Proofs (not) from The Book

S. Tabachnikov*

We all know about The Book in which, according to Erdős, God keeps the most elegant proof of each mathematical result. It is interesting to speculate whether every theorem has a "Book proof" (say what about the Four Color Theorem or the Collatz $3n+1$ Conjecture?) or whether a "Book proof" of a theorem is unique.

I believe many mathematicians have their own private collections of proofs from The Book. In fact, such a collection, and a highly successful one, was published by M. Aigner and G. M. Ziegler [1] (see [2] for a complement).

This article contains several proofs, not included in [1], that, in my opinion, are serious contenders for inclusion in The Book. Needless to say, the selection reflects my taste and mathematical interests.

Most of the theorems discussed here are 100–150 years old, but these proofs are considerably newer. This is another issue to mull over: how long does it take for a Book proof of a theorem to emerge?

Pick's Formula

Pick's Formula gives the area of a plane polygon whose vertices are points of the standard lattice $\mathbb{Z}^2$.

Theorem 1 (G. Pick, 1899 [21]) *Let I be the number of lattice points inside a simple polygon, B the number of lattice points on its boundary, including the vertices, and A its area. Then*

$$A = I + \frac{B}{2} - 1.$$

For an example, see Figure 1.

*Department of Mathematics, Penn State University, University Park, PA 16802, and ICERM, Brown University, Box 1995, Providence, RI 02912. e-mail: tabachni@math.psu.edu

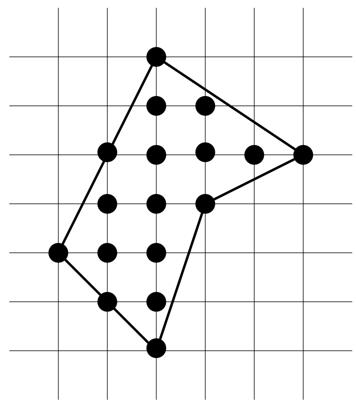

Figure 1: $I = 10, B = 7$, and $A = 12.5$.

There are many proofs of this classical result. The following one is due to C. Blatter [7].

Place a unit cube of ice at each lattice point in the plane and let the ice melt. The water will evenly distribute in the plane and, in particular, the amount of water inside the polygon will equal its area.[1]

Where does this water come from?

Consider the segment between two consecutive boundary points. The midpoint of this segment is a symmetry center of the lattice, so at each instant the water flow is centrally symmetric with respect to this midpoint. Therefore the total flow of water across the edge is zero, that is, the amount of water in the polygon does not change with time. Hence the final amount of water within the polygon comes from the interior and boundary lattice points.

The interior points contribute a unit of water each. A boundary point interior to an edge contributes half-a-unit of water, and it remains to account for the vertices of the polygon. A vertex contributes $\alpha/(2\pi)$ units of water where α is the interior angle at this vertex. Since the sum of the interior angles of an n-gon is $(n-2)\pi$, the total contribution of the vertices is

$$\frac{(n-2)\pi}{2\pi} = \frac{n}{2} - 1,$$

implying Pick's Formula.

[1]Not quite so, as the referee pointed out: the density of ice is lower than that of water. Strictly speaking, one should use cubes of ice of size about 1.09, the ratio of the densities of water and ice.

Pick's Formula does not extend to higher dimensions, but many results on lattice points in polytopes are known [6]. One wonders whether the idea of the above proof can be used in higher-dimensional setting.

Sperner's Lemma

Sperner's Lemma is a theorem in combinatorial geometry, a discretization of Brouwer's fixed point theorem. The statement of Sperner's Lemma is as follows.

Theorem 2 (E. Sperner, 1928 [24]) *Consider a triangulation of an n-dimensional simplex Δ whose vertices are labelled $0, 1, \dots, n$. Assume that the vertices of the triangulation are also "colored" $0, 1, \dots, n$, subject to the following constraint: the vertices on every facet of Δ do not use the color of the vertex opposite to this facet. Then the number of simplices of the triangulation colored in all $n + 1$ colors is odd; in particular, there is at least one such simplex.*

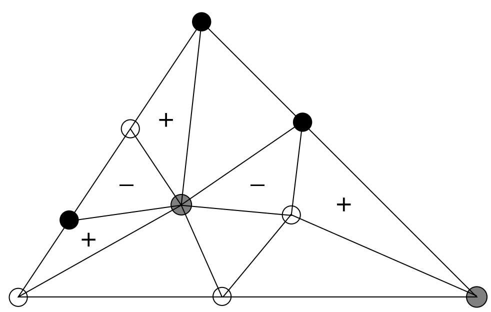

Figure 2: There are three positively and two negatively oriented 3-colored triangles. The colors are black, white, and gray.

