Six proofs of the Fáry–Milnor theorem

Anton Petrunin and Stephan Stadler

Introduction

The following problem was posted by Karol Borsuk [4].

Show that the total curvature of any nontrivial knot is at least $4\cdot\pi$.

It is known for many proofs based on different ideas. We sketch several solutions, one solution per section; each can be read independently.

This problem also has a number of refinements and generalizations; in particular, a strict inequality holds — this is the famous *Fáry–Milnor theorem*. However, for the sake of simplicity, we stick to the original formulation.

In order to continue, we need to agree on a definition of knot and explain what a *nontrivial knot* is. Despite the intuitive idea of a knot as a *cyclic rope*, the simplest formal definition uses polygonal curves. It turns out that this definition is also best suited for our purposes.

A *knot* (more precisely *tame knot*) is a simple closed polygonal curve in the Euclidean space $\mathbb{R}^3$ (*simple* means *no self-intersections*).

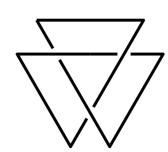
The solid triangle with vertices a, b, and c will be denoted by $\blacktriangle abc$. It is defined as the convex hull of the points a, b, and c; the points a, b and c are assumed to be distinct, but they might lie on one line.

We define a *triangular isotopy* of a knot to be the generation of a new knot from the original one by means of the following two operations:

Assume $[p,q]$ is an edge of the knot and x is a point such that the solid triangle $\blacktriangle pqx$ has no common points with the knot except for the edge $[p,q]$. Then we can replace the edge $[p,q]$ with the two adjacent edges $[p,x]$ and $[x,q]$.

We can also perform the inverse operation. That is, if for two adjacent edges $[p,x]$ and $[x,q]$ of a knot the triangle $\blacktriangle pqx$ has no common points with the knot except for the points on the edges $[p,x]$ and $[x,q]$, then we can replace $[p,x]$ and $[x,q]$ by one edge $[p,q]$.

Polygons that arise from one another by a finite sequence of triangular isotopies are called *isotopic*. A knot that is not isotopic to a triangle (that is, a simple polygonal curve with three vertices) is called *nontrivial*.



The trefoil knot shown on the diagram to the right gives a simple example of a nontrivial knot. A proof that it is nontrivial can be found in any textbook on knot theory. The most elementary and visual proof is based on the so-called *tricolorability* of knot diagrams [1, Section 1.5].

The total curvature of a smooth curve is usually defined as the integral of its curvature. For polygons, it is defined as the sum of its external angles. It is well known that the total curvature of a curve cannot be smaller than the total curvature of an inscribed polygonal curve (see, for example, [14]). In fact, the total curvature of a curve can be defined as the least upper bound on the total curvature of inscribed polygonal curves [2, 17]. This definition agrees with the definition given for smooth curves, and it makes sense for any simple curve. The total curvature of a curve α will be denoted by $\Phi(\alpha)$.

It follows that the problem allows the following reformulation which we are going to prove.

Main theorem. $\Phi(\alpha) \geqslant 4{\cdot}\pi$ *for any nontrivial knot* α.

1 Milnor–Fenner

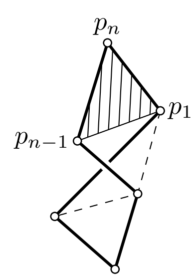
One of the first solutions to the problem was found by John Milnor [12]. In this section, we present an amusing interpretation of his proof found by Stephen Fenner [10]. Just like the original version, it is based on the following sufficient condition for the triviality of a knot.

1.1. Proposition. *Assume that a height function* $(x, y, z) \to z$ *has only one local maximum on a closed simple polygonal curve* α *and all the vertices of the polygonal curve are at different heights. Then* α *is a trivial knot.*

The proof is a straightforward construction of a triangular isotopy.

Proof. Let $\alpha = p_1 \dots p_n$ be the closed simple polygonal curve such that the height function $(x, y, z) \to z$ has one local maximum and all vertices have different heights. Note that the height function also has a unique local minimum, and α can be divided into two arcs from the min-vertex to the max-vertex with a monotonic height function.

Consider the three vertexes with the largest height; they have to include the max-vertex and two more. Note that these three vertices are consequent in the polygonal curve; without loss of generality, we can assume that they are p_{n-1}, p_n, and p_1.



Note that the solid triangle $\blacktriangle p_{n-1}p_np_1$ does not intersect any edge α except the two adjacent edges $[p_{n-1}, p_n] \cup [p_n, p_1]$. Indeed, if $\blacktriangle p_{n-1}p_np_1$ intersects $[p_1, p_2]$, then, since p_2 lies below $\blacktriangle p_{n-1}p_np_1$, the edge $[p_1, p_2]$ must intersect $[p_{n-1}, p_n]$; the latter is impossible since α is simple.

