The Erdős-Nagy Theorem and its Ramifications

Godfried Toussaint * School of Computer Science McGill University Montréal, Québec, Canada

May 11, 1999

Abstract

Given a simple polygon in the plane, a flip is defined as follows: consider the convex hull of the polygon. If there are no pockets do not perform a flip. If there are pockets then reflect one pocket across its line of support of the polygon to obtain a new simple polygon. In 1934 Paul Erdős conjectured that every simple polygon will become convex after a finite number of flips. The result was first proved by Béla Nagy in 1939. Since then it has been rediscovered many times in different contexts, apparently, with none of the authors aware of each other's work. The purpose of this paper is to bring to light this "hidden" work. We review the history of this problem, provide a simple elementary proof of the theorem and consider variants, generalizations and applications of interest in computational knot theory and molecular biology. We also uncover an incorrect "walk" algorithm in the knot theory literature and show how it can be fixed with the Erdős-Nagy theorem and other more efficient methods. We close with several open problems.

1 Introduction

Let $A = A_1A_2A_3A_4$ be a nonconvex quadrilateral in the two-dimensional xy-plane with A_3 as its reflex vertex (refer to Figure 1). Furthermore, assume that



Figure 1: Rotating a reflex vertex using the third dimension.

the quadrilateral (although planar) is embedded in the 3D space with axes x, y and z, that the vertices are *ball-joints* which allow rotations in all directions in 3D. Finally, assume the links (edges) are rigid line segments with $A_1A_2 = A_1A_4$ and $A_2A_3 = A_3A_4$. If we lift vertex A_3 off the xy-plane into the third dimension z (leaving the other three vertices fixed) by rotating it about the line through A_2 and A_4 until it returns to the xy-plane at position B_3 , then the quadrilateral has been convexified with one simple motion. This rotation motion in 3D is equivalent to a reflection transformation in the xy-plane: B_3 is the reflection of A_3 across the line through A_2 and A_4 .

A generalization of this problem has been discovered and re-discovered independently by several mathematicians, biologists, physicists and computer scientists dating back to 1935. Computer scientists are motivated by practical robotics problems with linkages. Molecular biologists and polymer physicists are interested in unravelling large molecules (modeled

^{*}This research was supported by NSERC and FCAR. e-mail: godfried@cs.mcgill.ca



Figure 2: Flipping the pockets of a polygon.

as polygons) such as circular DNA [7]. Mathematicians are curious about the geometric properties of polygons and simple closed curves.

The first person to propose this problem appears to be Paul Erdős in 1935 [6] in the context of planar polygons. Consider the simple polygon P in Figure 2 (a). If we subtract this polygon from its convex hull we obtain the convex deficiency: a collection of connected regions. Each such region together with its boundary is itself a polygon, often called a *pocket* of P. The polygon P in Figure 2 (a) has two pockets P_1 and P_2 . Each pocket has an edge which coincides with a convex hull edge of P (shown in the figure by dotted lines). Such an edge is called the pocket *lid*.

Erdős defined a reflection operation on P as a simultaneous reflection of all the pockets of P about their corresponding pocket lids. Applying a reflection operation to polygon P in Figure 2 (a) yields the new polygon P' in Figure 2 (b). In 1935 Erdős conjectured that given any simple polygon, a finite number of such reflection steps will convexify it. The first proof of Erdős' conjecture was provided in 1939 by Béla Nagy [15]. First Nagy observed that reflecting all the pockets in one step can lead from a simple polygon to a non-simple one. One such example due to Nagy is shown in Figure 3. Therefore he modified Erdős' problem slightly by defining one step to be the reflection of only *one* pocket. Since a pocket is reflected into a previously empty half-plane, no collisions can occur with such a motion. Let us call such an operation a flip.

Figure 4 shows a polygon being convexified after four flips. The pockets at each flip are shown in white before flipping and shaded after the flip is completed.



Figure 3: Flipping two pockets simultaneously may lead to a crossing polygon.



Figure 4: The polygon on the left is convexified after four flips.

Nagy then proceeded to prove that any simple polygon can be convexified by a *finite* number of flips.

