Locked and Unlocked Chains of Planar Shapes

Robert Connelly* Erik D. Demaine† Martin L. Demaine* Sándor Fekete‡
Stefan Langerman§ Joseph S. B. Mitchell¶ Ares Ribó‖ Günter Rote‖

August 3, 2004

Dedicated to Godfried Toussaint on the occasion of his 60th birthday

Abstract

If we attach to each bar in a polygonal chain a rigid shape whose inward normals all hit the attached bar, and the resulting hinged chain of shapes does not overlap itself, then this chain can always be unfolded into a straight or convex configuration (with respect to the underlying polygonal chain). On the other hand, we construct locked hinged chains of isosceles triangles with any desired apex angle $< 90°$, which is the threshold beyond which the inward-normal property no longer holds.

1 Introduction

We consider chains of nonoverlapping rigid planar shapes that are hinged together sequentially at rotatable joints. A chain forms either a ring where each shape is connected to two neighboring shapes, or an open chain in which the first and the last shape have only one neighbor. Such a chain can be *folded* by rotating the hinges in such a way that no two shapes ever properly overlap (though we allow shapes to touch along their boundaries).

Our goal is to understand when such a chain of shapes is *locked* in the sense that it cannot be folded from one configuration to another. Our work generalizes the study of locked and unlocked polygonal chains, which is the special case in which each shape is just a line segment. It is not surprising that locked chains of shapes exist, in contrast to planar polygonal chains. Nonetheless, we show in this paper how many of the tools built for understanding when polygonal linkages lock in the plane can establish both locked and unlocked chains of shapes.

*Department of Mathematics, Cornell University, Ithaca, NY 14853, USA. connelly@math.cornell.edu
†MIT Computer Science and Artificial Intelligence Laboratory, 32 Vassar St., Cambridge, MA 02139, USA. {edemaine, mdemaine}@mit.edu
‡Abteilung für Optimierung, Technische Universität Braunschweig, 38106 Braunschweig, Germany. s.fekete@tu-bs.de
§Chercheur qualifié du FNRS, Université Libre de Bruxelles, Département d'informatique, ULB CP212, Belgium. Stefan.Langerman@ulb.ac.be
¶Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600, USA. jsbm@ams.sunysb.edu
‖Institut für Informatik, Freie Universität Berlin, Takustraße 9, D-14195 Berlin, Germany. {ribo, rote}@inf.fu-berlin.de

1.1 Adorned Chains

Another way to look at a chain of shapes is to consider the polygonal chain of line segments connecting successive joints. (For an open chain, there is some freedom in choosing the endpoints for the first and the last bar.) On the one hand, these line segments can be viewed as *bars* that move rigidly with the shapes to which they belong. On the other hand, the shapes can be viewed as "adornments" to the bars of an underlying polygonal chain. This view leads to the concept of an "adorned polygonal chain" which we now proceed to define more precisely.

An *adornment* is a simply connected compact region in the plane, called the *shape*, together with a line segment ab connecting two boundary points, called the *base*. There are two *boundary arcs* from a to b which enclose the shape, called *sides*. We require that the base is contained in the shape. (In other words, the base must be a chord of the shape.) An adornment is *convex* if its shape is convex.

An *adorned polygonal chain* is a set of nonoverlapping adornments whose bases form a polygonal chain. We permit the shapes to touch on their boundary and to slide along each other. However, we require the underlying polygonal chain to be strictly simple: a hinge is not allowed to touch a point on the base of a shape to which it is not directly attached.

The viewpoint of a chain of shapes as an adorned polygonal chain is useful for two reasons. First, we can more easily talk about the kind of shapes, and their relation to the locations of the incident hinges, in a family of chains: this information is captured by the adornments. Second, the underlying polygonal chain provides a mechanism for folding the chain of shapes, as well as a natural *unfolding* goal: straighten the underlying open chain or convexify the underlying closed chain. Indeed, we show that, in some cases, unfolding motions of the polygonal chain induce valid unfolding motions of the chain of shapes.

1.2 Slender Adornments

The class of adornments that we treat have additional regularity requirements on their boundary: We require that each boundary arc is differentiable except at a countable set of points, and that one-sided (left and right) tangents exist everywhere. This class includes all convex shapes and also all piecewise smooth shapes with countably many smooth pieces.

At every point x of the boundary curve, there is a left tangent and a right tangent. (At a smooth point x, these two tangents coincide.) The *inward normals* are the rays starting at x, perpendicular to these tangents, pointing to the side on which the interior of the shape lies.

We call an adornment *slender* if every inward normal of the shape intersects the base (possibly at the base's endpoints). The main unfoldability result of this paper is that every adorned polygonal chain consisting of (possibly many different) slender convex adornments can be unfolded, and thus no such adorned polygonal chain is locked. This result is based on the existence of expansive motions for polygonal chains [CDR03].

