

On Folding and Unfolding with Linkages and Origami

by

Zachary Abel

B.A., Harvard University (2010)

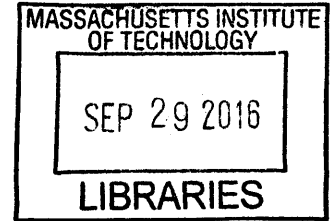
Submitted to the Department of Applied Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2016



© Massachusetts Institute of Technology 2016. All rights reserved.

Signature redacted

Author

Department of Applied Mathematics
September 20, 2016

Signature redacted

Certified by

Erik D. Demaine
Professor of Computer Science and Engineering
Thesis Supervisor

Signature redacted

Accepted by

Jonathan Kelner
Mark Hyman, Jr. Career Development Associate Professor of Applied Mathematics
Chairman, Department Committee on Graduate Theses

On Folding and Unfolding with Linkages and Origami

by
Zachary Abel

Submitted to the Department of Applied Mathematics
on September 20, 2016, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We revisit foundational questions in the kinetic theory of linkages and origami, investigating their folding/unfolding behaviors and the computational complexity thereof.

In Chapter 2, we exactly settle the complexity of realizability, rigidity, and global rigidity for graphs and linkages in the plane, even when the graphs are (1) promised to avoid crossings in all configurations, or (2) equilateral and required to be drawn without crossings (“matchstick graphs”): these problems are complete for the class $\exists\mathbb{R}$ defined by the Existential Theory of the Reals, or its complement. To accomplish this, we prove a strong form of Kempe’s Universality Theorem for linkages that avoid crossings.

Chapter 3 turns to “self-touching” linkage configurations, whose bars are allowed to rest against each other without passing through. We propose an elegant model for representing such configurations using infinitesimal perturbations, working over a field $\mathbb{R}\langle\varepsilon\rangle$ that includes formal infinitesimals. Using this model and the powerful Tarski-Seidenberg “transfer” principle for real closed fields, we prove a self-touching version of the celebrated Expansive Carpenter’s Rule Theorem.

We switch to folding polyhedra in Chapter 4: we show a simple technique to continuously flatten the surface of any convex polyhedron without distorting intrinsic surface distances or letting the surface pierce itself. This origami motion is quite general, and applies to convex polytopes of any dimension. To prove that no piercing occurs, we apply the same infinitesimal techniques from Chapter 3 to formulate a new formal model of self-touching origami that is simpler to work with than existing models.

Finally, Chapter 5 proves polyhedra are hard to edge unfold: it is NP-hard to decide whether a polyhedron may be cut along edges and unfolded into a non-overlapping net. This edge unfolding problem is not known to be solvable in NP due to precision issues, but we show this is not the only obstacle: it is NP complete for orthogonal polyhedra with integer coordinates (all of whose unfolding also have integer coordinates).

Thesis Supervisor: Erik D. Demaine

Title: Professor of Computer Science and Engineering

Acknowledgements

What a journey this has been—rewarding, enlightening, amazing, occasionally detrimental to my health, and worth every moment. While these pages represent a cross-section of my academic output as a graduate student, and while I am proud to present them as such, they do little to capture the truly collective effort that enabled their creation. Everything is a collaboration, and I am indebted to so many for their roles in making this possible. My heartfelt gratitude extends to...

...**my advisor Erik and grand-advisor (of a sort) Marty:** For your unending generosity of time, energy, advice, expertise, and resources. For your contagious energy, flexibility, and curiosity. And for your continued warmth and friendship.

...**my many other academic mentors and role models during graduate school, including Greg Aloupis, Henry Cohn, Noam Elkies, Sándor Fekete, Chuck Hoberman, Tom Hull, Jon Kelner, Robert Lang, Stefan Langerman, Anna Lubiw, Joseph O'Rourke, Daniela Rus, Sug Woo Shin, Diane Souvaine, Tomohiro Tachi, Godfried Toussaint, and Ryuhei Uehara:** For your welcoming attitude, free and fruitful exchange of ideas, and especially your generously offered wisdom and patience. As I met many of you through problem sessions organized by Erik and/or Godfried (at MIT or Barbados), another thank-you goes to them for the invitations and facilitation.

...**my many other coauthors and collaborators—those on papers included in this thesis, namely Sarah Eisenstat, Jin-ichi Itoh, Chie Nara, Jayson Lynch, and TB Schardl; other frequent collaborators, notably Hugo Akitaya, Pesto Hesterberg, Jason Ku, and Andrew Winslow; and many others besides:** for comprising the core of my daily work and social interactions as a researcher, and for always making those interactions so thoroughly fun.

...**my family—especially Mom, Dad, Matt, Panjola, Alissa, David, Don, Jan, Den, Sue, Max, and Gracie—and extended family—Scott, Jannis, Hannah, Ivan, Ruthi, Will, and Jon:** for the unconditional love, encouragement, and support that only a family can provide, even from a physical distance. For keeping me grounded when my head was (is?!) far away in the clouds, and for helping me climb out of some difficult health issues by stepping in and climbing right there beside me.

...**Grant Mindle and Barbara Currier:** for setting me on this path. For reaching far beyond the call of duty by kindling, nurturing, and stoking my love of math at a time that made all the difference.

...**department administrators Barbara Peskin, Michele Gallarelli, and Patrice Macaluso, and Assistant Dean Jason McKnight:** for unabating helpfulness and for working tirelessly to accommodate my unique circumstances.

...**and many others, whether omitted by space constraints or just spaciness:** truly, thank you.

Contents

1	General Introduction	17
1.1	Kempe’s Universality Theorem	18
1.2	The Carpenter’s Rule Theorem	18
1.3	Polyhedron Flattening	20
1.4	Polyhedron Unfolding	20
2	Hardness of Plane Graph Rigidity	23
2.1	Introduction	23
2.2	Description of the Main Construction	26
2.2.1	Definitions: Linkages and Graphs	26
2.2.2	Constrained Linkages	27
2.2.3	Drawing with Linkages and Graphs	28
2.2.4	Specification of the Main Theorem	29
2.2.5	Roadmap	30
2.3	Preliminaries on Semialgebraic Sets and $\exists\mathbb{R}$	31
2.3.1	Semialgebraic Sets as Projections of Algebraic Sets	31
2.3.2	Existential Theory of the Reals	32
2.4	Globally Noncrossing Graphs and Linkages	33
2.4.1	Rigidifying Polygons	33
2.4.2	Rigidifying Polyominoes	34
2.4.3	Simulation with Globally Noncrossing Linkages	36
2.4.4	Hardness and Universality of Globally Noncrossing Linkages	37
2.5	Unit-Distance and $\{1, 2\}$ -Distance Graphs and Linkages	38
2.5.1	Simulation with Unit- and $\{1, 2\}$ -Distance Linkages	39
2.5.2	Hardness and Universality of Unit-Distance and $\{1, 2\}$ -Distance Graphs	40
2.6	Matchstick Graphs and Linkages	42
2.6.1	Simulation with Matchstick Linkages	42
2.6.2	Hardness and (Weak) Universality of Matchstick Linkages	44
2.7	Extended Linkages and the Main Construction	45
2.7.1	Defining Extended Linkages	45
2.7.2	Detailed Overview of Strategy	46
2.7.3	The Gadgets	48
2.7.4	Combining the Gadgets	59
2.7.5	Sliceform and Angle Constraint Gadgets	62
2.7.6	Implementing Extended Linkages with Partially Rigidified Linkages	64
2.7.7	Modifications for Strong Matchstick Universality	66

