Convexifying Polygons in 3D: a Survey

Michael Soss and Godfried T. Toussaint

Abstract. To convexify a polygon is to reconfigure it with respect to a given set of operations until the polygon becomes convex. The problem of convexifying polygons has had a long history in a variety of fields, including mathematics, kinematics and physical chemistry. We survey its history throughout these disciplines.

1. Introduction

Suppose we constructed a physical model of a graph using stiff rods for edges and connecting them at the vertices with freely moving joints. The graph could then be reconfigured, or moved, in any continuous manner from one configuration, or embedding, to a second configuration as long as no physical constraint of the structure were violated. In this case we could insist that no joint was broken, that the lengths of the bars remained fixed, and that no two bars intersected throughout the motion. We refer to such a flexible structure as a linkage. Alluding to this physical example, we will often refer to the edges of the graph as links and the vertices as joints. A possible reconfiguration of a linkage is illustrated in figure 1.

The geometry community has long considered the problem of determining if it is possible to continuously reconfigure a linkage from one configuration to another while maintaining the edge lengths and avoiding self-intersections during the motion. We can simplify this problem by instead asking if a linkage can be brought to some canonical form. Suppose a linkage can be brought from a configuration A to some canonical form C, and also can be brought from a configuration B to C.

Figure 1. Reconfiguring a linkage by a rotation at the grey joint.
Because all the motions are reversible, it follows that the linkage can be continuously reconfigured from $A$ to $B$. Thus the problem of unfolding linkages is born. In particular, this question is most often asked for the classes of linkages in figure 2:

- Given a chain linkage, can it be straightened?
- Given a cyclic (polygon) linkage, can it be convexified (brought to a convex position)?
- Given a tree linkage, can it be flattened?

The answers, of course, depend on the motions allowed and the space in which the linkage lies. We will briefly examine the history surrounding these problems in the next four sections. We first survey the problem of unfolding linkages in the plane, focusing on work primarily by mathematicians. In section 4, we discuss results concerning linkages in three dimensions. In the last two sections, we detail related work by kinematic engineers and by physicists. Remarkably, physicists and mathematicians have independently duplicated a great deal of research in this area.

2. Mathematicians unfolding linkages in the plane

Augustin Cauchy appears to have been the first mathematician to consider the problem of straightening a chain linkage in the plane. As a lemma for his celebrated theorem regarding the rigidity of polyhedra, he attempted to prove the following in 1813.

**Lemma 1.** (due to Cauchy [16]) *If some joints of a convex chain are opened, the distance between the endpoints increases and the chain remains convex.*

We consider a joint to be opening if its angle approaches $\pi$, in other words, if the two edges incident to it approach a straight angle. Such a motion is illustrated in figure 3, where the gray vertices have been opened.

Cauchy’s original proof was incorrect, although its flaw was unnoticed until 1934 when Steinitz and Rademacher [70] published a correction. Lyusternik [43] independently published a similar proof in 1966, and in the following year, Zaremba [65] presented a proof which was so short and elegant that Aigner and Ziegler [3] decided to include it in their collection, *Proofs from THE BOOK.* In fact, these last two proofs also hold for linkages on the sphere. As a lemma in a paper on intersections of planes with polytopes, O’Rourke [54] has since further generalized Cauchy’s lemma to include specific non-convex configurations. The survey paper by Toussaint [76] details Cauchy’s proof and a short history of the problem.

In the computer science community, reconfiguration of chain linkages first appeared in the context of modeling the possible motions of robot arms. John Hopcroft, Deborah Joseph, and Sue Whitesides [34] considered the case where the
linkage is permitted to self-intersect, since the limbs of a robot arm can pass under one another. They designed algorithms for deciding whether two configurations are mutually accessible when the chain is contained inside some convex polygon, and further proved that determining whether a chain could be flattened onto a line segment of specified length, a problem known as Ruler Folding (illustrated in figure 4), is NP-complete.

The question of whether a planar chain linkage can be straightened or a planar cyclic linkage can be convexified, without self-intersections, has arisen several times since. Robert Connelly, Erik Demaine, and Günter Rote [19] researched the origins of the problem and found that it was independently posed by Stephen Schanuel in the early 1970’s, by Ulf Grenander [29] in 1987, by William Lenhart and Sue Whitesides [41, 42, 85] in 1991, and by Joseph Mitchell in 1992. Grenander, Chow, and Keenan [30] also considered a cyclic linkage whose angles are fixed but whose lengths are variable, and proved that any embedding could be reconfigured into any other.