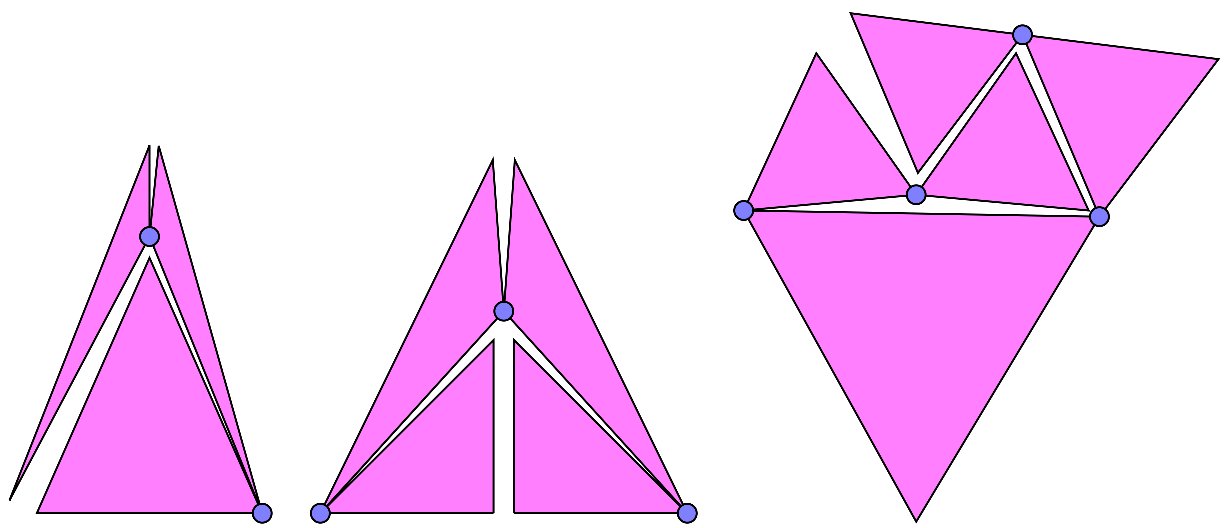
An example is a square with the diagonal as the base. Thus, as a consequence of our main theorem, every sequence of squares hinged together at diagonally opposite points can be unfolded, provided that no two hinges coincide. (The last assumption is probably only a technicality which can be removed by special arguments.) We do not know whether chains of squares connected at adjacent endpoints can be locked. Slender shapes can be nonconvex, but currently our proof does not extend to nonconvex adornments.

A boundary point of the shape has either one or two inward normals, depending on whether it is a smooth point or a "vertex". At a vertex, we can consider the whole range of inward normals as the tangent direction makes a continuous transition from the left tangent to the right tangent.

These normals sweep over an angle of at most 180°; thus, if the two extreme normals intersect the base, all intermediate "normals" will intersect the base as well (and vice versa).

1.3 Locked Adorned Chains

One of the earliest examples of a locked adorned polygonal chain was discovered by M. Demaine at a Barbados workshop in January 1998; it consists of just three triangles, shown in Figure 1(a). Figure 1(b) shows a locked variation on this example. In both examples of these examples, two of the triangles are obtuse. Figure 1(c) shows another example of a locked adorned polygonal chain in which each adornment is an acute—indeed, equilateral—triangle. However, the triangles are of substantially different size.



(a) A locked chain of three triangles.

(b) A locked chain of four triangles.

(c) A locked chain of equilateral triangles of different sizes.

Figure 1: Some examples of locked chains of triangles.

The main lockability result of this paper is a locked chain of equilateral triangles of equal size, shown in Figure 5. If the initial configuration is not allowed to have touching triangles, then the triangles have size arbitrarily close to equal. Furthermore, for any $\theta < 90°$, the locked chain can be modified to consist entirely of adornments whose shape is an isosceles triangle with apex angle θ and whose base is the nonequal side of the triangle, as shown in Figure 6. Among all such adornments, $\theta < 90°$ is precisely when the adornment is not slender. Therefore, we provide a tight characterization of when a chain of isosceles triangles can lock when the base of each triangle is its nonequal side.

2 Expansive Motions

Unfoldability of polygonal chains is established in [CDR03] using "expansive" motions. A motion of a linkage is *expansive* if the distance between every pair of vertices is a nondecreasing function over time, and furthermore, at all times, some vertex-vertex distance strictly increases.

Theorem 1 ([CDR03, Theorem 1]) *Every polygonal chain has an expansive motion to a straight or convex configuration.*

Our proofs of unfoldability of chains of slender adornments use this theorem, as well as the following stronger property of expansive motions:

Lemma 2 ([CDR03, Corollary 1]) *An expansive motion of a linkage never decreases the distance between any two points on the linkage, where each of the two points can be chosen to be a vertex or a fixed position along a bar.*

3 Slender Adornments

In this section we prove some basic results about when adornments are slender.