2 Rediscoveries of the Erdős-Nagy Theorem

Branko Grünbaum [8] described some of the strange history of this problem and uncovered several rediscoveries of the theorem. He also provided his own version of a proof which is similar to Nagy's proof with one of the main differences being that at each step he flips the pocket that has maximum area (if more than one pocket exists). Since [8] is rather inaccessible, here we first briefly outline his findings and then add some more rediscoveries and variants to the history of this problem.

As mentioned previously, in 1939 Béla Nagy changed Erdős' problem slightly by reflecting only one pocket of the polygon at each step so that simplicity is maintained during the convexification process. As we shall see later, maintaining simplicity during the process is not necessary if the definition of a flip is suitably modified. In 1957 there appeared two Russian papers by Reshetnyak [16] and Yusupov [23] proving the theorem with variants of basically the same proof.

In 1959 Kazarinoff and Bing [12] announced the problem with a solution. Two years later a proof appeared in a paper by Bing and Kazarinoff [3] and also in Kazarinoff's book [11]. They also conjectured that every simple polygon will be convex after at most 2n flips.

In 1973 two students of Grünbaum at the University of Washington, R. R. Joss and R. W. Shannon worked on this problem but did not publish their results. An account of the unfortunate circumstances surrounding this event is given by Grünbaum [8]. They found a counter-example to the conjecture of Bing and Kazarinoff (unaware of the conjecture of course). They showed that given any positive integer k, there exist simple polygons (indeed quadrilaterals suffice) that cannot be convexified with fewer than k flips.

In 1981 Kaluza [10] poses the problem again and asks if the number of flips could be bounded as a function of the number of vertices of the polygon.

In 1993 Bernd Wegner [22] takes up Kaluza's challenge and solves both problems. His proof of convexification in a finite number of flips is quite different from the others but his example for unboundednes is the same as that of Joss and Shannon.

In 1999 Biedl et al., [2] rediscover the problem again and obtain the same results as Wegner. Their proofs of convexification are remarkably similar and their unboundedness example is the same.

3 A Proof of the Erdős-Nagy Theorem

Some of the published proofs of the Erdős-Nagy theorem are long and technical, others make references to higher mathematics, and some have gaps. Therefore it is appropriate to borrow the best features of the existing proofs, fill in the gaps, and present a simple, clear, elementary and short proof of the theorem. In this section we present such a proof. First we consider a simple lemma that will be used in the proof.



Figure 5: The convexity tolerance disk of a vertex.

Lemma 1 Given a convex polygon, there exists a positive real number ϵ such that if some or all of the vertices are each moved by a distance less than ϵ , then the polygon remains convex.

Consider vertex A_i and its two adjacent Proof: vertices A_{i-1} and A_{i+1} (refer to Figure 5). Let L_i be the line passing through the midpoints of the two edges $A_i A_{i-1}$ and $A_i A_{i+1}$, and let r_i denote the minimum distance between A_i and L_i . Note that r_i is also the minimum distance between A_{i-1} and L_i as well as between A_{i+1} and L_i . Now construct disks D_i , D_{i-1} and D_{i+1} , all of the same radius r_i , centered at A_i , A_{i-1} and A_{i+1} , respectively. No matter where the vertices move, as long as each remains in the interior of its corresponding disk, their final positions B_i , B_{i-1} and B_{i+1} will have the property that B_i is separated by the line L_i from B_{i-1} and B_{i+1} . Therefore vertex B_i is convex. If we choose for the radius of our disk for every vertex the value $\epsilon = \min\{r_1, r_2, ..., r_n\}$ then all vertices will remain convex and since the polygon has no crossings it is convex.

This number ϵ is sometimes called the *convexity-tolerance* of the polygon [1]. It is a measure of how much the vertices of a convex polygon may be perturbed while guaranteeing that the polygon remains convex.

Theorem 1 Every closed simple polygon can be convexified with a finite number of flips.

Proof: Let $A^0 = A_1^0 A_2^0 \dots A_n^0$ denote the given polygon before any flips have taken place. After performing k flips we obtain the polygon $A^k = A_1^k A_2^k \dots A_n^k$ where vertex A_i^0 is taken to A_i^k for all i = 1, 2, ..., n.



Figure 6: A flip increases the distance from a fixed point and a vertex.