The same way, one can show that $\blacktriangle p_{n-1}p_np_1$ can not intersect $[p_{n-2}, p_{n-1}]$. The remaining edges lie below $\blacktriangle p_{n-1}p_np_1$, hence they cannot intersect this triangle.

Applying a triangular isotopy, to $\blacktriangle p_{n-1}p_np_1$ we get a closed simple polygonal curve $\alpha' = p_1 \dots p_{n-1}$ which is isotopic to α.

Since all the vertices p_i have different heights, the assumption of the proposition holds for α'.

Repeating this procedure $n - 3$ times we get a triangle. Hence α is a trivial knot. $\square$

Proof of the main theorem. Let $\alpha = p_1 \dots p_n$ be a nontrivial polygonal knot. Denote by v_i the unit vector in the direction of $p_{i+1} - p_i$; we assume that $p_n = p_0$. Consider the set U_i formed by all unit vectors u such that $\measuredangle(u, v_i) \geqslant \frac{\pi}{2}$ and $\measuredangle(u, v_{i-1}) \leqslant \frac{\pi}{2}$. Note that $u \in U_i$ if and only if the function $x \mapsto \langle u, x\rangle$ has a local maximum at p_i on α.

Let us choose (x, y, z)-coordinates in the space so that z-axis points in the direction of u. Then according to 1.1, the function $p \mapsto \langle u, p\rangle$ has at least two local maxima on α. It follows that the sets $U_1, \dots, U_n$ cover each point on the unit sphere $\mathbb{S}^2$ twice.

Recall that area $\mathbb{S}^2 = 4{\cdot}\pi$. Observe that $\varphi_i = \measuredangle(v_{i-1}, v_i)$ is the external angle of α at p_i. Note that U_i is a slice of the sphere between two meridians meeting at angle φ_i, therefore U_i occupies a $\frac{\varphi_i}{2\cdot\pi}$ portion of the whole sphere; so, area $U_i = 2{\cdot}\varphi_i$. Since the sets $U_1, \dots, U_n$ cover $\mathbb{S}^2$ twice, we get

$$\Phi(\alpha) = \varphi_1 + \dots + \varphi_n = \tfrac{1}{2}{\cdot}(\operatorname{area} U_1 + \dots + \operatorname{area} U_n) \geqslant \tfrac{2}{2}{\cdot} \operatorname{area} \mathbb{S}^2 = 4{\cdot}\pi. \qquad \square$$

2 Fáry

In this section, we sketch the solution of István Fáry [9] which was published before Milnor's proof.

We start with Crofton-type formulas for total curvature. Given a curve α in $\mathbb{R}^3$ and a unit vector u, denote by $\alpha_{u^\perp}$ and α_u the projections of α to the plane perpendicular to u and the line parallel to u respectively. Given a function $f\colon \mathbb{S}^2 \to \mathbb{R}$, let us denote its average value by $\overline{f(u)}$.

2.1. Crofton-type formula. *Let α be a polygonal curve in $\mathbb{R}^3$. Then*

$$\Phi(\alpha) = \overline{\Phi(\alpha_{u^\perp})} = \overline{\Phi(\alpha_u)}.$$

Proof. Observe that it is sufficient to check the identities for polygonal curves made of two edges and in this case, it boils down to very straightforward calculations. $\square$

The original version of Milnor's proof used the identity $\Phi(\alpha) = \overline{\Phi(\alpha_u)}$; in Fenner's version of his proof it was hidden under the carpet.

Fáry's proof is based on the identity $\Phi(\alpha) = \overline{\Phi(\alpha_{u^\perp})}$ and the following inequality for total curvature. Suppose $\alpha = p_1 \dots p_n$ is a simple closed polygonal curve in $\mathbb{R}^3$ and $o \notin \alpha$. Let us define the *angular length* of α with respect to o as the sum

$$\Psi_o(\alpha) = \measuredangle p_1 o p_2 + \dots + \measuredangle p_{n-1} o p_n + \measuredangle p_n o p_1.$$

2.2. Proposition. *For any closed simple polygonal curve and any $o \notin \alpha$, we have*

$$\Psi_o(\alpha) \leqslant \Phi(\alpha).$$

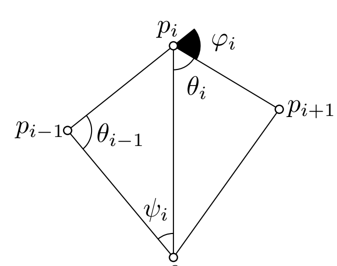
Proof. Let $\alpha = p_1 \dots p_n$; for each i, set

$$\varphi_i = \pi - \measuredangle p_{i-1} p_i p_{i+1}, \qquad \psi_i = \measuredangle p_{i-1} o p_i, \qquad \theta_i = \measuredangle o p_i p_{i+1}.$$

Here we assume that indexes are taken modulo n; in particular, $p_n = p_0$.