An equivalent characterization of slender adornments concerns the distance between an endpoint of the base and a point along the boundary of the shape:

Lemma 3 *An adornment is slender if and only if, for a point moving on one side of the shape, the distance to each endpoint of the base changes monotonically.*

Proof: Let $x(t)$ be a clockwise arc-length parameterization of a side, starting at one base endpoint a and ending at the other base endpoint b. By the assumptions, the distance function $f(t) = \|x(t) - a\|$ is smooth except at a countable set of points. The directions of the right and left tangents at $x(t)$ determine the sign of the left and right derivatives $f'_-(t)$ and $f'_+(t)$, via a straightforward geometric relation (see Figure 2): $f'_+(t) \geq 0$ if and only if the inward normal to the right tangent at x points into the half-space to the right of the line ax, and similarly for $f'_-(t)$.

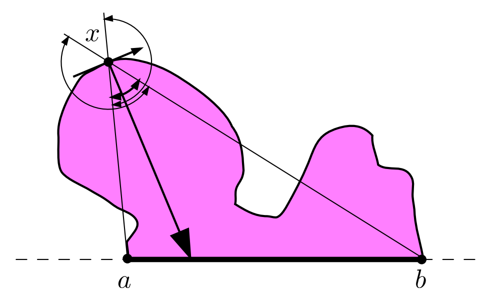


Figure 2: The monotone-distance property for slender shapes

Thus, monotonicity of f is equivalent to the condition that the inward normals lie in the 180° range of directions on one side of the line xa. Similarly, monotonicity of the distance from b means that the inward normals point into the half-space to the left of the line bx. Together, this is equivalent to the condition that the inward normals intersect the base segment ab. □

This lemma allows us to determine the largest (in the partial order defined by set containment) possible shape of a slender adornment:

Corollary 4 *The shape of a slender adornment is contained in the crescent enclosed by the two circles centered at one endpoint of the base and passing through the other endpoint of the base.*

Proof: Refer to Figure 3. Let x and y be the two endpoints of the adornment's base. Let p be a point moving from x to y along one of the adornment's sides. At the beginning, $p = x$, the distance $\|x - p\|$ is 0. At the end, $p = y$, the distance $\|x - p\|$ is $\|x - y\|$. By Lemma 3, at all times in between, $\|x - p\|$ must be between 0 and $\|x - y\|$. Therefore, all points p along the boundary of the adornment's shape must be inside or on the circle centered at x and passing through y. By

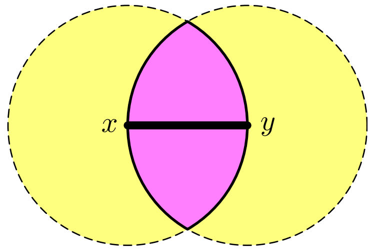


Figure 3: The largest slender adornment with a given base is a crescent.

symmetry, all such points p must also be inside or on the circle centered at y and passing through x. Therefore all such points p must be inside or on the boundary of the crescent enclosed by these two circles. □

4 Slender Adornments Cannot Lock

Now we can prove our main unfoldability result:

Theorem 5 *A strictly simple polygonal chain adorned by slender convex adornments can always be straightened or convexified.*

Proof: We claim that any expansive motion of the underlying polygonal chain straightens or convexifies the adorned chain without self-intersection.

For the adorned chain to self-intersect, a boundary point p_A of one adornment A must enter the interior of another adornment B. Consider the time t just before such an event might occur, at which p_A touches a boundary point p_B of B. See Figure 4. We refer to time t as the "current time". Because A and B are convex, there is a common tangent ℓ passing through p that separates the interiors of A and B. Because A and B are slender, their inward normals at p_A and p_B perpendicular to ℓ meet their bases, say at points a and b. Rotate the diagram so that ℓ is vertical and A is left of B; thus a, p_A, p_B, and b are oriented left to right along a horizontal line.

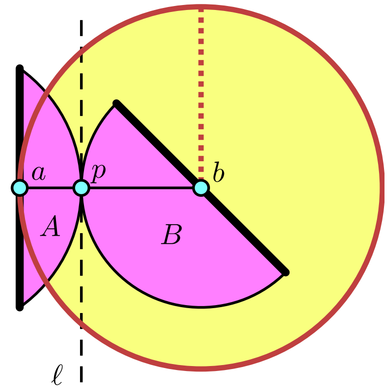


Figure 4: Two slender convex adornments A and B touching at a point p.