For a 2-dimensional example, see Figure 2.

Most of the proofs of Sperner's Lemma are of homological nature. The proof presented here is taken from [17].

Let each vertex of the triangulation move, with constant speed, directly toward the vertex of Δ that has the same color. The speeds are chosen so that each vertex of the triangulation reaches the respective vertex of Δ after 1 unit of time.

Consider the oriented volume of a simplex of the triangulation as a function of time t. The volume of a simplex is given by a determinant involving its vertices: if the vertices are $P_0, P_1, \ldots, P_n$ then

$$\mathrm{Vol} = \frac{1}{n!} \det(P_1 - P_0, P_2 - P_0, \ldots, P_n - P_0).$$

If the vertices move with constant velocities, the volume is a polynomial in t (of degree equal to the dimension of the ambient space).

Consider the sum of volumes of the simplices of triangulation as a function of time; denote this sum by $V(t)$. Since scaling and reorienting do not affect our considerations, assume that the volume of Δ is 1. For small values of t, we have $V(t) = 1$. This is due to the constraint: since each vertex on a facet remains on this facet, for small values of t we still have a triangulation of Δ. Since the sum of volumes is a polynomial in t we have $V(t) = 1$ for all t.

What about $t = 1$? Each vertex has reached its destination, so the volume of the resulting simplex vanishes, unless all vertices were colored in different colors. In the latter case, the volume is ± 1, depending on the orientation. Since $V(0) = 1$, we have $V(1) = 1$ as well.

This implies that the difference between the number of positively and negatively oriented simplices of the triangulation, colored in all $n+1$ colors, is one. In particular, the number of simplices of the triangulation colored in all $n+1$ colors is odd.

Barbier's Theorem

Barbier's Theorem concerns curves of constant width. Recall that a convex curve has constant width if the distance between a pair of parallel support lines to the curve does not depend on the direction of these lines.

Theorem 3 (J.-É. Barbier, 1860 [5]) *The perimeter length of a curve of constant width w equals πw.*

The proof described below is by way of rolling; I am not sure of its origin.

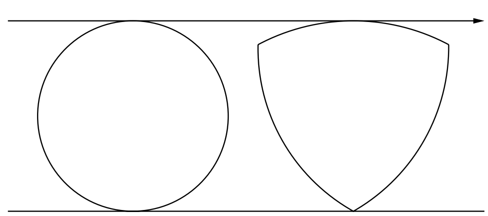

Figure 3: Figures of constant width as wheels.

Take two figures of the same constant width (say a disk and a Reuleaux triangle), and put them between two parallel horizontal lines. Let us roll the top line using the figures as wheels (or gears) see Figure 3.

The instantaneous motion of the wheel is rotation about its point of contact with the bottom line, and the angular speed of both wheels is the same. Indeed, the point of contact is instantaneously at rest, and the top point is moving with the speed of the moving line, say v, hence the angular speed is v/w.

The displacement of the moving line equals the perimeter length of the wheel times the number of turns. The latter is the same for both wheels since their angular speeds are equal, hence the former is also the same.

Fáry's Inequality

The average absolute curvature of a smooth closed plane curve γ is

$$\frac{\int_\gamma |\kappa(s)|\ ds}{L(\gamma)}$$

where s is an arc-length parameter, κ is the curvature, and $L(\gamma)$ is the perimeter length. The curve may self intersect.

Theorem 4 (Fáry, 1950 [8]) *If a closed curve is contained in the unit disk, then its average absolute curvature is not less than 1.*

I know four proofs of this results [26]; the one presented here originated at the Moscow Mathematical Olympiad in 1973.[2] The formulation of the olympiad problem was as follows: *A lion runs over a circular circus ring of radius 10 m. Moving along a polygonal line, the lion covers 30 km. Prove that the sum of the angles of all of the lions turns is not less than 2998 radians.*

The difference between this problem and Fáry's theorem is two-fold. Firstly, a smooth curve is replaced by a polygonal one, and the total curvature by the sum of the exterior angles of the polygonal line. Denote this sum by $C(\gamma)$. In the continuous limit, as the polygonal line approximates the smooth one, $C(\gamma)$, this discrete analog of total curvature, becomes $\int_\gamma |k(t)|\ dt$. Secondly, the trajectory of the lion is not closed; I leave it to the reader to deal with this minor issue and, in particular, to see why the answer to the olympiad problem is 2998 radians.

