L(\gamma)^2/4\pi M$ for all k. Similarly, the integral of w over $\{p:|w(p)|\geq M\}$ has absolute value $\leq L(\gamma)^2/4\pi M$. So the left side of (2.3) has absolute value $\leq L(\gamma)^2/2\pi M$. Since the left side does not depend on M, M can be increased without bound to yield the identity (2). ∎

References

1. P. S. Aleksandrov, *Combinatorial Topology*, vol. 1, Graylock Press, Rochester, N.Y., 1956 (translation by Horace Komm).
2. T. M. Apostol, *Mathematical Analysis*, Addison-Wesley, Reading, Mass., 1957.
3. T. F. Banchoff and W. F. Pohl, *A generalization of the isoperimetric inequality*, J. Differential Geometry **6** (1971), 175–192.
4. R. Courant and D. Hilbert, *Methods of Mathematical Physics*, vol. 1, Interscience Publishers, Inc., New York, 1953.
5. H. Federer, *Geometric Measure Theory*, Springer-Verlag, Berlin, 1969.
6. E. Hewitt and K. Stromberg, *Real and Abstract Analysis*, Springer-Verlag, New York, 1965.
7. M. H. A. Newman, *Elements of the Topology of Plane Sets of Points*, University Press, Cambridge, England, 1964.
8. R. Osserman, *The isoperimetric inequality*, Bull. Amer. Math. Soc. 84 **6** (1978), 1182–1238.
9. T. Radó, *A lemma on the topological index*, Fund. Math. **27** (1936), 212–225.
10. T. Radó, *The isoperimetric inequality and the Lebesgue definition of surface area*, Trans. Amer. Math. Soc. **41** (1947), 530–555.

DEPARTMENT OF MATHEMATICS
GEORGETOWN UNIVERSITY
WASHINGTON, D.C. 20057
U.S.A.

https://doi.org/10.4153/CMB-1981-026-8 Published online by Cambridge University Press