Now consider the motion of A relative to B (pinning B to the plane), with points p_A and a tracking the motion of A, and with points p_B and b remaining fixed. Lemma 2 implies that the distance between a and b cannot decrease during the (expansive) motion. By the triangle inequality, this distance $\|a - b\|$ is always at most the sum of the two distances $\|a - p_A\|$ and $\|p_A - b\|$. At the current time, this inequality holds with equality. During the motion, the first distance $\|a - p_A\|$ remains fixed because both a and p_A move relative to A, but the second distance $\|p_A - b\|$ may change. But because $\|a - b\|$ cannot decrease, $\|p_A - b\|$ cannot decrease. Put another way, p_A cannot enter the interior of the circle centered at b and passing through p_B, because then a would be forced to enter the interior of the circle centered at b and passing through the current position of a, because a has a fixed distance to p_A.

This argument shows that p_A cannot enter the interior of B to the first order, but we have not yet forbidden a second-order entering motion that proceeds vertically to the first order. Also note that the argument so far does not rely on the convexity of A or B.

In the special case that the base of B is horizontal, the inward normal at p_B in fact intersects the entire base of B. The argument above applies for every point b in the intersection, i.e., for every point b along B's base. We choose b to be the right endpoint of B's base. By Corollary 4, B is contained in the circle centered at b and passing through p_B (the left endpoint of B's base). Because p_A cannot enter the interior of this circle, p_A cannot enter the interior of B.

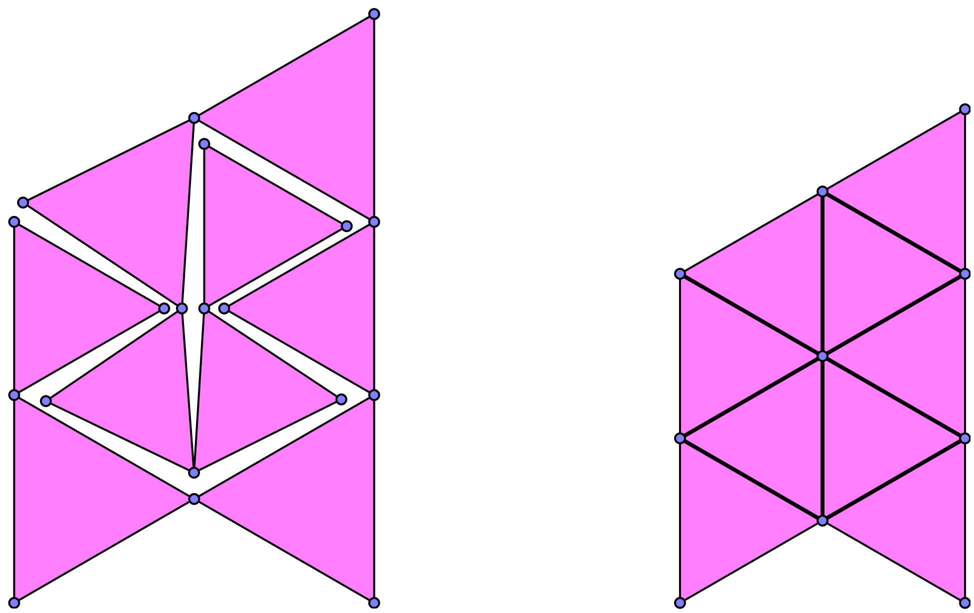
Now suppose that the base of B is not horizontal. In this case, b is uniquely defined as the intersection of a horizontal line with a nonhorizontal segment. There are two cases depending on whether b is an endpoint of B's base.

If b is not an endpoint of B's base, then there is a point b_+ along B's base just above b and a point b_- along B's base just below b. By Lemma 2, a cannot get closer to b_+ or b_-. In other words, a cannot enter the interior of the circle C_+ centered at b_+ and passing through the current position of a, nor can it enter the interior of the circle C_- centered at b_- and passing through the current position of a. Now the center b_+ of C_+ is above the horizontal line through p_B, b, and the current position of a. Thus, locally at the current position of a, circle C_+ proceeds up and left to the first order. Similarly, locally at the current position of a, C_- proceeds down and left to the first order. Therefore, locally, a must move to the left. The fixed distance $\|a - p_A\|$, currently horizontal, implies that p_A also must move to the left, locally. Hence, p_A cannot immediately move right of ℓ, so by convexity of B, p_A cannot immediately enter the interior of B.

Finally, suppose that b is an endpoint of B's base. Assume by symmetry that b is the lower endpoint of B's base. As in the previous case, choose a point b_+ along B's base just above b; a cannot enter the interior of the circle C_+ centered at b_+ and passing through the current position of a; and, locally at the current position of a, circle C_+ proceeds up and left to the first order. Therefore if a moves up or horizontally, it must move left, in which case p_A cannot enter the interior of B. If a moves down, it cannot enter the interior of the circle centered at b and passing through the current position of a, and as argued above, p_A cannot enter the interior of the circle C centered at b and passing through p_B. By Lemma 3, the portion of B's boundary that connects p_B and base endpoint b (and stays on a single side of B) is contained in the circle C. By convexity of B, this portion is precisely the portion of B's boundary that is below the horizontal line through p_B and b. Therefore, if a moves down, it cannot enter the interior of B. $\square$

5 Locked Chain of Sharp Triangles

An isosceles triangle with an apex angle of $\geq 90°$ and with the nonequal side as the base is a slender adornment. By Theorem 5, any chain of such triangles can be straightened. In this section we show that this result is tight: for any isosceles triangle with an apex angle of $< 90°$ and with the nonequal side as a base, there is a chain of these triangles that cannot be straightened.