Biedl et al. [10] demonstrated that not all tree linkages can be unfolded by exhibiting the tree in figure 5. This tree is said to be locked in position since it is impossible to unfold. Nevertheless, the question for chain linkages and cyclic linkages received a flurry of interest in the computational geometry community and remained open for several years. Its difficulty led researchers to consider special cases. In 1998, Everett, Lazard, Robbins, Schröder, and Whitesides [25] showed that all star-shaped polygons can be convexified, and in the following year Biedl, Demaine, Lazard, Robbins, and Soss [11] provided an algorithm to convexify monotone polygons.

Attempts were made to construct chains that were believed to be locked [51], but unfolding motions were found for each example. The difficulty of these folding problems was further illustrated by Arkin, Fekete, Mitchell, and Skiena [5], who demonstrated the intractability of folding problems which stem from wire bending and sheet metal folding, and again recently by Arkin, Bender, Demaine, Demaine, Mitchell, Sethia, and Skiena [4].

Using tools in rigidity theory, Connelly, Demaine, and Rote [19] were finally able to prove in January, 2000, that chains and cycles can be unfolded. However,
describing their motions involves integrating vector fields and thus is quite complicated. Several months later, Streinu [73] discovered a simpler method which computes $O(n^2)$ motions to convexify a polygon, although the complexity of computing each motion is still unclear.

A general algorithm for motion planning, which can determine if an object can be brought from a starting configuration to a target configuration, was developed in 1983 by Schwartz and Sharir [66], but the complexity is doubly exponential in the degrees of freedom. In the case of linkages, this is at least as large as the number of joints. This result complemented the proof by Reif [55] in 1979 that deciding if an arbitrary hinged object (which could include polyhedral segments) could be moved from a starting configuration to a target is PSPACE-complete. Clearly more specialized algorithms are necessary in the case of simple linkages.

3. Mathematicians convexifying polygons with flips and flipturns

In 1933, Paul Erdős [24] posed the following problem, illustrated in figure 6. Given a nonconvex simple polygon, consider its convex hull. Subtracting the polygon from its convex hull yields several polygons called pockets (shaded). Reflect each of these pockets across its lid, that is, the edge it shares with the convex hull. We call this operation a flip. Prove that after a finite number of flips the polygon will be convex.

The first of many proofs of Erdős' conjecture was published four years later by Béla Nagy [52]. Nagy showed that flipping all pockets simultaneously might lead to a nonsimple polygon, as in figure 7, and modified the problem so that only one pocket is flipped at a time.

This problem has since been independently discovered and solved by several mathematicians: in 1957 by Reshetnyak [58] and by Yusupov [86]; in 1959 by Kazarinoff and Bing [12, 36, 37]; in 1973 by Joss and Shannon [31]; in 1981 by Kaluza [35]; in 1993 by Wegner [82]; and in 1999 by Biedl et al. [9]. Grünbaum and Zaks [32, 33] in 1998 and Toussaint [77] in 1999 extended these results to hold for crossing polygons (where a flip is allowed if it does not introduce new
Figure 6. Flipping all pockets of a polygon.

Figure 7. Béla Nagy’s polygon where flipping all pockets leads to nonsimplicity.

Figure 8. Joss and Shannon’s quadrilateral which can require arbitrarily many flips.

self-intersections). The papers by Grünbaum [31] and by Toussaint [77] provide a more detailed account of the history of the problem and its solutions.

Joss and Shannon [31] demonstrated that although finitely many flips suffice for convexification, the number required cannot be bounded by a function of the number of edges in the polygon. They presented the quadrilateral in figure 8. Note that there is only one flip possible at any one time since there is at most one reflex vertex. For any integer $k$, one can make the smallest edge tiny enough as to require at least $k$ flips to convexify the polygon. Wegner [82] and Biedl et al. [9] also discovered the same quadrilateral.