Note that φ_i is the external angle at p_i; therefore

$$\Phi(\alpha) = \varphi_1 + \dots + \varphi_n$$



The directions of $p_i - p_{i-1}$, $o - p_i$, and $p_{i+1} - p_i$ make angles $\psi_i + \theta_{i-1}$, θ_i, and φ_i to each other. Applying the triangle inequality for these angles, we get

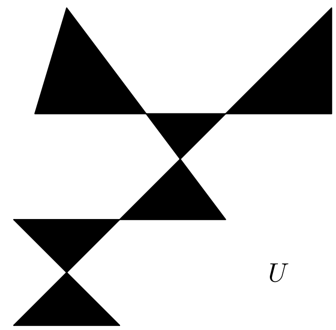
$$\varphi_i \geqslant \psi_i + \theta_{i-1} - \theta_i.$$

Summing up, we get

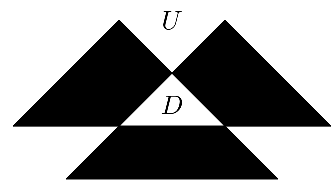
$$\varphi_1 + \dots + \varphi_n \geqslant \psi_1 + \dots + \psi_n,$$

and the result follows. $\square$

Proof of the main theorem. Consider a projection of the knot to a plane in *general position* (this time it means that the self-intersections of the projection are at most double and the projection of each edge is not degenerate). The obtained closed polygonal curve $\alpha_{u^\perp} = p_1 p_2 \dots p_n$ divides the plane into domains, one of which is unbounded, denote it by U, and the others are bounded.



First, note that all domains can be colored in a chessboard order; that is, they can be colored in black and white in such a way that domains with common borderline get different colors. If the unbounded domain is colored in white and every other domain is colored in black, then one can untie the knot by flipping these domains one by one.[1]



Therefore, among the bounded domains there is a white domain, denote it by D. The domain D cannot adjoin U, since they have the same color. Fix a point o in this domain.

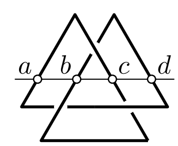
Since any ray from o crosses $\alpha_{u^\perp}$ twice, we get $\Psi_o(\alpha_{u^\perp}) \geqslant 4\cdot\pi$; that is, the angular length of $\alpha_{u^\perp}$ with respect to o is at least $4\cdot\pi$. By 2.2, we have

$$\Phi(\alpha_{u^\perp}) \geqslant 4\cdot\pi.$$

This is true for any u in general position. The remaining directions contribute nothing to the average value. It remains to apply the Crofton-type formula $\Phi(\alpha) = \overline{\Phi(\alpha_{u^\perp})}$. $\square$

3 Pannwitz–Hopf–Schmitz–Denne

Fáry's paper [9] is ended by the following note: *I just received a letter from Mr Borsuk, which says that Theorem 3 has been proved independently by Mr H. Hopf. It uses the theorem of Miss Pannwitz, which ensures that for any knot there is a line that crosses it at least in four points.* This proof is the subject of this section.



The main step in the proof is the existence of the so-called *alternated quadrisecant* of a knot α; that is, there is a line ℓ that shares with α four points a, b, c and d that appear on ℓ in the same order and have cyclic order a, c, b, d on α.

The question about existence of quadrisecant was asked by Otto Toeplitz and answered by Erika Pannwitz for tame knots in *general position* [13]. Elizabeth Denne had generalized this result to all knots [6, 7]. But for us tame knots in general position will be sufficient; namely, we need the following:

3.1. Proposition. *Any nontrivial knot α in general position admits an* alternated quadrisecant.

The precise meaning of general position will be clear from the proof; what is important is that any knot is arbitrary close to a knot in general position. Erika Pannwitz only proved the existence of quadrisecant, but the existence of alternated quadrisecant can be extracted from her proof.

Proof of the main theorem modulo 3.1. Let $\alpha = p_1 \dots p_n$ be a knot in general position. Suppose a, b, c and d be the points as in the definition of an alternated quadrisecant. Then $acbd$ is an inscribed quadrangle with all external angles equal to π. Therefore $\Phi(abcd) = 4\cdot\pi$.

[1] It is instructive to give a formal proof of the last statement; that is, *show that if there is only one white region, then α is trivial.*

Since the total curvature of an inscribed polygonal curve cannot be larger than the total curvature of the original curve, we get that $\Phi(\alpha) \geqslant 4{\cdot}\pi$.