(a) Drawn loosely. Separations should be smaller than they appear.

(b) Drawn tightly, with no separation, as a self-touching configuration.

Figure 5: A locked chain of equilateral triangles.

Figure 5(a) shows the construction for equilateral triangles (of slightly different sizes). This figure is drawn with the pieces loosely separated, but the actual construction has arbitrarily small separations and arbitrarily closely approximates the self-touching geometry shown in Figure 5(b). Stretching the triangles in this self-touching geometry, as shown in Figure 6, defines our construction for any isosceles triangles with an opposite angle of any value less than 90°. In this case, however, our construction uses two different scalings of the same triangle.

5.1 Theory of Self-Touching Configurations

This view of the construction as a slightly separated version of a self-touching configuration allows us to apply the program developed in [CDR02] for proving a configuration locked. This theory allows us to study the rigidity of self-touching configurations, which is easier because vertices cannot move even slightly, and obtain a strong form of lockedness of non-self-touching perturbations drawn with sufficiently small (but positive) separations.

To state this relation precisely, we need some terminology from [CDR02]. Call a linkage configuration *rigid* if it cannot move at all. Define a *δ-perturbation* of a linkage configuration to be a repositioning of each vertex within distance δ of its original position, without regard to preserving edge lengths (better than $\pm 2\delta$), but consistent with the combinatorial information of which vertices are on which side of which bar. Call a linkage *locked within* ε if no motion that leaves some bar pinned to the plane moves any point by more than ε. Call a self-touching linkage configuration *strongly locked* if, for any desired $\varepsilon > 0$, there is a $\delta > 0$ such that all δ-perturbations are locked within ε. Thus, if a self-touching configuration is strongly locked, then the smaller we draw the

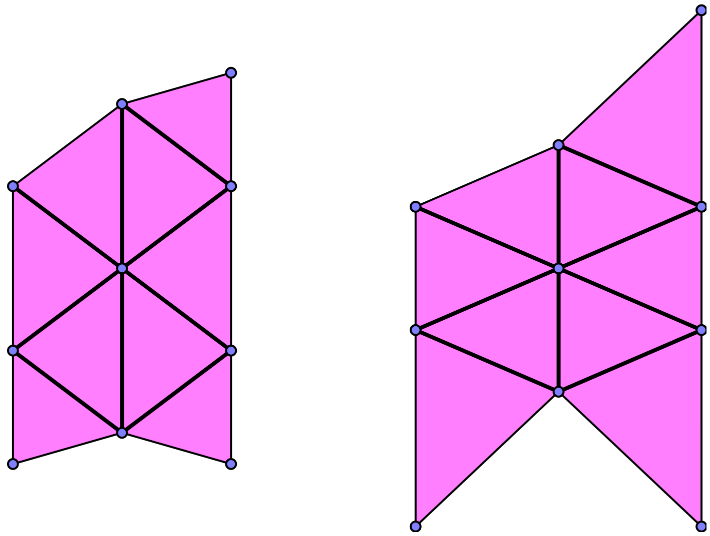


Figure 6: Variations on the self-touching configuration from Figure 5(b) to have any desired angle $< 90°$ opposite the base of each triangle.

separations in a non-self-touching perturbation, the less the configuration can move. In particular, if we choose ε small enough, the linkage must be locked in the standard sense of having a disconnected configuration space.

Theorem 6 ([CDR02, Theorem 8.1]) *If a self-touching linkage configuration is rigid, then it is strongly locked.*

Therefore, if we can prove that the self-touching configuration in Figure 5(b) (and its variations in Figure 6) are rigid, then sufficiently small perturbations along the lines shown in Figure 5(a) are rigid.

The theory of [CDR02] also provides tools for proving rigidity of a self-touching configuration. Specifically, we can study *infinitesimal motions* which just define the beginning of a motion to the first order. Call a configuration *infinitesimally rigid* if it has no infinitesimal motions.

Lemma 7 ([CDR02, Lemma 6.1]) *If a self-touching linkage configuration is infinitesimally rigid, then it is rigid.*

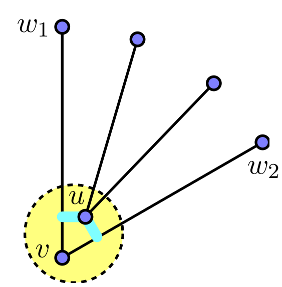


Figure 7: Two zero-length connections between vertices u and v.