We will call polygon A^m a descendant of A^k if m > k. Consider any point x in A^0 . Since for all k, A^{k+1} contains A^k , point x remains in all the descendants of A^0 . We are interested in the distance betwen point x and a vertex of the k-th descendant of A^0 , $d(x, A_i^k)$. After the next flip A_i^k either remains fixed or is reflected across a line of support L (refer to Figure 6). In the latter case this line is the perpendicular bisector of the segment $A_i^k A_i^{k+1}$. Let x' denote the intersection of line L with segment xA_i^{k+1} . Then we have

$$d(x, A_i^{k+1}) = d(x, x') + d(x', A_i^{k+1}).$$

Since x' is equidistant from A_i^k and A_i^{k+1} we obtain

$$d(x, A_i^{k+1}) = d(x, x') + d(x', A_i^k).$$

It follows from the triangle inequality that

$$d(x, A_i^{k+1}) \ge d(x, A_i^k).$$

Therefore the distance function $d(x, A_i^k)$ is a monotonically non-decreasing function of k. Furthermore, since the edges are rigid, the perimeter of every descendant of A^0 remains constant after every flip. Therefore the distance $d(x, A_i^k)$ is bounded from above by half the perimeter of A^0 . From these two observations it follows that the sequence $\{A_i^0 A_i^1 A_i^2 ...\}$ has a limit. Let us denote the limit of A_i^k , as k goes to infinity, by A_i^* and let $A^* = A_1^* A_2^* ... A_n^*$ denote the limit polygon.

Firstly we remark that the limit polygon A^* must be a simple polygon. In other words, different vertices cannot converge to one and the same limit vertex. This follows from the observation above that $d(x, A_i^k)$ is a monotonically non-decreasing function of k, where the role of x is now played by another vertex A_j^k where $j \neq i$. If both A_j^k and A_i^k move with the next flip then

$$d(A_i^{k+1}, A_i^{k+1}) = d(A_i^k, A_i^k)$$

If only A_i^k moves, then

$$d(A_j^{k+1}, A_i^{k+1}) \ge d(A_j^k, A_i^k).$$

Therefore two vertices of A^k cannot move closer together when we flip A^k .

Secondly we note that the limit polygon A^* must be convex, for otherwise, being a simple polygon, another flip would alter its shape contradicting that it is the limit polygon.

Thirdly, some vertices of A^* will have interior angles equal to π and others less than π . Note also that whenever a vertex A_i^k becomes straight it remains straight for all descendants of A^k . Therefore we may ignore straight vertices in the analysis.

It remains to show that the sequence $\{A^0, A^1, \dots, A^k\}$ where $A^k = A^*$ is finite. To this end let us now construct around each vertex A_i^* whose interior angle is less than π a disk D_i of radius ϵ , the convexity tolerance of $A^*.$ Consider the sequence of positions of the *i*-th vertex $\{A_i^0 A_i^1 A_i^2 \dots\}$. Since A_i^m converges to A_i^* as *m* approaches infinity, there must exist a finite number c_i of flips after which $A_i^{e_i}$ first enters disk D_i . Furthermore, once it enters D_i it stays there. This follows from the fact that A^{c_i} is contained in A^* and L_i separates $A_i^{c_i}$ from $A_{i+1}^{c_i}$ and $A_{i-1}^{c_i}$. In fact, $A_i^{c_i}$ cannot be contained in the interior of any pockets of subsequent descendants and is immobilized. If we let $c^* = max\{c_1, c_2, ..., c_n\}$, then after c^* flips every vertex has entered its limit disk and since the vertices must then remain in their respective disks A^{c^*} must be convex. Hence A^k is convex for $k = c^*$ flips.

4 Variants and Generalizations

4.1 Mouth Flips

Knot theorists are interested in polygons in 3D (knots). In particular, for the computer analysis of

knot spaces (or exploring the respective variety) they are interested in "walk" algorithms that will take one knot into another. Millet [14] rediscoverd a special case of the Erdős-Nagy theorem when the polygons are (1) star-shaped, (2) equilateral (all edges have equal length) and (3) a flip is made not on a complete pocket of the polygon but only on a reflex vertex reflected across the line joining its adjacent vertices. We will call such a flip a mouth-flip. Millet proves that ultimately enough mouth-flips convexify the polygon. However, one can prove with an argument similar to that in [8] that not only will the polygon be convexified after a finite number of mouthflips but this number can be bounded as a function of n because the polygon is equilateral. To see this note that the *before-after* positions of a mouth form a parallelogram. Therefore no new slopes (aside from the slopes of the edges of the original polygon) are ever introduced by mouth-flipping. But the area strictly increases after each mouth-flip. Therefore each new polygon generated on the path towards convexity is composed of a new permutation of the edges (no permutation is revisited during this walk). Therefore the number of mouth-flips is bounded by the number of permutations. We therefore have the following theorem.