3	Self-Touching Linkages and the Expansive Self Touching Carpenter’s Rule Theorem	69
3.1	Introduction	69
3.1.1	Our Results	70
3.1.2	Motivation for Our Infinitesimal Model	70
3.1.3	Two Prior Models of Self-Touching Linkages	72
3.2	Background on $\mathbb{R}\langle\varepsilon\rangle$ and Real Closed Fields	74
3.2.1	Real Closed Fields	74
3.2.2	Adjoining Infinitesimals	74
3.2.3	Sizes of Elements in $R\langle\varepsilon\rangle$	75
3.2.4	Projecting and Lifting	75
3.3	Expansive Self-Touching Carpenter’s Rule Theorem	75
3.3.1	Self-Touching Linkages via Infinitesimals	75
3.3.2	Expansive Self-Touching Carpenter’s Rule Theorem in the Infinitesimal Model	76
3.4	Gradual Functions	77
3.5	Translating to the Combinatorial and Limiting Models	79
3.5.1	From Infinitesimal to Combinatorial	80
3.5.2	From Infinitesimal To Limiting	84
3.5.3	Expansive Self-Touching Carpenter’s Rule Theorem	85
3.6	Linkages with Slender Adornments	85
3.6.1	Linkage Adornments	86
3.6.2	Slender Adornments	87
3.6.3	Chains with Slender Adornments Can Always Open	87
3.7	Extensions and Future Work	88
4	Continuously Flattening Polyhedra Using Straight Skeletons	91
4.1	Introduction	91
4.1.1	Flattening Polyhedra with Origami	91
4.1.2	Theoretical Foundations of Origami: Challenges of Self-Touching	92
4.1.3	Acknowledgements	93
4.2	Positive Hyperplane Arrangements	93
4.3	Orderly Squashing	94
4.4	Computing Orderly Squashing	98
4.4.1	The Algorithm	98
4.4.2	An Implementation	99
4.5	Challenges of Layer Ordering	100
4.5.1	Origami via Infinitesimals	100
4.6	Diminished Orderly Squashing	102
4.7	Folding Tessellations	104
4.7.1	Formal Setup for Spiderweb Folding	105
4.7.2	Layer Ordering for Tessellation Folding	107
5	Hardness of Polyhedron Edge Unfolding	109
5.1	Introduction	109
5.2	Unique Coordinate Square Packing	109
5.3	Overview	110
5.4	Polyhedron With Boundary	111

5.4.1	Atoms and Universal Wire Unfolding	111
5.4.2	The Construction	114
5.4.3	Wiring the Tower	115
5.4.4	Unfolding Surface B	118
5.5	Eliminating the Boundary	119

List of Figures

1-1	Two famous linkages for constraining straight-line motion. Watt’s Linkage (left) has three bars, and if a and b are fixed in place, the midpoint c of the middle bar moves <i>approximately</i> along a vertical line segment. Similarly, in the Peaucellier-Lipkin linkage (right), points a and b are fixed in place, and point c moves along an <i>exact</i> vertical line segment.	17
1-2	Snapshots of an expansive motion for a closed chain, computed by the authors of [32] according to the algorithm from that paper. Image used with permission.	19
1-3	A polyhedron that has no edge unfolding [4]. Figure taken from [4], with permission.	21
2-2	Any connected polyomino of 1440×1440 squares can be turned into a globally rigid graph with integer coordinates and integer edge lengths.	35
2-3	Thickening a rigidified subtree (H, C_H) into a globally rigid polygon.	36
2-4	The trace of an NX-constrained linkage need not be closed.	42
2-5	Left: Edge polyiamonds used to simulate edges of integer length. Right: Edge polyiamonds braced at 90°	43
2-6	The region described in Lemma 2.7.5 (gray) is bounded by four circular arcs and contains the square $[-\theta/2, \theta/2]^2$	47
2-7	Creating parallelograms with extended linkages.	49
2-8	Left: The Parallel Gadget allows e to move freely in a neighborhood of its initial position while forcing ef to remain parallel to ab . Right: a schematic representation of the same gadget.	50
2-9	Gadgets for manipulating angles in the Main Construction. In these figures we use the convention that $\text{Diff}_\Theta(C) = \theta$, $\text{Diff}_\Lambda(C) = \alpha$, and $\text{Diff}_\Gamma(C) = \beta$, possibly with subscripts. Angle chains marked with a solid gray sector are frozen, i.e., have tolerance 0. Unfrozen angle chains at cell edge midpoints are assigned tolerance δ . All remaining angle chains have tolerance ε . Vertices surrounded by squares are pinned, and those marked with an “x” are sliceform vertices. The pins shown here at the vertices a_i are for clarification only; in the overall construction, these nodes are forbidden from moving by other means, so these explicit pins are unnecessary.	51
2-10	Snapshots of the configuration space of the Copy Gadget (cf. Figure 2-9a, Lemma 2.7.10), shown for angle offsets of 3° , 0° and -3° respectively.	52
2-11	Snapshots of the Angle Average Gadget’s configuration space (cf. Lemma 2.7.10, Figure 2-9c), shown for each pair of values $\theta_1, \theta_3 \in \{-8^\circ, 0^\circ, 8^\circ\}$ (where $\theta_i = \text{Diff}_{\Theta_i}$). In each configuration, it may be seen that $\theta_2 = (\theta_1 + \theta_3)/2$. Edge lengths have been altered from those in Figure 2-9c to exaggerate the gadget’s movement.	53