Finally, for any knot $\beta = q_1 \dots q_n$ there is an arbitrarily close knot $\alpha = p_1 \dots p_n$ in general position; in particular, for any $\varepsilon > 0$ we can assume that $\Phi(\beta) > \Phi(\alpha) - \varepsilon$. It follows that $\Phi(\beta) > 4{\cdot}\pi - \varepsilon$ for any positive ε — hence the result. $\square$

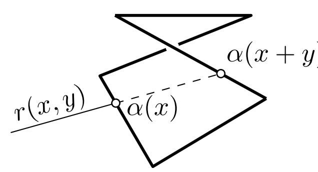
In the proof of the proposition, we will use the following characterization of the trivial knot.

3.2. Lemma. *A knot α is trivial if there exists a piecewise linear map F from the disc $\mathbb{D}$ to $\mathbb{R}^3$ such that the restriction of F to the boundary $\partial\mathbb{D} = \mathbb{S}^1$ is a degree-one map to α and F does not map interior points of $\mathbb{D}$ to α.*

This statement can be deduced from the so-called *Dehn's lemma* — a heavy weapon of 3-dimensional topology proved by Christos Papakyriakopoulos. For those who are familiar with Dehn's lemma, it would be an exercise; otherwise, we suggest taking it for granted. In the following proof, we follow closely the presentation of Carsten Schmitz [15] and the original argument of Erika Pannwitz.

Proof of the proposition. We may assume that α comes with a 1-periodic piecewise linear parametrization by $\mathbb{R}$; so the space of oriented chords of α can be identified with the open cylinder $\mathbb{S}^1 \times (0,1)$, where $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$. Namely, we assume that a pair $(x,y) \in \mathbb{S}^1 \times (0,1)$ corresponds to the oriented chord with the ends at $\alpha(x)$ and $\alpha(x+y)$.

Choose a pair $(x,y) \in \mathbb{S}^1 \times (0,1)$. Let us denote by $r(x,y)$ the ray that starts at $\alpha(x)$ and goes in the direction opposite to $\alpha(x+y)$. We write $(x,y) \in C_3$ if $r(x,y)$ crosses α at another point.



The points $\alpha(x)$ and $\alpha(x+y)$ divide α into two open arcs $\alpha|_{(x,x+y)}$ and $\alpha|_{(x+y,x+1)}$. If $(x,y) \in C_3$ and $r(x,y)$ crossed the first arc, then we write $(x,y) \in C_3^+$; if it crosses the second arc, then $(x,y) \in C_3^-$. Note that $C_3^+ \cup C_3^- = C_3$.

Observe that if C_3^+ and C_3^- intersect, then the proposition follows.

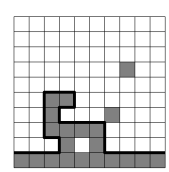
The set $C_3^\pm$ might be not closed; denote by $\bar{C}_3^\pm$ its closure. Suppose

$$(x,y) \in (\bar{C}_3^+ \cup \bar{C}_3^-) \backslash (C_3^+ \cup C_3^-).$$

Observe that $\alpha(x)$ is a vertex, say p_i, of α. Further, the line containing p_i and $\alpha(x+y)$ lies in the plane spanned by p_{i-1}, p_i, and p_{i+1}; and it intersects yet another point of α. The latter is not possible if α in general position. That is, since α is in general position, $C_3^+ \cap C_3^- = \varnothing$ implies that $\bar{C}_3^+ \cap \bar{C}_3^- = \varnothing$.

Note that there is $\varepsilon > 0$ such that for any $x \in \mathbb{S}^1$ we have the following: if $y < \varepsilon$, then $(x,y) \notin \bar{C}_3^+$, and if $y > 1 - \varepsilon$, then $(x,y) \notin \bar{C}_3^-$.

Arguing by contradiction, assume $\bar{C}_3^+ \cap \bar{C}_3^- = \varnothing$. Note that in this case there is a large n such that any $\frac{1}{n} \times \frac{1}{n}$-square in $\mathbb{S}^1 \times (0,1)$ does not intersect both $\bar{C}_3^+$ and $\bar{C}_3^-$. Let us cut $\mathbb{S}^1 \times (0,1)$ into n^2 such squares; each square defines a closed subset of $\mathbb{S}^1 \times (0,1)$. Color the union of squares that intersect $\bar{C}_3^+$ or lie in the lowest row of squares for $y \leqslant \frac{1}{n}$. Note that we did not color squares in the upper row and the boundary of the colored set contains a simple curve $t \mapsto (x(t), y(t))$ that cuts the cylinder $\mathbb{S}^1 \times (0,1)$ into two cylinders.



Note that $(x(t), y(t)) \notin \bar{C}_3$ for any $t \in \mathbb{S}^1$ and the curve $t \mapsto (x(t), y(t))$ runs along coordinate lines. Consider the one-parameter family of line segments in $r(x(t), y(t))$ that start at $\alpha(x(t))$ and end on the surface of a large cube that contains α in its interior. This way we obtain a piecewise linear annulus that connects the curve $t \mapsto \alpha(x(t))$ to a curve on the surface of the cube. The latter curve can be contracted by a piecewise linear disc in the surface of the cube. It might have self-intersection, but it cannot contains points of α.