A final tool we need from [CDR02] is for proving infinitesimal rigidity. For each vertex u wedged into a convex angle between two bars $\{v, w_1\}$ and $\{v, w_2\}$, we say that there are two *zero-length connections* between u and v, one perpendicular to each of the two bars $\{v, w_i\}$.[1] See Figure 7. These connections must increase to the first order because u must not cross the two bars $\{v, w_i\}$. In proving infinitesimal rigidity, we can choose to discard any zero-length connections we wish, because ignoring some of the noncrossing constraints only makes the configuration more flexible. Together, the bars and the zero-length connections are the *edges* of the configuration. Define a *stress* to be an assignment of real numbers (*stresses*) to edges such that, for each vertex v, the vectors with directions

[1] The definition of such connections in [CDR02] is more general, but this definition suffices for our purposes.

defined by the edges incident to v, and with magnitudes equal to the cor-
responding stresses, sum to the zero vector. We denote the stress on a bar $\{v, w\}$ by ω_{vw}, and we denote the stress on a zero-length connection between vertex u and vertex v perpendicular to $\{v, w\}$ by $\omega_{u,vw}$.

Lemma 8 ([CDR02, Lemma 7.2]) *If a self-touching configuration has a stress that is negative on every zero-length connection, and if the configuration is infinitesimally rigid when every zero-length connection is treated as a bar pinning two vertices together, then the self-touching configuration is infinitesimally rigid.*

5.2 Simplifying Rules

We introduce two rules that significantly restrict the allowable motions of the self-touching configuration of isosceles triangles.

Rule 1 *If a bar b is wedged against another bar b' of equal length, and the bars incident to b' form angles less than $90°$ on the same side as b, then any infinitesimal motion must keep b pinned against b'. See Figure 8.*

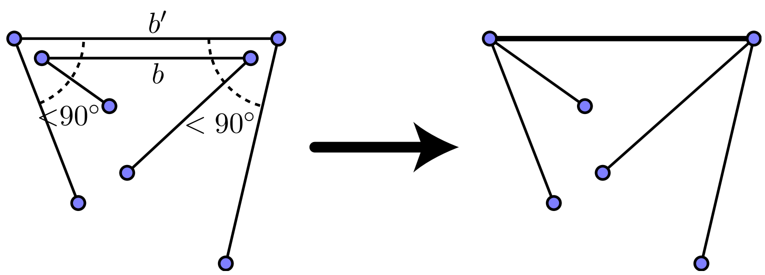


Figure 8: Rule 1 for simplifying self-touching configurations.

Proof: The noncrossing constraints at the endpoints of b and b' prevent b from moving relative to b' until the angles at the endpoints of b' open to $\geq 90°$, which can only happen after a positive amount of time. □

We can apply this rule to the region shown in Figure 9, resulting in a simpler linkage with the same infinitesimal behavior. Although the figure shows positive separations for visual clarity, we are in fact acting on the self-touching configuration of Figure 5(b).

Rule 2 *If a bar b is wedged against an incident bar b' whose other incident bar b'' forms a convex angle with b' surrounding b, then any infinitesimal motion must keep b pinned against b'. See Figure 10.*

Proof: The noncrossing constraints at the endpoint of b surrounded by the convex angle formed by b' and b'' prevent b from moving relative to b' until the convex angle opens to $\geq 90°$, which can only happen after a positive amount of time. □

We can apply this rule twice, as shown in Figure 11, to further simplify the linkage.

The final simplification comes from realizing that the central quadrangle gap between triangles is effectively a triangle because the right pair of edges are a rigid unit. Thus the gap forms a rigid linkage (though it is not infinitesimally rigid, because a horizontal movement of the central vertex would maintain distances to the first order), so we can treat it as part of a large rigid block. Figure 12 shows a simplified drawing of this self-touching configuration, which is rigid if and only if the original self-touching configuration is rigid.