2-12	Gadgets for manipulating vectors in the Main Construction. In these figures we use the conventions that $\text{Diff}_\Lambda(C) = \alpha$ and $\text{Diff}_\Gamma(C) = \beta$, possibly with subscripts, in addition to the conventions of Figure 2-9.	56
2-13	Vector Sum Gadget	57
2-14	The Vector Sum Gadget is built out of more primitive gadgets from Lemmas 2.7.10 and 2.7.11.	57
2-15	Computing the sum $\gamma = \beta_1 + \alpha_2$. Cells with “S” are start gadgets; those with “x” are crossover gadgets; and those with α_j , β_j , or γ are copy gadgets.	60
2-16	The Sliceform gadget keeps w_1, v, w_3 and w_2, v, w_4 collinear.	63
2-17	Angle Restrictor Gadget, $\mathcal{L}_{\text{restrict}}$, shown in full (left) and closeup (right). . .	64
2-18	The Crossing End Gadget creates a crossing at $g = h$ precisely when $\alpha = \beta = 0$, assuming g remains on the line $y = Q/2$	67
3-1	A visual representation of a self-touching linkage in our infinitesimal model. The precise vertex locations are listed in equation (3.1). Spaces between touching bars are exaggerated; they are actually infinitesimal.	71
3-2	A self-touching linkage configuration as represented by the combinatorial and limiting models.	73
3-3	For a node u near chain $v_1v_2v_3$, u lies to the left of the chain when the unit vectors in the directions of v_2v_1 , v_2w , and v_2v_3 respectively are arranged in counterclockwise cyclic order.	81
3-4	Two motions allowed by our Vertex-Chain condition but not that in [33]. These motions are forbidden by the Vertex-Edge constraint, however, so this weaker Vertex-Chain condition does not affect the model overall.	81
3-5	Both halves of condition $L_{\text{EE}}(c, u_1u_2, v_1v_2)$ need not hold simultaneously. . . .	82
3-6	When $ u_1v_1 $ and u_2v_2 remain small, midpoint v_m is constrained to remain near the midpoint of edge u_1u_2	83
3-7	An example motivating the modification of our definition of adornment overlap. In the figure, edge uw is touching along the left side of edge uv and therefore does not intersect the interior of adornment D (only its boundary). But D and uw should still be considered overlapping.	86
3-8	Half-lenses are the fundamental examples of slender adornments; half-lens(p, uv) is shown. Slender adornments are precisely those that are expressible as a union of half-lenses.	87
3-9	A self-touching linkage configuration in \mathbb{R}^3 that has no nontouching perturbations (in either \mathbb{R}^3 or $\mathbb{R}\langle\varepsilon\rangle^3$) without artificially inserting more vertices. . .	88
4-1	Examples of 2D positive hyperplane arrangements: (a) positive cell P_5 is shaded; (b) a nonconvex polygonal subset is shown in bold.	94
4-2	Orderly squashing of a tetrahedron. The left pane shows the ordering of faces and the crease pattern, with primary creases in thick dark red and other creases in thin light red. The other panes show an animation of the flattening. Face 2, originally at the back, ends up on top.	95
4-3	Orderly squashing of a convex polyhedron: (top) the ordering of front faces; (middle row) front view of the creases and two stages of the flattening; (bottom row) back view of the creases and two stages of the flattening. Primary creases are shown in thick dark red and other creases in thin light red. . . .	96

4-4	Orderly squashing of a 3D positive hyperplane arrangement: (from left to right) the ordering of hyperplanes; the crease pattern, with primary creases in thick dark red and others in thin light red; two frames of the flattening.	97
4-5	A convex polyhedron where any face ordering produces $\Theta(n^2)$ primary creases: (left) the polyhedron; (middle) the primary creases for one face ordering; (right) all creases for that face ordering.	99
4-6	Combinatorial complexity of crease patterns (left column) and performance of our implementation (right column) for the primary crease pattern (top row) and full crease pattern (bottom row).	100
4-7	Origami tessellations made from a spiderweb graph through a limiting orderly squash. Small offsets (left) result in elegant tessellations, while large offsets (right) have more interaction.	105
5-1	A depiction of an “atom,” the polyhedral surface with 9×9 square boundary that enables universal wire unfolding as in Theorem 5.4.1. An atom is composed of 125 unit-square faces.	111
5-2	An illustration of Theorem 5.4.1: a path of atoms that turns right, right, left, straight in sequence is unfolded into the path of flatoms with the entirely different turn sequence right, left, straight, left.	112
5-3	All nine required unfoldings of an atom inside a flatom. Each unfolding is labeled with $[X, Y, t, u]$ as described in the proof of Theorem 5.4.1. Solid black lines indicate cuts, dotted lines are valley folds, and dashed lines are mountain folds. Gray lines are uncreased edges.	113
5-4	Detailed depiction of polyhedral surface B	114
5-5	The three steps in the construction of molecule wires W_i of Lemma 5.4.3. Each wire W_i connects molecule u_i to $v_{\sigma(i)}$, where u_1, \dots, u_n are lined up along the bottom edge from left to right, and each molecule v_i is just under the lower left corner of brick b_i . The figures correspond to $\sigma(1) = 3$, $\sigma(2) = 1$, $\sigma(3) = 2$, and $\sigma(4) = 4$	116
5-6	A closeup of the bottom-left corner of T_{bottom} illustrating how to write G as a wire of atoms as in the proof of Lemma 5.4.4. This wire, X_0 , is formed by traversing the four atoms of wire X'_0 , then walking around the spanning tree S , and finally following atom-wire X''_0	117
5-7	The bricks in a unique-coordinate square packing are attached to the bottom-left corner of the cage via non-overlapping wires of flatoms, as follows: First draw wires directly up and right. Then resolve wire-brick overlaps by detours as shown here. Finally, make each wire zig-zag in the small dotted region to bring its length up to $4L$	118
5-8	The polyhedron $C = C(d, (a_1, \dots, a_n))$ without boundary.	119

List of Tables

1.1	This Thesis's four topics categorized by behavior (folding or unfolding) and object (linkage or polyhedron).	19
2.1	Summary of our results (bold) compared with earlier results (cited).	24

Chapter 1

General Introduction

This section provides a gentle, high-level overview of the contents of this thesis, prioritizing clarity over technical precision. A more rigorous and thorough introduction accompanies each separate chapter as new questions and concepts are, well, introduced.

Mathematical linkages are an abstraction of physical mechanisms—mechanical linkages—that are built from rigid members or supports (such as metal rods) joined with rotatable hinges, prevalent in machinery and other physical devices. Famous examples include Watt’s linkage and the Peaucellier-Lipkin Inversor for (approximately or exactly) converting rotational motion to linear motion (Figure 1-1), an essential constraint in the movement of pistons, for example. We needn’t look far for familiar examples—in many vehicles, mechanical linkages govern pivotal functions like suspension (often with Watt’s Linkage [9]), steering, windshield wipers, and many others. The theory of mathematical linkages is unsurprisingly widely applied in machinery, engineering, and robotics, but recently it has also found deep connections to protein folding [42, 70].