Observe that $t \mapsto \alpha(x(t))$ defines a degree-one map $\mathbb{S}^1 \to \alpha$. Applying the lemma, we get the result. $\square$

4 Alexander–Bishop

Here we sketch the proof given by Stephanie Alexander and Richard Bishop [3]. This proof was designed to work for more general ambient spaces. As a result, it is more elementary.

In the proof, we construct a total-curvature-decreasing deformation of a given knot into a doubly covered bigon. The statement follows since the latter has total curvature $4\cdot\pi$.

Proof of the main theorem. Let $\alpha = p_1 \dots p_n$ be a nontrivial knot; that is, one can not get a triangle from α by applying a sequence of triangular isotopies defined in the previous section.

If $n = 3$ the polygonal curve α is a triangle. Therefore, by definition, α is a trivial knot — nothing to show.

Consider the smallest n for which the statement fails; that is, there is a nontrivial knot $\alpha = p_1 \dots p_n$ such that

$$\Phi(\alpha) < 4\cdot\pi. \quad ❶$$

We use the indexes modulo n; that is, $p_0 = p_n$, $p_1 = p_{n+1}$ and so on. Without loss of generality, we may assume that α is in *general position*; this time it means that no four vertexes of α lie on one plane.

Set $\alpha_0 = \alpha$. If the solid triangle $\blacktriangle p_0p_1p_2$ intersects α_0 only in the two adjacent edges, then applying the corresponding triangular isotopy, we get a knot α_0' with $n-1$ edges that is inscribed in α_0. Therefore

$$\Phi(\alpha)_0 \geqslant \Phi(\alpha)_0'.$$

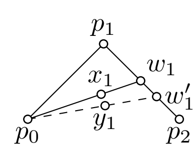
On the other hand, by the induction hypothesis

$$\Phi(\alpha_0') \geqslant 4\cdot\pi,$$

which contradicts ❶.

Choose the first point w_1' on the edge $[p_1, p_2]$ so that the line segment $[p_0, w_1']$ intersects α_0. Denote a point of intersection by y_1.

Choose a point w_1 on $[p_1, p_2]$ a bit before w_1'. Denote by x_1 the point on $[p_0, w_1]$ that minimizes the distance to y_1. This way we get a closed polygonal curve $\alpha_1 = w_1p_2 \dots p_n$ with two marked points x_1 and y_1. Denote by m_1 the number of edges in the arc $x_1w_1 \dots y_1$ of α_1.



Note that α_1 is isotopic to α_0; in particular, α_1 is a nontrivial knot.

Now let us repeat the procedure for the adjacent edges $[w_1, p_2]$ and $[p_2, p_3]$ of α_1. If the solid triangle $\blacktriangle w_1p_2p_3$ intersects α_1 only at these two adjacent edges, then we get a contradiction the same way as before. Otherwise, we get a new knot $\alpha_2 = w_1w_2p_3 \dots p_n$ with two marked points x_2 and y_2. Denote by m_2 the number of edges in the polygonal curve $x_2w_2 \dots y_2$.

Note that the points x_1, x_2, y_1, y_2 cannot appear on α_2 in the same cyclic order; otherwise the polygonal curve $x_1x_2y_1y_2$ can be made to be arbitrary close to a doubly covered bigon which again contradicts ❶.

Therefore we can assume that the arc $x_2w_2 \dots y_2$ lies inside the arc $x_1w_1 \dots y_1$ in α_2 and therefore $m_1 > m_2$.

Continuing this procedure we get a sequence of polygonal curves $\alpha_i = w_1 \dots w_i p_{i+1} \dots p_n$ with marked points x_i and y_i such that the number of edges m_i from x_i to y_i decreases as i increases. Clearly $m_i > 1$ for any i and $m_1 < n$. Therefore it requires less than n steps to arrive at a contradiction. □

5 Ekholm–White–Wienholtz

In this section, we discuss a solution of the problem based on a theorem of Tobias Ekholm, Brian White, and Daniel Wienholtz [8]. This theorem was a breakthrough in minimal surface theory at the time. Yet, it was based on an elementary idea that we are going to explain now.

We start with a polygonal curve α with total curvature less than $4{\cdot}\pi$; show that an area-minimizing disc spanned by α has no self-intersections, and therefore α has to be a trivial knot. So in a way the equation for area-minimizing surfaces solves our problem, we only need to understand it.

The main hero in this proof is the so-called *extended monotonicity theorem.* We will also apply the *Douglas–Rado theorem* on the existence of area-minimizing discs and reuse the inequality between total curvature and angular length from Fáry's proof; see 2.2.