[2]That year that I graduated from high school in Moscow. I participated in the olympiad, tried, and failed, to solve this problem.

So, consider a closed polygonal line γ with sides e_i of length l_i. Starting with $i = 1$, rotate the side e_{i+1} about its common end-point v_i with e_i so that e_{i+1} becomes the extension of e_i. The rotation angle is equal to the exterior angle α_i of the polygonal line γ at the vertex v_i. In this way one unfolds γ into a straight line, as if it was a carpenter's rule, see Figure 4. In other words, one rolls γ along a straight line.

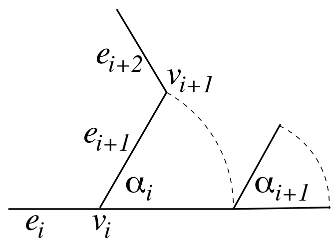

Figure 4: Unfolding a polygonal line.

Let the plane roll along with γ. The total displacement of the center O of the unit disk is a horizontal segment of length $\sum l_i = L(\gamma)$. The trajectory of the point O consists of arcs of circles of radii not greater than 1 (subtending the angles α_i); the length of such an arc does not exceed α_i. Clearly the length of the trajectory of O is not less than its total displacement, that is,

$$C(\gamma) = \sum \alpha_i \geq L(\gamma),$$

as needed.

Some readers may prefer a continuous version of the rolling argument, when the curve γ is smooth. It goes as follows.

As before, we think of γ as a kind of a wheel and roll it along a horizontal line once. The plane containing γ rolls along, and we consider the trajectory of the center of the disk O. The length of this trajectory is not less than $L(\gamma)$.

Let v be the instantaneous speed of the point of contact of γ with the horizontal line, ω be the angular velocity of the "wheel", and R the radius of curvature at the contact point. Then $v = R\omega$. We may assume that $v = 1$, and hence $\omega = 1/R = |\kappa|$ where κ its curvature at the contact point.

The instantaneous speed of the center of the disk O is $D\omega$, where D is the distance from O to the point of contact. Therefore the length of the trajectory of O is $\int D\omega \; ds$ where s is the length parameter along the horizontal line. Since $D \leq 1$, we have:

$$L(\gamma) \leq \int D\omega \; ds \leq \int \omega \; ds = \int |\kappa| \; ds,$$

as needed.

Interestingly, the Fáry's Inequality extends to the case when γ is contained inside a convex closed curve, say Γ. In this case, the claim is that the average absolute curvature of Γ is not greater than that of γ.[3] The proof of this generalized inequality is much more involved: see [14, 19, 15].

Altitudes of a spherical triangle

A well known theorem of Euclidean geometry asserts that the altitudes of a plane triangle are concurrent. The same result holds in spherical (and hyperbolic) geometry.

Theorem 5 *The altitudes of a spherical triangle intersect at one point.*

The proof presented here is due to V. Arnold who published a similar, albeit more involved, argument in hyperbolic geometry [4]; see [12] for a detailed exposition.

In spherical geometry, one has a one-to-one correspondence between points and oriented great circles ("lines"): this is the relation between the pole and the equator (the choice of two poles for an oriented equator is made by the right hand rule). This spherical duality interchanges points and lines and preserves the incidence relation. In particular, three lines are concurrent if and only if their poles are collinear.

Consider a spherical triangle ABC see Figure 5. Let P be the pole of the line AB. Then the altitude dropped from C to AB is the line PC (meridians are perpendicular to the equator).

Assume that the sphere has radius 1 and is centered at the origin. The point P is given by the cross-product $A \times B$, normalized to have length 1. The pole of the line PC is $P \times C = (A \times B) \times C$, again normalized. Likewise for the other two altitudes of the triangle ABC.