Wegner [82] also posed the inverse problem. Select a line that passes through the polygon in exactly two vertices, and reflect one of the two subchains across this line. If the resulting polygon does not self-intersect, the operation is called a deflation. This is the inverse of the flip operation, since the reflected subchain is now a pocket of the convex hull of the resulting polygon and the original line is its lid. Alluding to the concept that convex polygons do not admit flips, Wegner defined a deflated polygon as a polygon that does not admit any deflations. He conjectured that all polygons would be deflated after finitely many deflations. Figure 9 illustrates the deflation operation, and figure 10 illustrates a polygon which is deflated. A counterexample to Wegner’s conjecture was demonstrated by Fevens, Hernández, Mesa, Morin, Soss, and Toussaint [27].
Joss and Shannon [31] also considered a variant of the flip operation. Instead of reflecting a pocket, one can also rotate the pocket by 180 degrees about the center of its lid, as illustrated in figure 11. We call this operation a flipturn. Note that after a flipturn, each edge is rotated by 180 degrees. Unlike in a flip, each edge retains its original slope. If we consider each edge as a vector, a flipturn has the effect of permuting the order of the edges. Since each flipturn increases the area of the polygon, we will never visit the same permutation twice. Therefore the polygon will be convex after at most $(n-1)!$ flipturns, where $n$ is the number of edges of the polygon. Joss and Shannon conjectured that $n^2/4$ flipturns would suffice. Biedl [8] showed in 2000 that for some polygons, one can find a sequence of $\Omega(n^2)$ flipturns which convexify it (although shorter sequences exist for her examples). The same year, Ahn et al. [1] proved that a polygon is convex after any sequence of $n(n-s)/(2-s)$ flipturns, where $s$ is the number of distinct slopes of the edges. For arbitrary polygons, this value is $(n^2 - 3n)/2$. They consider only nondegenerate polygons; that is, polygons with no pockets determined by collinear edges. Aichholzer et al. [2] removed this condition and proved that there always exists a convexifying sequence of at most $5(n-4)/6$ flipturns for orthogonal polygons.

4. Mathematicians unfolding linkages in three dimensions

The geometry of three-space is quite different from that of two-space, and indeed not all three-dimensional chains can be straightened. Cantarella and Johnston [14] and Biedl et al. [9] independently proved that the chain of figure 12 is locked. This chain is often referred to as knitting needles due to its resemblance to the tools of the same name. If the first and last links are long enough, the endpoints
cannot be brought near the other four joints. One can then prove that the linkage behaves much like a trefoil knot.

In the case of three-dimensional polygons, the goal is to place the polygon into a planar convex configuration. Clearly a knotted polygon cannot be convexified, but Biedl et al. [9] demonstrated a class of locked unknots by joining two knitting needles to form the polygon of figure 13. Cantarella and Johnston [14] also proved that there exist locked unknots and presented the class of locked unknotted hexagons illustrated in figure 14. They conjectured that the configuration space of unknotted hexagons had three classes: the unknot and the left-handed and right-handed versions of their locked polygon. Toussaint [78] discovered the hexagon illustrated in figure 15, bringing the conjectured number of classes to five.

Biedl et al. [9] have shown that planar polygons can be convexified by motions in three-space. They rediscovered the flip operation with the idea of rotating a pocket in three-space about its lid, illustrated in figure 16. Realizing that a polygon may require an unbounded number of flips with respect to its number of edges, they described a linear-time algorithm to convexify a planar polygon using more complex motions. Several months later, Aronov, Goodman, and Pollack [6] simplified their
Figure 14. The locked hexagon of Cantarella and Johnston.

Figure 15. The locked hexagon of Toussaint.

Figure 16. Performing a flip in three-space.

algorithm and demonstrated that the result could be generalized to include crossing polygons. (They considered a reconfiguration of a crossing polygon to be valid if it introduces no new crossings during the motion.)

We can generalize the flip operation by choosing two points of a polygon and rotating one of the subchains about the line through these two points. This operation was first used in 1945 by Choquet [17] in an application known as curve stretching. A curve $c'$ is a stretched version of $c$ if for every two points on $c$, the arc length between them is maintained and the Euclidean distance between them is either maintained or increased. One can imagine $c$ as a rope, and $c'$ as a position of the same rope, but "spread apart." A polygon which is convexified by flips is therefore a stretched version of the original polygon, since each flip either maintains or increases the pairwise distances between points in the polygon.