Theorem 2 A star-shaped equilateral polygon with n vertices can be convexified with at most (n - 1)!mouth-flips.

4.2 **Pivots and Hyperplane Flips**

One way to generalize the original Erdős-Nagy flip is to consider any two vertices of the polygon and to reflect one of the polygonal chains they determine across the line they define. An additional generalization is obtained if the selected chain is not reflected but rotated (about the line as axis) by some angle (assuming the polygon is embedded in 3D). Finally, a third further generalization is to polygons in d dimensions. Combining all three ideas leads to a motion which in knot theory is called a *pivot* [14]. Erdős-Nagy flips may be considered as special cases of pivots with planar polygons in 3D where the pairs of vertices are determined by lines of support of the polygon and each rotation has angle π . Indeed, under this generalization one may ask Erdős' question for arbitrary (possibly crossing) planar polygons. Grünbaum and Zaks [9] have shown that every planar polygon (possibly crossing) may be convexified after a finite number of pivots.

Another special case of pivots which is a natural generalization of Erdős-Nagy flips is as follows. Let P be a polygon in \mathbb{R}^d and let H be a hyperplane supporting the convex hull of P and containing at least two vertices of P. Reflect one of the resulting polygonal chains across H. Let us call such motions hyperplane-flips. The first person to propose these hyperplane-flips appears to be Gustave Choquet [5] in 1945 for applications to curve stretching, a topic to be discussed below. He claimed in [5] (but published no proof) that after a suitable choice of a countable number of hyperplane-flips the polygons. These results were rediscovered in 1973 by Sallee [19].

In 1994 Millet [14], in connection with exploring varieties, proposed a "walk" algorithm (sequence of pivots) to take any equilateral polygon (knot) in 3D into any other. The interest in equilateral polygons comes from molecular biology where homogeneous macromolecules such as DNA are modelled by polygons with equal length edges. Here the vertices correspond to the atoms and the edges to the bonding force between them. To establish the walk Millet proposed taking an arbitrary polygon P in 3D to a planar regular polygon. His algorithm consists of three parts: (1) convert P to a *planar* star-shaped polygon P', (2) convert P' to a convex polygon P'' and (3) convert P'' to a regular polygon. Part (2) is done using the mouth-flips discussed above. However, his algorithm for part (1) does not always work correctly. His procedure may yield non-simple planar polygons in which all turns are right turns and the winding number is high thus invalidating step (2) of the algorithm. However, we can obtain a walk algorithm by modifying (1) and applying the Erdős-Nagy theorem for (2). Furthermore, this modification generalizes Millet's theorem to polygons in d dimensions with no restrictions on edge lengths. Consider the first four vertices of P. They determine a possibly skew quadrilateral. Rotate one of the triangles so that the

quadrilateral is planar (one pivot). If the quadrilateral is not convex apply Erdős flips (pivots) to it until it is convex. Note that some of these pivots may carry the remaining polyon with them. Now advance to the next vertex of P, pivot this triangle so it is co-planar with the quadrilateral and again apply flips to the pentagon if it is not convex. Continuing this process leads to convexification with pivots only. Therefore we have the following result.

Theorem 3 In dimensions higher than two any polygon can be convexified with a finite number of pivots.

Of course it follows from our previous discussion that the number of pivots in Theorem 3 cannot be bounded as a function of n. However, convexification is possible in a polynomial number of moves if we are willing to use more complicated motions. For example in 1995 Lenhart and Whitesides [13] showed that (in any dimension greater than two) a polygon may be convexified in O(n) time with O(n) 5*joint* line-tracking motions. Each such motion rotates five joints with two cooperating "elbows". In 1973 Sallee [19] proved that this can be accomplished with 4-joint line-tracking motions. He gives no complexity analysis in [19] but examination of his algorithm reveals that it can be accomplished in O(n) time with $O(n^2)$ such motions. We can improve Sallee's number of motions and the complexity of the line-tracking motions of Lenhart and Whitesides [13] by maintaining a convex quadrilateral at each step to obtain the following result.