We want to show that the altitudes are concurrent or, equivalently, that their poles are collinear. Great circles are the intersections of the sphere with planes through the origin. Thus we want to prove that the position vectors of the poles lie in the same plane, that is, satisfy a linear relation. This relation is

$$(A \times B) \times C + (B \times C) \times A + (C \times A) \times B = 0,$$

the Jacobi identity for the cross-product!

[3] The curve γ inside Γ resembles DNA inside a cell which is a very long "curve" packed inside a small domain; for this reason, the generalized Fáry's Inequality is sometimes called the DNA inequality.

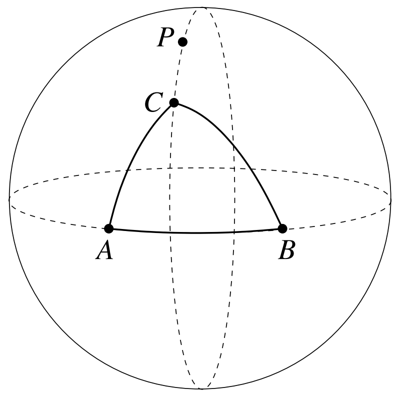

Figure 5: Altitude of a spherical triangle.

(As the referee pointed out, the above argument leaves aside some degenerate cases, for example, when $(A \times B) \times C = 0$. These degenerate cases can be obtained as limits of non-degenerate ones, so the theorem holds in the limit.)

The cross-product defines the structure of Lie algebra in $\mathbb{R}^3$, and this Lie algebra is isomorphic to $so(3)$, the Lie algebra of motions of the unit sphere. Likewise, the argument in the hyperbolic case involves $sl(2, \mathbb{R})$, the Lie algebra of motions of the hyperbolic plane. Curiously, there seems to be no such Lie algebraic argument in the Euclidean case, although the Euclidean theorem can be deduced from the spherical (or the hyperbolic) one as a limiting case, as the curvature goes to zero.

Sturm-Hurwitz Theorem

A trigonometric polynomial of degree N is a function

$$f(x) = c + \sum_{k=1}^{N} a_k \cos kx + b_k \sin kx.$$

We consider f as a function on the circle of length 2π.

It is well known that a trigonometric polynomial of degree N has at most $2N$ roots. A lesser-known result gives the lower bound on the number of roots.

Theorem 6 (C. Sturm, 1836 [25], A. Hurwitz, 1903 [11]) *Let*

$$f(x) = \sum_{k=n}^{N} a_k \cos kx + b_k \sin kx. \tag{1}$$

Then the number of sign changes of f is at least $2n$.

More generally, and in words: *the number of roots of a periodic function is not less than that of its first harmonic.*

As with other results discussed here, there are many proofs of this theorem, see [23] and [20]. Following [13, 16], we present a proof by way of Rolle's theorem. This proof can be extended from trigonometric polynomials to smooth functions; to avoid technicalities, we do not dwell on this generalization.[4]

Denote by $Z(f)$ the number of sign changes of a function f defined on the circle. Rolle's theorem asserts that $Z(f') \geq Z(f)$: indeed, the derivative changes sign between consecutive sign changes of a function.

Introduce the operator D^{-1}, the inverse derivative, defined on the space of functions with zero average:

$$(D^{-1}f)(x) = \int_0^x f(t)\,dt.$$

The constant of integration is chosen so that the inverse derivative again had zero average (we need this assumption for the inverse derivative again to be periodic). Rolle's theorem then reads: $Z(f) \geq Z(D^{-1}f)$.

Consider the sequence of functions

$$f_m = (-1)^m \left(nD^{-1}\right)^{2m} f$$

where f is the trigonometric polynomial (1); explicitly,

$$f_m(x) = (a_n \cos nx + b_n \sin nx) + \sum_{k=n+1}^{N} \left(\frac{n}{k}\right)^{2m} (a_k \cos kx + b_k \sin kx)\,.$$

By Rolle's theorem, for every $m \geq 1$, one has: $Z(f) \geq Z(f_m)$.

As $m \to \infty$, the function $f_m(x)$ gets arbitrarily close to $a_n \cos nx + b_n \sin nx$. This pure harmonic, if not identically zero, changes sign exactly $2n$ times, hence so does f_m, for m large enough. Therefore $Z(f) \geq 2n$.

The inverse derivative operator is a discrete analog of the heat operator, and indeed, one can prove the Sturm-Hurwitz theorem using the heat flow, see [22]. The argument goes as follows.