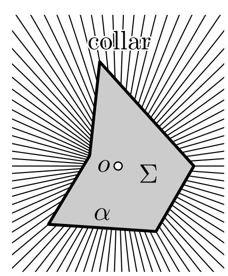
The image of a map from a domain of $\mathbb{R}^2$ to $\mathbb{R}^3$ will be called a *surface*; it might have self-intersections and singularities, but we assume it is reasonable, say locally Lipschitz; so we can talk about its area. A point on the surface might refer to a point in $\mathbb{R}^3$, or the corresponding point in the domain of parameters in $\mathbb{R}^2$; it should be easy to guess from the context.

We denote by $\mathbb{D}$ the closed disc in the plane. A surface defined by a map $f\colon \mathbb{D} \to \mathbb{R}^3$ will be called a *disc*. The restriction $f|_{\partial\mathbb{D}}$ will be called the *boundary line* of the disc.

A disc Σ is called area-minimizing if it has the smallest area among the surfaces with the given boundary line. The following statement about area-minimizing discs is easy to believe, but not easy to prove; see [19].

5.1. Douglas–Rado theorem. *Given a simple closed polygonal curve α in $\mathbb{R}^3$, there is an area-minimizing disc Σ with boundary line α; it is a smooth surface, possibly with self-intersections and isolated singularities.*

Moreover, if Σ has no self-intersections, then it is an embedded smooth surface with no singularities (in this case α is a trivial knot).



Choose a disc Σ in $\mathbb{R}^3$ with boundary line α. Given a point $o \notin \alpha$, let us consider *collared* Σ with respect to o; it is a new surface that will be denoted by $\hat\Sigma_o$; it includes Σ and the *collar* formed by all rays that start at points of α and go in the direction opposite to o. Note that $\hat\Sigma_o$ admits a natural parametrization by the whole plane.

5.2. Extended monotonicity theorem. *Let Σ be an area-minimizing disc with boundary line α. Given a point $o \notin \alpha$, consider the function*

$$W_o(r) = \operatorname{area}(\hat\Sigma_o \cap \bar B_r(o)),$$

where $\bar{B}_r(o)$ denotes the ball of radius r centered at o. Then the function $r \mapsto \frac{W_o(r)}{r^2}$ is nondecreasing. Moreover,

(a) $\lim_{r\to\infty} \frac{W_o(r)}{r^2} = \frac{1}{2}\cdot\Psi_o(\alpha)$, *where $\Psi_o(\alpha)$ denotes the angular length of α with respect to o; see Section 2.*

(b) *If $o \in \Sigma$, then* $\lim_{r\to 0} \frac{W_o(r)}{r^2} \geqslant \pi$.

The classical monotonicity theorem states that the same holds for balls that do not touch the boundary of the surface. The stated version is due to Brian White [18]. The same statement holds for minimal surfaces [8]; its proof requires deeper diving into differential geometry. At the same time, the original formulation admits a generalization to a wider class of ambient spaces [16].

Proof. Denote by $\lambda_o(r)$ the curve of intersection of the sphere $\partial B_r(o)$ with $\hat{\Sigma}_o$; set $\ell(r) = \mathrm{length}[\lambda_o(r)]$. Observe that

$$W_o'(r) \geqslant \ell(r)$$

for almost all r. (Formally speaking, this inequality follows from the so-called *coarea formula*.)

Set $\Delta_r = \hat{\Sigma}_o \cap \bar{B}_r(o)$; it is a surface bounded by $\lambda_o(r)$. Let Δ_r' be the cone over $\lambda_o(r)$ with the center at o. Note that Δ_r' differs from Δ_r only inside Σ. Since Σ is area-minimizing, we get that

❶ $$\operatorname{area}\Delta_r' \geqslant \operatorname{area}\Delta_r$$

for any $r > 0$. Observe that

$$\operatorname{area}\Delta_r = W_o(r), \qquad \operatorname{area}\Delta_r' = \tfrac{1}{2}\cdot r\cdot\ell(r).$$

Applying ❶, we get

$$r\cdot\ell(r) \geqslant 2\cdot W_o(r).$$

Therefore, we get

$$r\cdot W_o'(r) \geqslant 2\cdot W_o(r)$$

for almost all r. If W_o is smooth, then this inequality implies the main statement. In general, the argument shows that the absolutely continuous part of $W_o(r)/r^2$ is nondecreasing which suffices.

(a). Observe that up to a fixed error we have that $W_o(r)$ is the area of the ball of radius r in the cone over α with the tip at o. It follows that $W_o(r)/r^2$ approaches the area of the unit ball in this cone as $r \to \infty$ — hence the result.

(b). The statement is evident for smooth points of Σ. Since smooth points are dense in Σ, and $o \mapsto W_o(r)$ is a continuous function, the main part of the theorem implies that $W_o(r) \geqslant \pi\cdot r^2$ for *any* point $o \in \Sigma$ — hence the result. □

Proof of the main theorem. Suppose that $\Phi(\alpha) < 4\cdot\pi$. Consider an area-minimizing surface Σ with the boundary line α; it exists by 5.1. If Σ has no self-intersections, then α is a trivial knot.