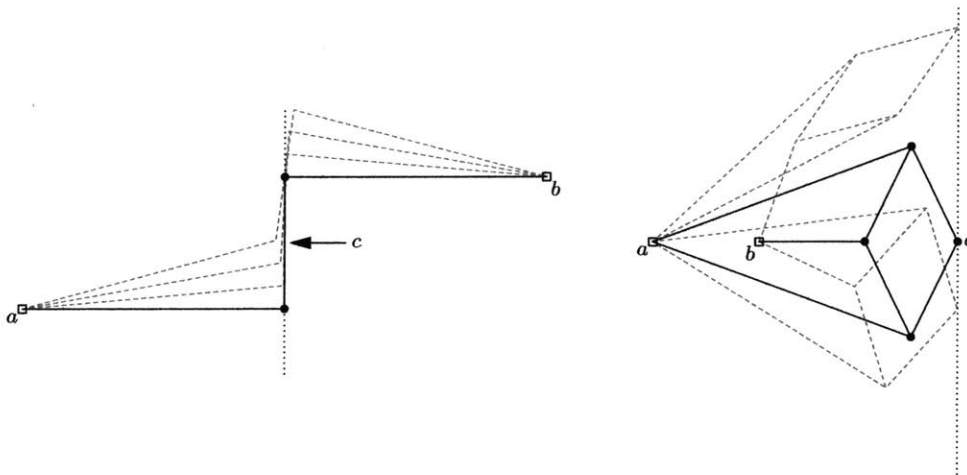


Figure 1-1: Two famous linkages for constraining straight-line motion. Watt’s Linkage (left) has three bars, and if a and b are fixed in place, the midpoint c of the middle bar moves *approximately* along a vertical line segment. Similarly, in the Peaucellier-Lipkin linkage (right), points a and b are fixed in place, and point c moves along an *exact* vertical line segment.

Origami—the art of paper folding—has similarly found diverse uses beyond the artistic. Origami’s graceful yet sturdy kinematics have facilitated advances in self-folding robotics [68] and space satellites [85], to name a few. When applied to enclosed three-dimensional volumes, i.e., polyhedra, origami techniques have also proven useful for airbag (un)folding [64], deployable architecture [82], and surgical implants [63].

To enable and support these far-reaching consequences, a thorough development of the underlying mathematical theory of folding and unfolding is essential, and progress is ongoing in the active research communities of mathematical/computational geometry and graph rigidity theory. Current definitions, techniques, results, and unsolved problems relating to mathematical folding and unfolding are thoroughly surveyed in *Geometric Folding Algorithms* by Demaine and O’Rourke [43].

This thesis addresses four fundamental questions, each discussed at length in [43], that may be broadly categorized along two axes: objects of study—either linkages or polyhedra—and type of behavior—either folding or unfolding. Pleasingly, these four questions fit perfectly into these four categories, one from each, as depicted in Table 1.1. We briefly describe each problem below, with more thorough coverage to be found in the appropriate chapters.

1.1 Kempe’s Universality Theorem

In the Linkage Folding corner of Table 1.1, we revisit the classical Kempe’s Universality Theorem, which provides a positive answer to a question of mechanism design: given any desired trajectory of movement in the plane, it is possible to construct a linkage that causes a pen to follow this movement as closely as you wish. This has been summarized informally as “there’s a linkage to sign your name”, but until now there has been a disconnect between the mathematical and physical interpretations of this result: Kempe’s Universality Theorem from 1876, and all more recent improvements thereof, allow the edges of their mathematical linkages to pass through each other. There is currently no known method for “phasing” a physical metal bar through another, so these results cannot be directly physically realized.

In Chapter 2, we show that these edge crossings are unnecessary: Kempe’s Universality Theorem is true even for linkages that cannot and do not self-intersect, thus enabling physical construction of such mechanisms (though actual physical construction of such mechanisms remains as future work.) In addition to solving this decade-old open problem, our result also has implications for the *computational* aspect of linkage kinematics: given the specification of a linkage, algorithmically determining *whether* or *how* this linkage may fold is known to be a computationally demanding task, even harder than the intractable NP-complete problems. We show that this remains intractable even for linkages that do not intersect themselves, settling an open question from computational graph rigidity theory.

1.2 The Carpenter’s Rule Theorem

In Chapter 3 we turn to a question in Linkage Unfolding: can every chain linkage be unfolded? Specifically, suppose we are given a linkage in the shape of a chain—simply a path of edges joined end to end—that has been drawn in the plane with no self-intersection, even though it may come close to touching itself as in the example of Figure 1-2. The well-known Carpenter’s Rule Theorem shows that it is always possible to continuously unfold such a chain to a fully opened position without changing edge lengths or letting any part of the chain intersect another part, as illustrated in Figure 1-2.

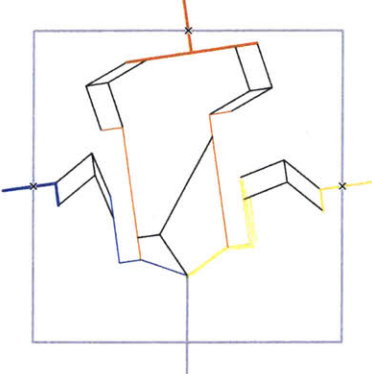
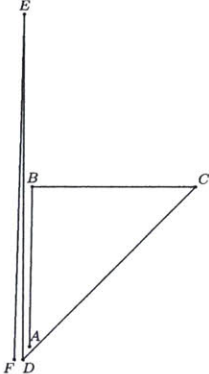
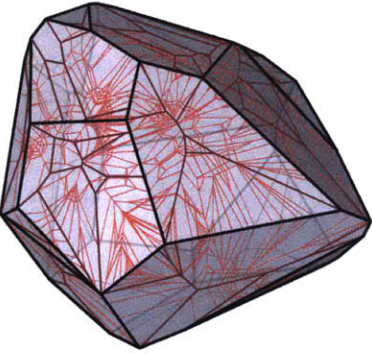
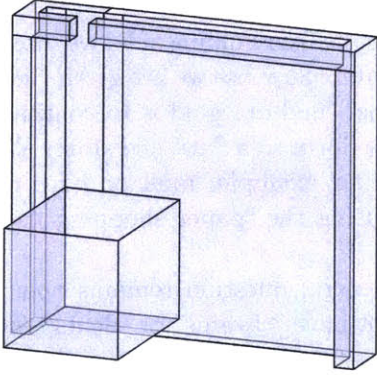
	Folding	Unfolding
Linkages	 <p>Kempe's Universality Theorem (Chapter 2)</p>	 <p>The Carpenter's Rule Theorem (Chapter 3)</p>
Origami Polyhedra	 <p>Polyhedron Flattening (Chapter 4)</p>	 <p>Polyhedron Unfolding (Chapter 5)</p>

Table 1.1: This Thesis's four topics categorized by behavior (folding or unfolding) and object (linkage or polyhedron).

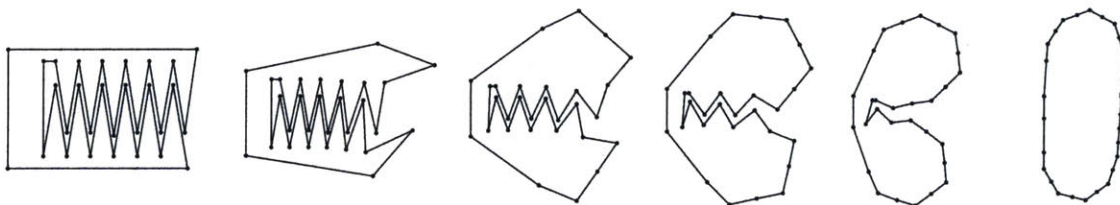


Figure 1-2: Snapshots of an expansive motion for a closed chain, computed by the authors of [32] according to the algorithm from that paper. Image used with permission.