Let $f(x)$ be the initial distribution of heat on the circle. Consider the propagation of heat described by the heat equation

$$\frac{\partial F(x,t)}{\partial t} = \frac{\partial^2 F(x,t)}{\partial x^2}, \qquad F(x,0) = f(x).$$

[4] As the referee pointed out, this disclaimer may disqualify the proof from The Book.

The number of sign changes of $F(x,t)$, considered as a function of x, does not increase with t: an iceberg can melt down in a warm sea but cannot appear out of nowhere (this is the maximum principle in PDE).

On the other hand, one can solve the heat equation explicitly:

$$F(x,t) = \sum_{k \geq n} e^{-k^2 t} \left(a_k \cos kx + b_k \sin kx \right).$$

The rest of the argument is as before: the higher harmonics tend to zero faster than the first non-trivial one. Thus, $F(x,t)$ has at least $2n$ zeroes for t large enough.

Four-Vertex Theorem

A vertex of a plane curve is a local extremum of its curvature. By an oval we mean a closed smooth strictly convex curve.

Theorem 7 (S. Mukhopadhyaya, 1909 [18]) *A plane oval has at least four vertices.*

Since its publication, the Four-Vertex theorem has generated a vast literature that includes numerous proofs and generalizations. The presented proof is due to R. Thom [27]; we follow the exposition in [9]. For the concept of curvature via the osculating circles, see [10].

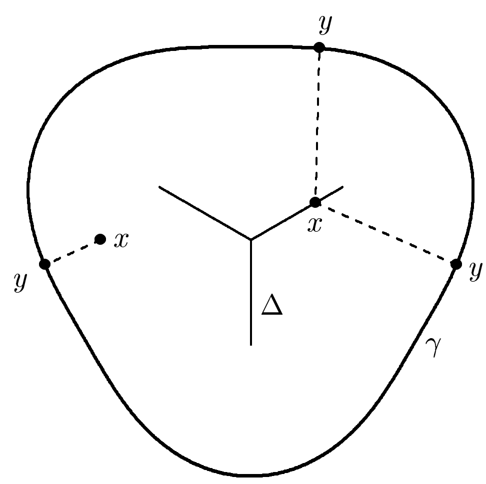

Figure 6: The symmetry set of an oval

For every point x inside the oval γ, consider the closest point y on the oval. Of course, for some interior points x, the closest boundary point y is not unique. The locus of such points x is called the *symmetry set*; denote it by Δ. For example, for a circle, Δ is its center, and for an ellipse, Δ is the segment between the two centers of maximal curvature. For a generic oval, Δ is a graph, and its vertices of valence 1 are the centers of local maximal curvature of γ see Figure 6.

Let us justify the last claim. It is clear that the vertices of Δ of valence 1 are the centers of extremal curvature (where two points labeled y in Figure 6 merge together). But why not centers of minimal curvature? This is because an osculating circle of minimal curvature locally lies outside of the curve γ. Therefore the distance from the center of such a circle to the curve is less than its radius and hence its center does not belong to the symmetry set Δ.

Delete the symmetry set from the interior of γ. What remains can be continuously deformed to the boundary oval by moving every point x toward the closest point y. Hence the complement of Δ is an annulus, and therefore Δ has no loops (and consists of only one component, for that matter). Thus Δ is a tree which necessarily has at least two vertices of valence 1. It follows that the curvature of the oval has at least two local maxima, as needed.

Let us conclude with remarks relating the last two results.

One of the proofs of the Four-Vertex theorem deduces it from the Sturm-Hurwitz theorem for $n = 2$, applied to the support function $p(x)$ of the oval. Namely, the vertices correspond to zeros of $p' + p'''$. The Fourier expansion of the function $p' + p'''$ starts with the second harmonics (the constant term and the first harmonics are annihilated by the differential operator $d/dx + d^3/dx^3$), hence $p' + p'''$ has at least four zeros.

And another proof of the Four-Vertex theorem makes use of curve shortening, an analog of the heat flow for curves, see [3].

Acknowledgments. I am grateful to numerous colleagues for sharing their favorite Book proofs and criticizing mine. In particular, many thanks to B. Khesin, M. Levi, V. Ovsienko, I. Scherbak, R. Schwartz, T. Tokieda, and G. M. Ziegler for their criticism of the first drafts of this article. The numerous thoughtful suggestions of the referee helped to improve the exposition. This work was partially supported by the NSF grant DMS-1105442.