Suppose Σ has a self-intersection at a point o. In this case, the intersection $B_r(o) \cap \Sigma$ is covered by two or more small area-minimizing subdiscs of Σ. By 5.2b, we get

$$\lim_{r\to 0} \frac{W_o(r)}{r^2} \geqslant 2\cdot\pi.$$

Applying the main statement in the monotonicity theorem, we get $\frac{W_o(r)}{r^2} \geqslant 2\cdot\pi$ for any $r > 0$. By 5.2b and Proposition 2.2 in Fáry's proof, we get

$$\Phi(\alpha) \geqslant \Psi_o(\alpha) \geqslant 2\cdot\frac{W_o(r)}{r^2} \geqslant 4\cdot\pi$$

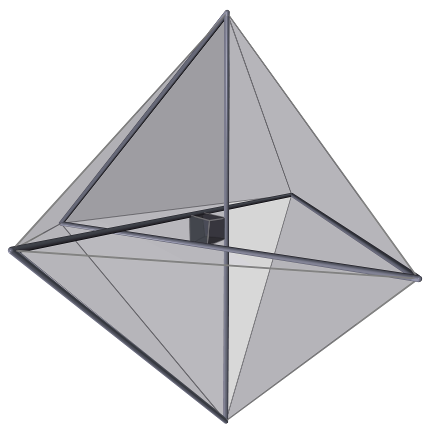
— a contradiction. □

6 Cantarella–Kuperberg–Kusner–Sullivan

The following proof is due to Jason Cantarella, Greg Kuperberg, Robert Kusner, and John Sullivan [5]; this is the only proof in our collection that uses knot theory a bit beyond the basic definitions.

Choose a polygonal curve $\alpha = p_1 \dots p_n$. Suppose a plane Π is in general position; that is, it does not contain vertices of α. Let us denote by $\operatorname{cross}_\alpha \Pi$ the number of intersections of Π and α. Extend the function $\Pi \mapsto \operatorname{cross}_\alpha(\Pi)$ to the minimal upper semicontinuous function defined for all planes; in other words, $\operatorname{cross}_\alpha(\Pi)$ is the maximal integer k such that there is a plane Π' arbitrary close to Π that intersects α at k points. The number $\operatorname{cross}_\alpha(\Pi)$ will be called the *crossing number* of Π. Note that for any plane Π, the crossing number is even, and it cannot exceed n.

It is easy to see that the *convex hull* $h_1(\alpha)$ of α can be defined the following way: $x \in h_1(\alpha)$ *if* $\operatorname{cross}_\alpha(\Pi) \geqslant 2$ *for any plane* $\Pi \ni x$. This observation suggests the following definition of the *second hull*: $x \in h_2(\alpha)$ *if* $\operatorname{cross}_\alpha(\Pi) \geqslant 4$ *for any plane* $\Pi \ni x$.



First and second hull of a trefoil.

6.1. Theorem. *The second hull of any nontrivial knot α has a nonempty interior.*

To show that this theorem implies our main theorem, we will apply the spherical Crofton formula: *for any spherical polygonal curve γ we have*

$$(*) \qquad \operatorname{length}\gamma = \pi\cdot\overline{n},$$

where $\overline{n}$ denotes the average number of intersections of γ with equators. To prove this formula, check it for an arc on an equator and sum it up for all edges of γ.

Proof of the main theorem modulo 6.1. Choose a point $o \in h_2(\alpha)$; we can assume that $o \notin \alpha$. Consider the radial projection α^* of α to the sphere centered at o; observe that

$$\operatorname{length}\alpha^* = \Psi_o(\alpha),$$

where $\Psi_o(\alpha)$ denotes the angular length of α with respect to o; see Section 2.

By 2.2, it is sufficient to show that

$$(**) \qquad \operatorname{length}\alpha^* \geqslant 4\cdot\pi.$$

Since o is in the second hull, α^* crosses every equator in general position at least 4 times. It follows that the average number of crossings is at least 4. Applying $(*)$, we get $(**)$. □

Suppose that a plane Π divides a knot α into two arcs, one on each side; in particular, Π intersects α at two points, say p and q. Then we can create two knots α_1 and α_2 by joining the ends of the two arcs by the line segment $[p, q]$. In this case, we say that α is a *connected sum* of α_1 and α_2.

6.2. Claim. *Suppose that a knot α is a connected sum of knots α_1 and α_2. If at least one of the knots α_1 or α_2 is nontrivial, then so is α.*

This claim has an amusing proof via the so-called *infinite swindle* [11].