The question we revisit is this: what happens if the linkage *does* rest against itself at the start without passing all the way through itself (much like a piece of zero-thickness paper that has been folded onto itself)? For example, perhaps the “teeth” in Figure 1-2 are pushed up against each other, and perhaps also closed together into a single, multi-layered line segment. Are all such self-touching linkages also continuously unfoldable without crossing through themselves?

Again this answer is *yes*, as previously demonstrated in [1] with difficult topological methods. In Chapter 3, we prove an even stronger result with a simpler proof. Our key insight is a new way to formalize the intuitive but tricky notion of “touching without piercing,” which has been notoriously challenging to work with in both the linkage and origami literature but is essential to understanding layer ordering in origami and other thin material settings. Our new formalization, termed the *infinitesimal model* of self-touching linkages, reveals that the Carpenter’s Rule Theorem for self-touching linkages is in fact no different from the prior nontouching case, and it paves the way for more effective handling of self-touching phenomena going forward.

1.3 Polyhedron Flattening

On the Polyhedron Folding side, we discuss in Chapter 4 the problem of flattening polyhedra using origami. Suppose we are given the surface of some three-dimensional polyhedron made out of paper, and our goal is to continuously bend, fold, crease, and otherwise manipulate the surface down to a flattened state. We are not allowed, however, to tear or rip the paper. As a familiar example, milk or juice cartons are often designed to be folded flat before disposal, using the “paper shopping bag” crease pattern. Can any shape be flattened like this?

This general question remains open, but the special case of *convex* polyhedra—those that do not have “caverns” or “dents”—was shown to always be possible by Itoh, Nara, and Vilcu [49]. In Chapter 4 of this thesis, we offer a different method of flattening convex polyhedra that improves on [49] in a number of ways. Our method, called *orderly squashing*, is extremely simple to describe and compute, in contrast to Itoh et al.’s reliance on Alexandrov’s Gluing Theorem, a powerful but opaque tool for reconstructing the 3D shape of a polyhedron from knowledge of its surface. Our resulting flattening better adheres to the polyhedron’s intrinsic geometry, following the so-called *straight skeleton gluing*, solving a conjecture from [43]. Finally, our technique generalizes to a certain class of nonconvex polyhedra (but, alas, the general problem for nonconvex polyhedra remains open), and even to origami with higher-dimensional polytopes.

To analyze orderly squashing, we adapt our infinitesimal model of self-touching linkages from Chapter 3 to the origami setting, thus providing a new formal framework for self-touching origami that we argue is vastly simpler to reason about. As evidence, we provide a fully formal analysis of both the geometry *and* layer ordering of orderly squashing as an origami process.

1.4 Polyhedron Unfolding

Finally, Chapter 5 considers the famous topic of Unfolding Polyhedra: for a given three-dimensional polyhedron, is it possible to cut the surface and unfold it into a single, connected piece in the plane *that does not overlap itself*? The search for such polyhedral unfoldings

has a long history (beginning with Albrecht Dürer in 1525, if not earlier) and represents a simple method to cut out and physically construct the polyhedron from, say, flat paper or sheet metal. The non-overlapping condition expresses the fact that two different parts of the surface cannot be cut from the same portion of material. Many variations on this question are especially applicable to manufacturing and architecture.

There is special interest in *edge unfoldings*, which are non-overlapping unfoldings that cut the surface only along its edges, leaving each face intact. Edge unfoldings for convex polyhedra such as the Platonic and Archimedean solids are readily available, but it is unknown whether *every* convex polyhedron has an edge unfolding. What is known is that not all *nonconvex* polyhedra have an edge unfolding: for some nonconvex polyhedra, no matter which set of edges you choose to cut (without completely separating the surface), the surface will always overlap itself when unfurled in the plane. One such “ununfoldable” example is the polyhedron formed by overlapping two triangular prisms as in Figure 1-3.

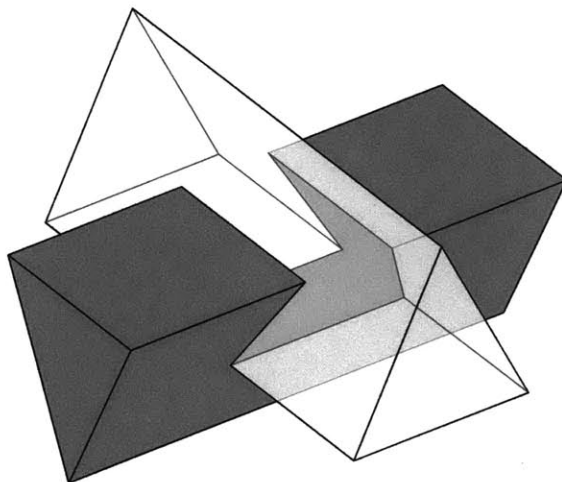


Figure 1-3: A polyhedron that has no edge unfolding [4]. Figure taken from [4], with permission.

Since some nonconvex polyhedra cannot be edge unfolded, the natural next question is algorithmic: how hard is it to write a computer program that, when given the description of some (possibly nonconvex) polyhedron, can decide whether or not an edge unfolding exists? In Chapter 5 we show that, unfortunately, the answer is “very hard.” Indeed, we show that this problem is NP-hard, making it very unlikely that an efficient algorithm exists.

Chapter 2

Hardness of Plane Graph Rigidity

Chapter Summary. We exactly settle the complexity of graph realization, graph rigidity, and graph global rigidity as applied to three types of graphs: “globally noncrossing” graphs, which avoid crossings in all of their configurations; matchstick graphs, with unit-length edges and where only noncrossing configurations are considered; and unrestricted graphs (crossings allowed) with unit edge lengths (or in the global rigidity case, edge lengths in $\{1, 2\}$). We show that all nine of these questions are complete for the class $\exists\mathbb{R}$, defined by the Existential Theory of the Reals, or its complement $\forall\mathbb{R}$; in particular, each problem is (co)NP-hard.

One of these nine results—that realization of unit-distance graphs is $\exists\mathbb{R}$ -complete—was shown previously by Schaefer (2013), but the other eight are new. We strengthen several prior results. Matchstick graph realization was known to be NP-hard (Eades & Wormald 1990, or Cabello et al. 2007), but its membership in NP remained open; we show it is complete for the (possibly) larger class $\exists\mathbb{R}$. Global rigidity of graphs with edge lengths in $\{1, 2\}$ was known to be coNP-hard (Saxe 1979); we show it is $\forall\mathbb{R}$ -complete.