Theorem 4 In dimensions higher than two any polygon may be convexified in O(n) time with O(n) 4-joint line-tracking motions.

4.3 Curve Inflation

A generalized version of Erdős' problem for the case of arbitrary simple curves has also been discovered independently. In this context the operation is referred to as *inflation*. Flipping several arcs simultaneously as originally proposed by Erdős is called *full inflation* and flipping only one arc is called *partial inflation*. For sufficiently smooth curves Robertson [17] proves that they concerge to a convex curve after a suitable infinite sequence of flips. Robertson and Wegner [18] investigate the degree of smoothness of the limit curves obtained by flipping

4.4 Stretching

Let $A = A_1, A_2, ..., A_n$ be a polygon that is reconfigured to another $B = B_1, B_2, ..., B_n$. In other words the corresponding line segments have the same length and to each point on A there corresponds a point on B in the obvious way. If for every two points on A their corresponding points on B are further apart then we say that B is a *stretched* version of A. In 1973 Sallee [19] proved that for every polygon in d dimensions there exists a planar convex stretched version. Furthermore he gives an algorithm for carrying out the reconfiguration. Therefore these are stronger results than the convexification results mentioned earlier. The same results were aparently obtained as early as 1945 by Choquet [5].

4.5 Flipturns

Consider a planar polygon with a pocket determined by vertices A_i and A_j . Another generalization of the Erdős flip was considered in 1973 by Joss and Shannon [8] where after flipping the pocket we rotate it by 180 degrees about the center of the convex hull edge that determines the pocket. The effect of this kind of flip which they called a *flipturn* is that no new slopes are introduced after a flipturn. What was automatically obtained in the case of mouthflips for star-shaped equilateral polygons is obtained here for any simple polygon by flipping and "turning". Joss and Shannon proved that any simple polygon with n sides can be convexified by a sequence of at most (n-1)! flipturns. This bound is very loose and they conjectured that $(n^2)/4$ flipturns are always sufficient. Grünbaum and Zaks [9] showed that even crossing polygons could be convexified with a finite number of flipturns. In 1999 Therese Biedl discovered a polygon such that a bad sequence of flipturns leads to convexification only after $\theta(n^2)$ flipturns.

4.6 Non-Crossing Linkages

None of the work discussed above, apart from the original Erdős-Nagy theorem, is concerned with whether or not edges of the polygon cross each other during reconfiguration. However, in linkage analysis in robotics and in some molecular biology problems the edges are to be considered as physical barriers so that no crossings are allowed. Biedl et al., [2] explore the area of convexifying polygons under these constraints. A survey of this area can be found in [20]. In three dimensions unknotted polygons that cannot be convexified have been discovered independently by Biedl et al., [2] (with ten edges) and Cantarella and Johnston [4] (with six edges). Toussaint [21] discovered an additional class of stuck unknotted hexagons.

5 Conclusion and Open Problems

We conclude by mentioning several open problems in this area.

1. Wegner [22] proposed a very interesting variant of Erdős flips which can be considered the inverse problem which he called *deflation*. Given a simple polygon P in the plane, if there exists a pair of non-adjacent vertices A_i and A_j such that the line through A_i and A_j is not a line of support of P, the line intersects the boundary of the polygon only at A_i and A_j , and the polygonal chain $A_i, A_{i+1}, ..., A_j$ can be reflected about this line to lie inside the polygon then this reflection operation is called a *deflation*. If this cannot be done the polygon is called *deflated*. Wegner conjectured that every simple polygon can be deflated with a finite number of deflations.

2. Wegner also introduced two measures of convexity for simple polygons that are functions of the number of flips that will convexify the polygon. He called these the *maximal* and *minimal* inflation complexities. The former is the maximum number of flips that will convexify a polygon. The latter is the minimum number of flips. There are polygons (quadrilaterals) for which these two numbers are the same. What is the computational complexity of computing these numbers?