Suppose β is a polygonal curve *inscribed* in α; that is, the vertices of β lie on α and they appear in the same cyclic order on α and β. If a plane in general position intersects an edge of β, then it intersects the corresponding arc of α. Therefore we get the following:

6.3. Observation. *If a polygonal line β is inscribed in α. Then $h_2(\beta) \subset h_2(\alpha)$.*

Proof of 6.1. Assume the contrary; let α be a nontrivial polygonal knot with the least number of vertices, say n, such that $h_2(\alpha)$ has an empty interior. It is easy to see that $n \geqslant 6$; in other words, any simple space 5-gon is a trivial knot.

Suppose Π is a plane in general position that divides α in two arcs, so it defines a decomposition of α into a connected sum of two knots α_1 and α_2. By the observation, $h_2(\alpha_1)$ and $h_2(\alpha_2)$ have empty interior. It follows that one of these knots, say α_1, is trivial and therefore the other, respectively α_2, is isotopic to α. Indeed, if n_1 and n_2 denote the number of vertices in α_1 and α_2, then $n_1 + n_2 = n + 4$. If both knots are nontrivial, then $n_1 \geqslant 6$ and $n_2 \geqslant 6$. Therefore $n_1 < n$ and $n_2 < n$ which contradicts minimality of n.

The open half-space H bounded by Π and containing $\alpha_2 \setminus \Pi$ will be called *essential*. Intuitively, an essential half-space cuts off a trivial knot of α. (A half-space containing all of α will be considered essential as well.) A rather straightforward application of 6.2 implies that if H' is another essential half-space for α, then it is also essential for α_2. It follows that the intersection of all essential half-spaces, say W, has nonempty interior — roughly speaking, it has to contain the region where knotting of α takes place.

Finally observe that if $\operatorname{cross}_\alpha \Pi = 2$ for a plane Π in general position, then Π bounds an essential half-space. It follows that if a plane in general position intersects W, then it has crossing number at least 4, so $h_2(\alpha) \supset W$ — hence the result. □

References

[1] C. Adams. *The knot book.* An elementary introduction to the mathematical theory of knots. 1994.

[2] A. D. Aleksandrov, Yu. G. Reshetnyak. "Rotation of a curve in an n-dimensional Euclidean space". *Siberian Math. J.* 29.1 (1988), 1–16.

[3] S. Alexander, R. Bishop. "The Fary–Milnor theorem in Hadamard manifolds". *Proc. Amer. Math. Soc.* 126.11 (1998), 3427–3436.

[4] K. Borsuk. "Sur la courbure totale des courbes fermées". *Ann. Soc. Polon. Math.* 20 (1947), 251–265.

[5] J. Cantarella, G. Kuperberg, R. Kusner, J. Sullivan. "The second hull of a knotted curve". *Amer. J. Math.* 125.6 (2003), 1335–1348.

[6] E. Denne. *Alternating quadrisecants of knots.* Ph.D. thesis–UIUC. 2004.

[7] E. Denne. *Quadrisecants and essential secants of knots: with applications to the geometry of knots.* 2016. arXiv: 1608.02608 [math.GT].

[8] T. Ekholm, B. White, D. Wienholtz. "Embeddedness of minimal surfaces with total boundary curvature at most 4π". *Ann. of Math. (2)* 155.1 (2002), 209–234.

[9] I. Fáry. "Sur la courbure totale d'une courbe gauche faisant un nœud". *Bull. Soc. Math. France* 77 (1949), 128–138.

[10] S. Fenner. *The total curvature of a knot (long).* URL: `www.ics.uci.edu/~eppstein/junkyard/knot-curvature.html`.

[11] B. C. Mazur. "On embeddings of spheres". *Acta Math.* 105 (1961), 1–17.

[12] J. Milnor. "On the total curvature of knots". *Ann. of Math. (2)* 52 (1950), 248–257.

[13] E. Pannwitz. "Eine elementargeometrische Eigenschaft von Verschlingungen und Knoten". *Math. Ann.* 108.1 (1933), 629–672.

[14] A. Petrunin, S. Zamora Barrera. *What is differential geometry: curves and surfaces.* 2021. arXiv: `2012.11814 [math.HO]`.

[15] C. Schmitz. "The theorem of Fáry and Milnor for Hadamard manifolds". *Geom. Dedicata* 71.1 (1998), 83–90.

[16] S. Stadler. "The structure of minimal surfaces in CAT(0) spaces". *J. Eur. Math. Soc. (JEMS)* 23.11 (2021), 3521–3554.

[17] J. Sullivan. "Curves of finite total curvature". *Discrete differential geometry.* Oberwolfach Semin. 2008, 137–161.

[18] B. White. "Half of Enneper's surface minimizes area". *Geometric analysis and the calculus of variations.* Int. Press, Cambridge, MA, 1996, 361–367.

[19] B. White. *Lectures on Minimal Surface Theory.* 2013. arXiv: `1308.3325 [math.DG]`.