The majority of the paper is devoted to proving an analog of Kempe’s Universality Theorem—informally, “there is a linkage to sign your name”—for globally noncrossing linkages. In particular, we show that any polynomial curve $\varphi(x, y) = 0$ can be traced by a noncrossing linkage, settling an open problem from 2004. More generally, we show that the nontrivial regions in the plane that may be traced by a noncrossing linkage are precisely the compact semialgebraic regions. Thus, no drawing power is lost by restricting to noncrossing linkages. We prove analogous results for matchstick linkages and unit-distance linkages as well.

2.1 Introduction

The rise of the steam engine in the mid-1700s led to an active study of *mechanical linkages*, typically made from rigid bars connected together at hinges. For example, steam engines need to convert the linear motion of a piston into the circular motion of a wheel, a problem solved approximately by Watt’s parallel motion (1784) and exactly by Peaucellier’s inversor (1864) [43, Section 3.1]. These and other linkages are featured in an 1877 book called *How to Draw a Straight Line* [58] by Alfred Bray Kempe—a barrister and amateur mathematician in London, perhaps most famous for his false “proof” of the Four-Color Theorem [59] that nonetheless introduced key ideas used in the correct proofs of today [13, 72].

Graph type	Realization	Rigidity	Global rigidity	Universality
General	$\exists\mathbb{R}$ -complete [74]	$\forall\mathbb{R}$ -complete [74]	$\forall\mathbb{R}$-complete (CoNP-hard [73])	Compact semialg. [60]
Globally noncrossing (no configs. cross)	$\exists\mathbb{R}$-complete	$\forall\mathbb{R}$-complete	$\forall\mathbb{R}$-complete	Compact semialg.
Matchstick graph (unit + noncrossing)	$\exists\mathbb{R}$-complete (NP-hard [45])	$\forall\mathbb{R}$-complete	$\forall\mathbb{R}$-complete	Bounded semialg.
Unit edge lengths (allowing crossings)	$\exists\mathbb{R}$ -complete [74]	$\forall\mathbb{R}$-complete	Open (do they even exist?)	Compact semialg.
Edge lengths in $\{1, 2\}$ (allowing crossings)	$\exists\mathbb{R}$ -complete [74]	$\forall\mathbb{R}$-complete	$\forall\mathbb{R}$-complete (CoNP-hard [73])	Compact semialg.

Table 2.1: Summary of our results (bold) compared with earlier results (cited). The rows give the special types of graphs considered. The middle three columns give complexity results for the three natural decision problems about graph embedding; all completeness results are strong. The rightmost column gives the exact characterization of drawable sets.

Kempe’s Universality Theorem. Kempe wondered far beyond drawing a straight line by turning a circular crank. In 1876, he claimed a universality result, now known as Kempe’s Universality Theorem: every polynomial curve $\varphi(x, y) = 0$ can be traced by a vertex of a 2D linkage [57]. Unfortunately, his “proof” was again flawed: the linkage he constructs indeed traces the intended curve, but also traces finitely many unintended additional curves. Fortunately, his idea was spot on.

Many researchers have since solidified and/or strengthened Kempe’s Universality Theorem [54, 51, 60, 2, 74]. In particular, small modifications to Kempe’s gadgets lead to a working proof [2, 43, Section 3.2]. Furthermore, the regions of the plane drawable by a 2D linkage (other than the entire plane \mathbb{R}^2) are exactly compact semialgebraic regions* [60, 2]. By carefully constructing these linkages to have rational coordinates, Abbott et al. [2] showed how to reduce the problem of testing isolatedness of a point in an algebraic set to testing rigidity of a linkage. Isolatedness was proved coNP-hard [62] and then $\forall\mathbb{R}$ -complete[†] [74]; thus linkage rigidity is $\forall\mathbb{R}$ -complete.

Our results: no crossings. See Table 2.1 for a summary of our results in comparison to past results. Notably, all known linkage constructions for Kempe’s Universality Theorem (and its various strengthenings) critically need to allow the bars to cross each other. In practice, certain crossings can be made physically possible, by placing bars in multiple parallel planes and constructing vertices as vertical pins. Without extreme care, however, bars can still be blocked by other pins, and it seems difficult to guarantee crossing avoidance for complex linkages. Beyond these practical issues, it is natural to wonder whether allowing bars to cross is necessary to achieve linkage universality. Don Shimamoto first posed this problem in April 2004, and it was highlighted as a key open problem in the first chapter of *Geometric Folding Algorithms* [43].

*A compact planar region is **semialgebraic** if it can be obtained by intersecting and/or unioning finitely many basic sets defined by polynomial inequalities $p(x, y) \geq 0$.

[†]The class $\forall\mathbb{R}$ = co- $\exists\mathbb{R}$ consists of decision problems whose complement (inverting yes/no instances) belong to $\exists\mathbb{R}$. The class $\exists\mathbb{R}$ refers to the problems (Karp) reducible to the **existential theory of the reals** ($\exists x_1 : \dots \exists x_n : \pi(x_1, \dots, x_n)$ for a Boolean function $\pi : \mathbb{R} \rightarrow \{0, 1\}$), which is somewhere between NP and PSPACE (by [27]). The classic example of an $\exists\mathbb{R}$ -complete problem is pseudoline stretchability [69].

We solve this open problem by strengthening most of the results mentioned above to work for **globally noncrossing graphs**, that is, graphs plus edge-length constraints that alone force all configurations to be (strictly) noncrossing.* In particular, we prove the following universality and complexity results:

1. The planar regions drawable by globally noncrossing linkages are exactly (\mathbb{R}^2 and) the compact semialgebraic regions (Theorem 2.4.8), settling Shimamoto’s 2004 open problem.
2. Testing whether a globally noncrossing graph has any valid configurations is $\exists\mathbb{R}$ -complete (Theorem 2.4.6).
3. Testing rigidity is strongly $\forall\mathbb{R}$ -complete even for globally noncrossing linkages drawn with polynomially bounded integer vertex coordinates (Theorem 2.4.7).
4. Testing global rigidity (uniqueness of a given embedding) is strongly $\forall\mathbb{R}$ -complete even for globally noncrossing linkages drawn with polynomially bounded integer vertex coordinates (Theorem 2.4.7).

Our techniques are quite general and give us results for two other restricted forms of graphs as well. First, **matchstick graphs** are graphs with *unit* edge-length constraints, and where only (strictly) noncrossing configurations are considered valid. We prove in Section 2.6 the following universality and complexity results:

5. The planar regions drawable by matchstick graphs are exactly (\mathbb{R}^2 and) the bounded semialgebraic regions (Theorem 2.7.15). Notably, unlike all other models considered, matchstick graphs enable the representation of open boundaries in addition to closed (compact) boundaries.
6. Recognizing matchstick graphs is (strongly) $\exists\mathbb{R}$ -complete (Theorem 2.6.5). This result strengthens a 25-year-old NP-hardness result [45, 26], and settles an open question of [74].
7. Testing rigidity or global rigidity of a matchstick graph is strongly $\forall\mathbb{R}$ -complete (Theorems 2.6.6 and 2.6.7).