Sallee [64] proved that for a three-dimensional curve, there exists a stretched version which is planar and convex. Robertson and Wegner have also studied this operation (which they refer to as inflation) in the plane [61, 62, 82], and Wegner has explored stretching curves on the sphere [83, 84]. Millett [50] has also used similar motions in three dimensions to convexify knots. Several such studies (including those of Sallee and Millett) allowed the polygon to self-intersect during the motion.

In 2000, Calvo, Krizanc, Morin, Soes, and Toussaint [13] demonstrated a class of polygons which can be convexified in three-dimensions without crossings: polygons which admit a simple orthogonal projection. The algorithm works by first convexifying the planar projection by keeping the height of each coordinate fixed, and then with a linear-time algorithm to convexify a polygon with a convex orthogonal projection.
The following year, Demaine, Langerman, and O'Rourke [22] studied multiple linkages which could interlock with one another. They describe configurations of pairs of short three-dimensional chains such that the two chains cannot be separated. In the same year, Soss [67] independently discovered short chains which are interlocked under the restriction that the angles between edges are fixed.

In 1999, C ocean and O'Rourke [18] proved that all chains, polygons, and trees can be unfolded in dimensions four and higher.

5. Kinematic engineers unfolding linkages

The study of kinematic linkages appears to have been first mathematically codified in 1874 by the engineer Franz Reuleaux in his work Theoretische Kinematik [59]. Kinematics is the study of motion, and in the context of linkages, is the study of how certain joints move in concert with the motion of other joints.

Reuleaux defined a mechanism as “a closed kinematic chain, of which one link is thus made stationary [59, p. 47].” A major component of his work is in describing the mathematical relationships between the curves drawn by each joint as one moves the linkage in a specified fashion. Even the simple four-bar linkage [60] shown in figure 17 was the subject of intense study in the nineteenth century. (This linkage was often called a three-bar linkage since there was no consensus at that time on whether to count the stationary edge as part of the structure.) Here one can easily compute the path traveled by the points p and q as the angle $\theta$ is altered. Toussaint [79] provides a short history of four-bar linkages and presents several new proofs on some of their properties.

Perhaps the most famous early linkage is Peaucellier's Inversor of figure 18, developed in 1864 by the French engineer of the same name. If the points a and b are fixed in space such that the distance between a and b is the same as that between b and p, then as p moves along the circle, q moves along the line. Thus the linkage inverts radial motion to linear motion and vice versa.

The paper by Farouki [26] and the book by McCarthy [48] provide an in-depth historical background and expand on the connections between geometry and kinematics.

6. Polymer physicists unfolding linkages

Flexible polymer molecules are long chains and cycles of atoms, often called monomers. These chains are held together by chemical bonds, about which a certain amount of rotation is generally possible. Physicists have long attempted to simulate a sampling of random polymer configurations by treating the polymer as a linkage.
In 1934, Werner von Kuhn [38] proposed a model in which a long chain of unit length edges is allowed to assume any configuration, possibly self-intersecting. To obtain a random configuration, he chose each link to assume any orientation with uniform probability. Unfortunately, Kuhn's model did not perform very well in predicting polymer properties. Benoît [7] and Taylor [75] independently corrected his model so that successive elements of the polymer chain were not randomly oriented. They recognized that not only should the lengths of the links relate to the lengths of the monomers, but the angles at the joints of the linkage should relate to the bond angles between monomers. Rather than allowing a link to assume any orientation, it was bound to those that preserve the bond angle with the previous link. This resulted in a range illustrated in figure 19.

In 1943, Kuhn and Kuhn [39] recognized that the normal random-walk model was severely deficient, since it implied that the atoms could overlap. They realized that self-intersections, or even positions in which atoms are very close together, should not be allowed. Eighteen years later, Sykes [74] proposed modeling polymers as self-avoiding walks on a cubic lattice and began his work by counting the number of configurations for small numbers of links. His attempts to compute this number for large walks did not meet much success [23], and even today this well-known problem remains unsolved and a focus of mathematical research [46]. Since the number of configurations grows exponentially with the length of the linkage, it became evident that Monte Carlo methods would prove useful in sampling configurations.

Early methods of randomly generating self-avoiding walks took quite long to finish [63, 81]. The difficulty was due to the approach; a walk was generated
by choosing a random direction at each step. If a self-intersection arose, the entire configuration was discarded and the process begun anew. Note that it is not correct to simply remove the self-intersecting step and choose a new direction, as this would erroneously increase the likelihood of the portion of the configuration attained up to that point.