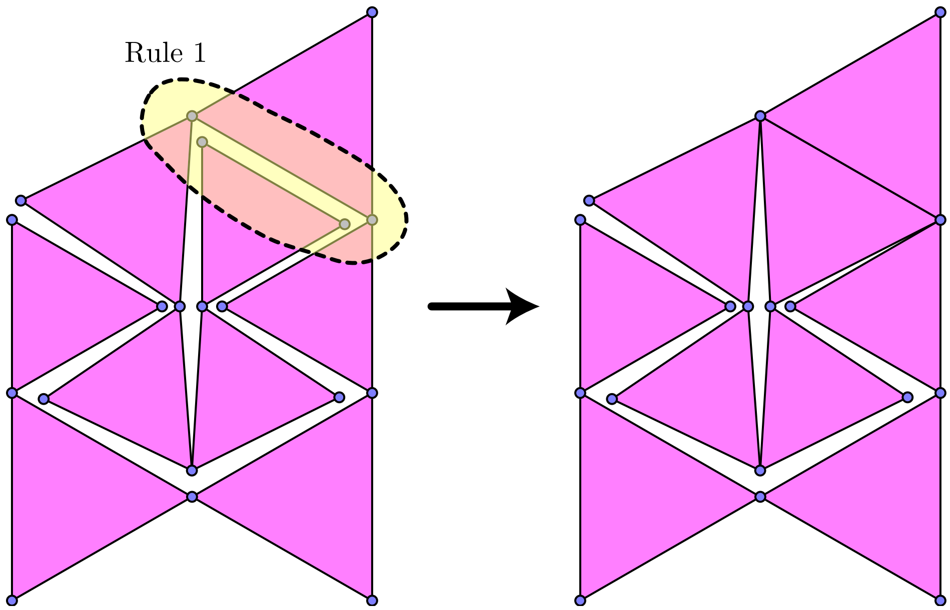


Figure 9: Applying Rule 1 to the chain of equilateral triangles from Figure 5.

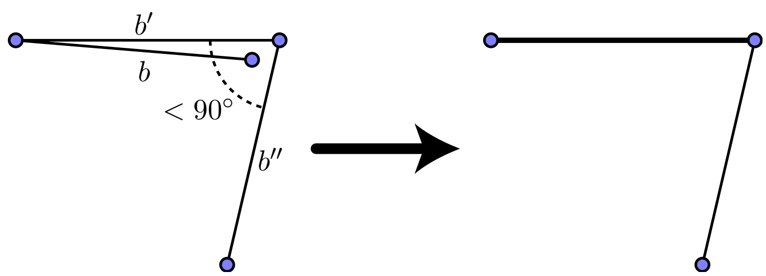


Figure 10: Rule 2 for simplifying self-touching configurations.

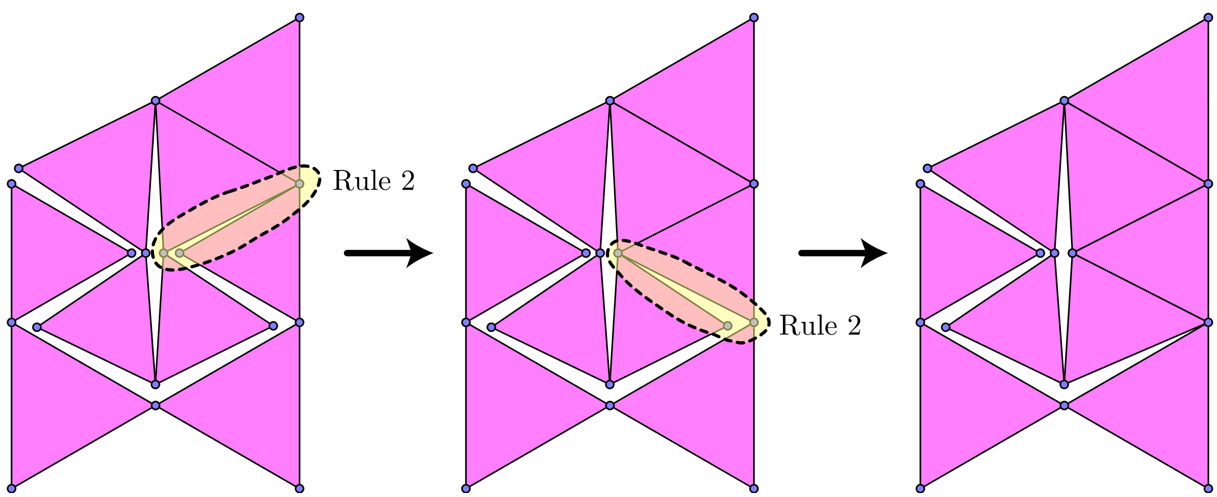


Figure 11: Applying Rule 2 twice to the configuration from Figure 9.

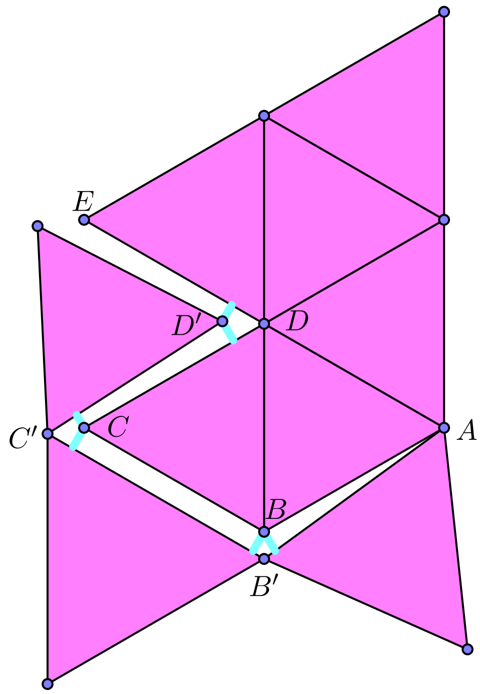


Figure 12: The simplified configuration from Figure 11.