Second, we consider restrictions on edge lengths to be either all equal (unit) or all in $\{1, 2\}$, but at the price of allowing crossing configurations. Recognizing unit-distance graphs is already known to be $\exists\mathbb{R}$ -complete [74]. We prove in Section 2.5 the following additional universality and complexity results:

8. The planar regions drawable by unit-edge-length linkages are exactly the compact semialgebraic regions (and \mathbb{R}^2) (Theorem 2.5.9), proving a conjecture of Schaefer [74].
9. Testing rigidity of unit-edge-length linkages is strongly $\forall\mathbb{R}$ -complete (Theorem 2.5.7), proving a conjecture of Schaefer [74].
10. Testing global rigidity of linkages with edge lengths in $\{1, 2\}$ is strongly $\forall\mathbb{R}$ -complete (Theorem 2.5.8). This result strengthens a 35-year-old strong-coNP-hardness result for the same scenario [73]. While it would be nice to strengthen this result to unit edge lengths, we have been unable to find even a single globally rigid equilateral linkage larger than a triangle.

*Thus, the noncrossing constraint can be thought of as being “required” or not of a configuration; in either case, the configurations (even those reachable by discontinuous motions) will be noncrossing.

We introduce several techniques to make noncrossing linkages manageable in this setting. In Section 2.7.1 we define *extended linkages* to allow additional joint types, in particular, requiring angles between pairs of bars to stay within specified intervals. Section 2.7.2 then shows how to draw a polynomial curve and obtain Kempe’s Universality Theorem with these powerful linkages while avoiding crossings, by following the spirit of Kempe’s original construction but with specially designed modular gadgets to guarantee no crossings between (or within) the gadgets. We simulate extended linkage with linkages that have chosen subgraphs marked as rigid. In turn, in Sections 2.4–2.6, we simulate these “partially rigidified” linkages with the three desired linkage types: globally noncrossing, unit-distance or $\{1, 2\}$ -distance, and matchstick.

This chapter is joint work with Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Jayson Lynch, and Tao B. Schardl. An extended abstract version appeared in the 2016 Symposium on Computational Geometry [6], and the present chapter (with minor modifications) will appear in the Journal of Computational Geometry’s special issue of selected papers from that Symposium [7].

2.2 Description of the Main Construction

The heart of this paper is a single, somewhat intricate linkage construction. In this section, we describe and discuss the properties of this construction in detail, after building up the necessary terminology.

2.2.1 Definitions: Linkages and Graphs

Unless otherwise specified, all graphs $G = (V(G), E(G), \ell_G)$ in this text are connected, edge-weighted with positive edge lengths $\ell_G(e) > 0$, and contain no self-loops. We first recall and establish notation for our primary objects of study, linkages.

Definition 2.2.1 (Linkages). An **abstract linkage**, or simply a **linkage**, is a weighted graph $G = (V(G), E(G), \ell_G)$ together with a choice of **pin locations** $P(w) \in \mathbb{R}^2$ for vertices w in a chosen subset $W \subset V(G)$ of **pinned vertices**.

Definition 2.2.2 (Linkage Configurations). A **configuration** of a linkage $\mathcal{L} = (G, W, P)$ is an assignment of vertex locations $C : V(G) \rightarrow \mathbb{R}^2$ respecting the edge-length and pin assignments: $|C(u) - C(v)| = \ell_G(u, v)$ for each edge (u, v) , and $C(w) = P(w)$ for each pinned vertex $w \in W$. Two configurations are **congruent** if they differ only by a Euclidean transformation, i.e., a translation, rotation, and possibly reflection.

The **configuration space** $\text{Conf}(\mathcal{L}) \subseteq (\mathbb{R}^2)^{|V(G)|}$ is the set of all configurations; it is a closed, semialgebraic subset of $(\mathbb{R}^2)^{|V(G)|}$. An abstract linkage is called **configurable** or **realizable** if its configuration space is nonempty, and a **configured linkage** is a linkage together with a chosen **initial configuration** $C_0 \in \text{Conf}(\mathcal{L})$. Configured linkage (\mathcal{L}, C_0) is **rigid** if there is no nontrivial continuous deformation of \mathcal{L} , i.e., C_0 is isolated in $\text{Conf}(\mathcal{L})$. Similarly, (\mathcal{L}, C_0) is **globally rigid** if C_0 is the *only* configuration in $\text{Conf}(\mathcal{L})$.

As a convenient abuse of notation, we often write v instead of $C(v)$ when configuration C is understood from context.

We consider **abstract (weighted) graphs** and **configured (weighted) graphs** as abstract or configured linkages without pins, $P = \emptyset$, with one key difference: rigidity and global rigidity for graphs are more liberally defined to allow Euclidean motions.

Definition 2.2.3 (Graph Rigidity and Global Rigidity). A configured graph (G, C_0) is **rigid** if the only continuous deformations of (G, C_0) are rigid Euclidean motions, i.e., if there is a neighborhood of C_0 in $\text{Conf}(G)$ consisting only of configurations congruent to C_0 . Likewise, (G, C_0) is **globally rigid** if all configurations are congruent to C_0 .

Definition 2.2.4 (Noncrossing Configurations). A configuration C of a linkage \mathcal{L} is **noncrossing** if distinct edges intersect only at common endpoints: for any pair of adjacent edges $(u, v) \neq (u, v') \in E(G)$ sharing a vertex u , the segments $C(u)C(v)$ and $C(u)C(v')$ intersect only at $C(u)$ in \mathbb{R}^2 , and for any pair $(u, v), (u', v')$ of disjoint edges in G , segments $C(u)C(v)$ and $C(u')C(v')$ are disjoint in \mathbb{R}^2 .

Let $\text{NConf}(\mathcal{L}) \subseteq \text{Conf}(\mathcal{L})$ denote the subset of noncrossing configurations. Then we say \mathcal{L} is **globally noncrossing** if all of its configurations are noncrossing, i.e., $\text{NConf}(\mathcal{L}) = \text{Conf}(\mathcal{L})$. If C is a noncrossing configuration, the **minimum feature size** of C is the shortest distance from a segment $C(u)C(v)$ to a point $C(w)$ for $(u, v) \in E(G)$ and $w \in V(G) \setminus \{u, v\}$. If linkage \mathcal{L} is globally noncrossing, its **global minimum feature size** is defined as the infimum of the minimum feature size of its configurations.*

Definition 2.2.5 (Combinatorial Embeddings, Angle Chains). A **combinatorial embedding** σ of a graph G consists of a cyclic ordering σ_v of v 's incident edges for each vertex $v \in V(G)$. A noncrossing configuration C of G **agrees with** σ if the counterclockwise cyclic ordering of edges $C(v)C(w)$ around $C(v)$ matches σ_v for each vertex v .