References

[1] M. Aigner, G. M. Ziegler. *Proofs from THE BOOK*. Fourth edition. Springer-Verlag, Berlin, 2010.

[2] M. Aigner, G. M. Ziegler. "Brillanten für das 'BUCH der Beweise'". `mathe-lmu.de`, Math. Dept., Ludwigs-Maximilians University, No. 7, 2003, 22–28.

[3] S. Angenent. "Inflection points, extatic points and curve shortening". Hamiltonian systems with three or more degrees of freedom, 3–10, NATO Adv. Sci. Inst. Ser. C Math. Phys. Sci., 533, Kluwer, 1999.

[4] V. Arnold. "Lobachevsky triangle altitudes theorem as the Jacobi identity in the Lie algebra of quadratic forms on symplectic plane". *J. Geom. Phys.* **53** (2005), 421–427.

[5] E. Barbier. "Note sur le problème de laiguille et le jeu du joint couvert". *J. Math. Pure Appl.* **5** (1860), 273–286.

[6] M. Beck, S. Robins. *Computing the continuous discretely. Integer-point enumeration in polyhedra.* Springer, New York, 2007.

[7] C. Blatter. "Another Proof of Pick's Area Theorem". *Math. Mag.* **70** (1997), 200.

[8] I. Fáry. "Sur certaines inégalites géométriques". *Acta Sci. Math. Szeged* **12** (1950), 117–124.

[9] D. Fuchs, S. Tabachnikov. *Mathematical omnibus. Thirty lectures on classic mathematics.* Amer. Math. Soc., Providence, RI, 2007.

[10] E. Ghys, S. Tabachnikov, V. Timorin. "Osculating curves: around the Tait-Kneser Theorem". Math. Intelligencer, 35 (2013), No 1, 61–66

[11] A. Hurwitz. "Über die Fourierschen konstanten integrierbarer funktionen". *Math. Ann.* **57** (1903), 425–446.

[12] N. Ivanov. "V. Arnol'd, the Jacobi identity, and orthocenters". *Amer. Math. Monthly* **118** (2011), 41–65.

[13] G. Katriel. "From Rolle's Theorem to the Sturm-Hurwitz Theorem". arXiv:math/0308159.

[14] J. Lagarias, T. Richardson. "Convexity and the average curvature of plane curves". *Geom. Dedicata* **67** (1997), 1–30

[15] E. Larson. "The DNA inequality in non-convex regions". *Adv. Geom.* **10** (2010), 221–248.

[16] Y. Martinez-Maure. "Les multihérissons et le théorème de Sturm-Hurwitz". *Arch. Math.* **80** (2003), 79–86.

[17] A. McLennan, R. Tourky. "Using volume to prove Sperner's lemma". *Econom. Theory* **35** (2008), 593–597.

[18] S. Mukhopadhyaya. "New methods in the geometry of a plane arc". *Bull. Calcutta Math. Soc.* **1** (1909), 32–47.

[19] A. Nazarov, F. Petrov. "On a conjecture of S. L. Tabachnikov". *St. Petersburg Math. J.* **19** (2008), 125–135.

[20] V. Ovsienko, S. Tabachnikov. *Projective differential geometry old and new. From the Schwarzian derivative to the cohomology of diffeomorphism groups.* Cambridge Univ. Press, Cambridge, 2005.

[21] G. Pick. "Geometrisches zur Zahlentheorie". *Sitzenber. Lotos (Prague)* **19** (1899), 311–319.

[22] G. Pólya. "Qualitatives über Wärmeausgleich". *Z. angew. Math. Mech.* **13** (1933), 125–128.

[23] G. Pólya, G. Szégő. *Problems and theorems in analysis.* Springer-Verlag, Berlin, 1998.

[24] E. Sperner. "Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes". *Abh. Math. Sem. Univ. Hamburg* **6** (1928), 265–272.

[25] C. Sturm. "Sur une classe dequations à différences partielles". *J. Math. Pure Appl.* **1** (1836), 373–444.

[26] S. Tabachnikov. "The tale of a geometric inequality". MASS Selecta, 257–262, Amer. Math. Soc., Providence, RI, 2003.

[27] R. Thom. "Sur le cut-locus d'une variété plongée". *J. Differential Geometry* **6** (1972), 577–586.