3. The Joss-Shannon conjecture that every simple polygon can be convexified with at most $(n^2)/4$ flipturns is still open. In fact, no upper bound lower than (n-1)! is known!

4. The results concerning stuck unknotted hexagons in [4] and [21] show that there exist at least five classes of nontrivial embeddings of the hexagon in 3D. It is conjectured that there are no more than five such classes.

References

- Manuel Abellanas, Ferran Hurtado, and Pedro Ramos. Tolerance of geometric structures. In Proc. 6th Canadian Conference on Computational Geometry, pages 250-255, 1994.
- [2] T. Biedl, E. Demaine, M. Demaine, S. Lazard, A. Lubiw, J. O'Rourke, M. Overmars, S. Robbins, I. Streinu, G. T. Toussaint, and S. Whitesides. Locked and unlocked polygonal chains in 3d. In Proc. 10th ACM-SIAM Symposium on Discrete Algorithms, pages 866-867, 1999.
- [3] R. H. Bing and N. D. Kazarinoff. On the finiteness of the number of reflections that change a nonconvex plane polygon into a convex one. *Matematicheskoe Prosveshchenie*, 6:205-207, 1961. In Russian.
- [4] Jason Cantarella and Heather Johnston. Nontrivial embeddings of polygonal intervals and unknots in 3-space. Journal of Knot Theory and its Ramifications, 7(8):1027-1039, 1998.
- [5] Gustave Choquet. Variétés et corps convexes. In Association Française pour l'Avancement des Sciences, 1945. Congrès de la Victoire-1945.
- [6] Paul Erdős. Problem number 3763. American Mathematical Monthly, 42:627, 1935.
- [7] Maxim D. Frank-Kamenetskii. Unravelling DNA. Addison-Wesley, 1997.
- [8] Branko Grünbaum. How to convexify a polygon. Geombinatorics, 5:24-30, 1995.

- [9] Branko Grünbaum and Joseph Zaks. Convexification of polygons by flips and by flipturns. Technical Report 6/4/98, Department of Mathematics, University of Washington, Seattle, 1998.
- [10] T. Kaluza. Problem 2: Konvexieren von Polygonen. Math. Semesterber., 28:153-154, 1981.
- [11] N. D. Kazarinoff. Analytic Inequalities. Holt, Rinehart and Winston, 1961.
- [12] N. D. Kazarinoff and R. H. Bing. A finite number of reflections render a nonconvex plane polygon convex. Notices of the American Mathematical Society, 6:834, 1959.
- [13] William J. Lenhart and Sue H. Whitesides. Reconfiguring closed polygonal chains in euclidean d-space. Discrete and Computational Geometry, 13:123-140, 1995.
- [14] K. Millet. Knotting of regular polygons in 3space. Journal of Knot Theory and its Ramifications, 3:263-278, 1994.
- [15] Béla Nagy. Solution of problem 3763. American Mathematical Monthly, 46:176-177, 1939.
- [16] Yu. G. Reshetnyak. On a method of transforming a nonconvex polygonal line into a convex one. Uspehi Mat. Nauk, 12(3):189-191, 1957. In Russian.
- [17] S. A. Robertson. Inflation of plane curves. In Geometry and Topology of Submanifolds-III, pages 264-275. World Scientific, 1991.
- [18] S. A. Robertson and B. Wegner. Full and partial inflation of plane curves. In *Intuitive Geometry*, *Colloquia Math. Soc. Janos Bolyai*, pages 389– 401. North-Holland, Amsterdam, 1994.
- [19] G. T. Sallee. Stretching chords of space curves. Geometriae Dedicata, 2:311-315, 1973.
- [20] G. T. Toussaint. Computational polygonal entanglement theory. In VIII Encuentros de Geometría Computacional, July 7-9 1999. Castellon, Spain.

- [21] G. T. Toussaint. A new class of stuck unknots in Pol₆. Technical Report SOCS-99.1, School of Computer Science, McGill University, April 1999.
- [22] Bernd Wegner. Partial inflation of closed polygons in the plane. Contributions to Algebra and Geometry, 34(1):77-85, 1993.
- [23] A. Ya. Yusupov. A property of simply-connected nonconvex polygons. Uchen. Zapiski Buharsk. Gos. Pedagog., pages 101-103, 1957. In Russian.