5.3 Stress Argument

Finally we argue that the simplified configuration of Figure 12 is infinitesimally rigid using Lemma 8. The configuration is clearly infinitesimally rigid if B is pinned against B', C is pinned against C', and D is pinned against D'. It remains to construct a stress that is negative on all length-zero connections. The stress we construct is nonzero only on the edges connecting points with labels in Figure 12; we also set $\Omega_{AD} = 0$.

We start by assigning the stresses incident to A. We choose $\omega_{AB} < 0$ arbitrarily, and set $\omega_{AB'} := -\omega_{AB} > 0$. A is now in equilibrium because these stress directions are parallel.

We symmetrically assign $\omega_{BC} := \omega_{AB} < 0$ and $\omega_{B'C'} = \omega_{A'B'} > 0$. The resulting forces on B and B' are vertical. They can be balanced by an appropriate choice of the stresses $\omega_{B,B'A} = \omega_{B,B'C'} < 0$, which, taken together, also point in the vertical direction.

Vertex D' has exactly three incident stresses—$\omega_{C'D'}$, $\omega_{D',DC}$, and $\omega_{D',DE}$—which do not lie in a halfplane. Thus there is an equilibrium assignment to these stresses, unique up to scaling, and the stresses all have the same sign. Because zero-length connections must be negative, we are forced to make all three of these stresses negative. We also choose this scale factor to be substantially smaller than the stresses that have been assigned so far.

By assigning $\omega_{CD} = -\omega_{C'D'}$, we establish equilibrium at vertex D as well: the forces at D are the same as at D', only with reversed signs.

Vertex C feels two stresses assigned so far—$\omega_{CD} > 0$ and $\omega_{BC} < 0$. By the choice of scale factors, the latter force dominates, leaving us with a negative force in the direction close to CB, and two stresses $\omega_{C,C'B'}$ and $\omega_{C,C'D'}$ which can be used to balance this force. The three directions do not lie in a halfplane. Therefore $\omega_{C,C'B'}$ and $\omega_{C,C'D'}$ can be assigned negative stresses.

Finally, vertex C' is also in equilibrium because $\omega_{B'C'} = -\omega_{BC}$, $\omega_{C'D'} = -\omega_{CD}$, and the stress from the zero-length connections are the same as for C but in the opposite direction.

In summary, we have shown the existence of a stress that is zero on all zero-length connections. By Lemma 8, the self-touching configuration is infinitesimally rigid, so by Lemma 7, the configuration is rigid. By the simplification arguments above, the original self-touching configuration is also rigid. By Theorem 6, the original self-touching configuration is strongly locked, so sufficiently perturbations are locked.

We remark that an argument similar as above, using an assignment of stresses, can also be used for proving Rules 1 and 2, with an appropriate modification of Lemma 8; however, the direct argument that we have given is simpler.

The argument relied on the isosceles triangles having an apex angle of $< 90°$ (but no more) in order to guarantee that particular triples of stress directions are or are not in a halfplane. It also relies on the symmetry of the configuration through a vertical line (excluding the triangle in the upper right). Thus the argument generalizes to all isosceles triangles sharper than $90°$.

6 Conclusion

We have seen that the theory of self-touching linkage configurations and their stresses play a key role in the analysis of locked chains of shapes (Section 5). However, it remains open whether chains of slender convex adornments in which the underlying polygonal chain is self-touching can lock. The barrier is that it remains open whether self-touching configurations of polygonal chains can always be straightened or convexified. The Maxwell-Cremona correspondence between stresses and liftings of planar graphs to three-dimensional terrains, which is instrumental for the strictly simple case (Theorem 1), is currently being extended to self-touching chains and linkages [RR04]. We hope that this extension will shed further light on locked and unlocked chains of shapes.

References

[CDR02] Robert Connelly, Erik D. Demaine, and Günter Rote. Infinitesimally locked self-touching linkages with applications to locked trees. In J. Calvo, K. Millett, and E. Rawdon, editors, *Physical Knots: Knotting, Linking, and Folding of Geometric Objects in 3-space*, pages 287–311. American Mathematical Society, 2002.

[CDR03] Robert Connelly, Erik D. Demaine, and Günter Rote. Straightening polygonal arcs and convexifying polygonal cycles. *Discrete & Computational Geometry*, 30(2):205–239, September 2003.

[RR04] Ares Ribó and Günter Rote. The Maxwell-Cremona correspondence for self-touching linkages. In preparation, 2004.