In 1969, Peter Verdier [80] proposed a faster method of computing random walks. Rather than randomly walking on the lattice and hoping for a sequence of non-intersecting steps, he began with a valid walk and performed a series of random reconfigurations. Any single joint could be moved to a vacant lattice point so long as the edges remained unit length and the chain did not self-intersect. These conditions imply that the only motions possible are those illustrated in figure 20.

An interesting problem arose concerning this new method of generating walks. It is not obvious that there exists a sequence of reconfigurations by which any configuration can be brought to any other, and thus it is not obvious whether all valid configurations have a nonzero probability of being generated. In the physics community, if all configurations are attainable, the model is said to be ergodic. Not surprisingly, Verdier's model lacks ergodicity. In the plane, the chain in figure 21 is locked; in three-dimensions, the chains in figure 22 are locked. Therefore neither could be generated by Verdier's simulations unless it was selected as the starting configuration.

The same year as Verdier, Moti Lal [40] proposed a different operation for randomly reconfiguring a walk which yields an ergodic model. His method was to select an edge at random, defining two subchains, then to reflect one of these subchains across the line through the edge, shown in figure 23. This operation has been since called a pivot throughout the literature. If the resulting chain intersected itself, the pivot was rejected and a new edge was randomly chosen. Otherwise, the procedure was repeated from the new configuration. Lal defined his model on the hexagonal lattice, although the pivot operation works just as well in a variety of
spaces. In order to prove that the model is ergodic, he demonstrated that any planar chain could be “straightened” (alternating right and left turns of 120°) by a series of pivots.

This algorithm was independently reinvented in 1976 by Olaj and Pelinka [53] and again nine years later by MacDonald et al. [44]. Its efficiency has led it to be a major focus of study during the past fifteen years [15, 44, 46, 47, 56, 57], although expanded by an operation. In addition to Lal’s original pivot, it is customary to select two points on the chain and cut out the portion between them. The portion is then replaced, either reflected about a line through its cut points or rotated by 180 degrees, as in figure 24. As long as the resulting chain does not self-intersect and lies on the lattice, the pivot is accepted. This new operation easily generalizes to polygons and is remarkably similar to flips and flipturns.

In order to demonstrate ergodicity, Madras and Sokal [47] proved that in any dimension all lattice chains can be straightened by pivots. Madras, Orlitsky, and Shepp [45] extended this result to prove that any $d$-dimensional lattice polygon can be convexified with $O(n^{6(d-2)})$ pivots for $d \geq 3$.

Many in the physics community were concerned that the restriction of the model to the lattice was too far removed from reality. So called off-lattice or continuum models were constructed. John Curro [20, 21] extended Lal’s idea of a pivot to continuous space by selecting an edge of the chain and rotating one portion of the chain about a line through the edge. As in Lal’s model, the joint-angle at the edge stays fixed, and if the result of the motion is not self-intersecting, it is accepted. This motion is often referred to as a dihedral rotation, illustrated in figure 25. This
model has been used with success by several researchers \cite{20, 21, 28, 49, 71, 72} since its introduction, in polymer physics and also in chemistry, where preserving the bond angles of molecules (or the joint-angles of linkages) is vital.

In 2000, Soss and Toussaint \cite{69} presented efficient algorithms for determining whether dihedral rotations about edges of three-dimensional chains result in self-intersecting (and thus illegal) conformations, and presented an $\Omega(n \log n)$ lower bound on the computation time required to solve the problem. The following year, Soss, Erickson, and Overmars \cite{68} proved that even with the aid of an arbitrary amount of preprocessing time of the chain’s structure, this lower bound of time per rotation could only be improved slightly if at all. Further underlining the computational difficulties of these problems, in 2001 Soss \cite{67} demonstrated the intractability of deciding whether it is possible to move a chain between two configurations by dihedral rotations without causing the chain to self-intersect.
References


Chemical Computing Group, 1010 Sherbrooke Street West, Montreal, QC, Canada H3A 2R7.

E-mail address: soss@chemcomp.com

School of Computer Science, McGill University, 3480 University Street #318, Montreal, QC, Canada H3A 2A7.

E-mail address: godfried@cs.mcgill.ca