Whenever edge (v, u) is followed by (v, w) in σ_v , the two-edge path $\Lambda = (u, v, w)$ is an **angle chain** of σ at v ; in the special case where v is adjacent to exactly one edge (v, u) , there is a single angle chain, (u, v, u) , at v . If C is a noncrossing configuration of G agreeing with σ , the **angles** of C at v are the angles $\angle C(\Lambda) := \angle C(u)C(v)C(w)$ for each angle chain $\Lambda = (u, v, w)$. When there is only one angle chain (u, v, u) at v , we define $\angle C(u)C(v)C(u)$ to have measure 360° . With this convention, the angles of C at v add to 360° no matter v 's degree.†

2.2.2 Constrained Linkages

We will make use of a number of special-purpose “constraints” or “annotations” that may be attached to linkages to artificially modify their behavior, such as “rigid constraints” that “rigidify” a subgraph into a chosen configuration while allowing the rest of the linkage to move freely. These annotations do not affect the linkage itself; instead, they merely indicate which configurations of the linkage they consider acceptable. The language of constraints allows us to separate a desired *effect* from the *implementation* or *construction* that enforces that effect. For example, Sections 2.4 through 2.6 develop three different techniques to force subgraphs to remain rigid—three different “implementations” of the rigidifying constraint—allowing a majority of the work, namely the Main Theorem (Theorem 2.2.13), to be reused in all three contexts. We now define constraints in general, and the rigid constraint in particular, more formally.

Definition 2.2.6. A **constraint** Con on an abstract linkage \mathcal{L} is specified by a subset of the configuration space, $\text{Con} \subseteq \text{Conf}(\mathcal{L})$, and we say the configurations $C \in \text{Con}$ **satisfy**

*In particular, if \mathcal{L} is not realizable, its global minimum feature size is $+\infty$. Note also that, if $\text{Conf}(\mathcal{L})$ is compact, this infimum is achieved and is strictly greater than 0.

†Recall that all graphs are connected, so $\deg(v) \geq 1$.

constraint Con . A **constrained linkage** \mathcal{L} is an abstract linkage \mathcal{L}_0 together with a finite set K of constraints on \mathcal{L}_0 , and the **constrained configuration space** is defined as $\text{Conf}(\mathcal{L}) := \text{Conf}(\mathcal{L}_0) \cap \bigcap_{\text{Con} \in K} \text{Con}$. In other words, constrained linkage \mathcal{L} simply ignores any configurations of \mathcal{L}_0 that don't satisfy all of its constraints.

All terms discussed in Section 2.2.1—realizability, rigidity, global rigidity, etc.—apply equally well to constrained linkages via their *constrained* configuration space.

Definition 2.2.7. A **rigid constraint** $\text{RigidCon}_{\mathcal{L}}(H, C_H)$ on a linkage $\mathcal{L} = (G, W, P)$ is specified by a connected subgraph* $H \subseteq G$ together with a configuration C_H of H . A configuration $C \in \text{Conf}(\mathcal{L})$ satisfies the rigid constraint when C induces a configuration $C|_H$ on H that is congruent to the given C_H , i.e., differs only by a (possibly orientation-reversing) Euclidean transformation. When a constrained linkage \mathcal{M} possesses constraint $\text{RigidCon}_{\mathcal{M}}(H, C_H)$, we say (H, C_H) is a **rigidified subgraph** of \mathcal{M} . A constrained linkage all of whose constraints are rigid constraints is called a **partially rigidified linkage**.

2.2.3 Drawing with Linkages and Graphs

Definition 2.2.8 (Linkage Trace and Drawing). For a linkage \mathcal{L} and a tuple $X = (v_1, \dots, v_k)$ of distinct vertices of \mathcal{L} , the **trace** of X is defined as the image $\pi_X(\text{Conf}(\mathcal{L})) \subset (\mathbb{R}^2)^k$, where π_X is the projection map sending a configuration $C \in \text{Conf}(\mathcal{L})$ to

$$\pi_X(C) := (C(v_1), \dots, C(v_k)).$$

A linkage (\mathcal{L}, X) is said to **draw**[†] its trace, and a set $R \subseteq (\mathbb{R}^2)^k$ is **drawable (by a linkage)** if it can be expressed as the trace of some k vertices of a linkage.

We single out some drawings as particularly nice:

Definition 2.2.9 (Rigid Drawing). Say (\mathcal{L}, X) draws its trace **rigidly** if the map π_X has finite fibers, i.e., for any $p \in \pi_X(\text{Conf}(\mathcal{L}))$, there are only finitely many configurations $C \in \text{Conf}(\mathcal{L})$ with $\pi_X(C) = p$.

In particular, if p is isolated in $\pi_X(\text{Conf}(\mathcal{L}))$, then any configuration C with $\pi_X(C) = p$ is rigid, because the discrete set $\pi_X^{-1}(p)$ contains no nonconstant continuous paths.

Definition 2.2.10 (Continuous Drawing). Say (\mathcal{L}, X) draws its trace **continuously** if the map π_X has the **path lifting property**: for any configuration $C \in \text{Conf}(\mathcal{L})$ and path $\gamma : [0, 1] \rightarrow \pi_X(\text{Conf}(\mathcal{L}))$ in the trace starting at $\gamma(0) = \pi_X(C)$, there is a path $\gamma' : [0, 1] \rightarrow \text{Conf}(\mathcal{L})$ starting at $\gamma'(0) = C$ and lifting γ , i.e., $\gamma = \pi_X \circ \gamma'$.

In particular, if a point $p \in \pi_X(\text{Conf}(\mathcal{L}))$ is *not* isolated, then any configuration C with $\pi_X(C) = p$ is *not* rigid, because a nontrivial continuous path beginning at p can be lifted to a nontrivial path beginning at C . We are especially interested in cases where (\mathcal{L}, X) draws both continuously and rigidly; these concepts were introduced in [2] for their usefulness in proving computational hardness of linkage rigidity, and we rely on them for similar purposes. We make use of an even stronger notion as well:

Definition 2.2.11 (Perfect Drawing). If the map π_X is a *homeomorphism* between $\text{Conf}(\mathcal{L})$ and the trace, we say (\mathcal{L}, X) draws **perfectly**.

*Our notion of “subgraph” requires the edge-lengths of H to agree with those in G .

†This “drawing” need not be continuous. For example, the linkage may have a disconnected trace.

Definition 2.2.12 (Linkage Simulation). When (\mathcal{L}, X) draws precisely the full configuration space $\text{Conf}(\mathcal{M})$ of another linkage \mathcal{M} , we say that (\mathcal{L}, X) **simulates** \mathcal{M} . It may **continuously, rigidly, or perfectly simulate** \mathcal{M} if it draws $\text{Conf}(\mathcal{M})$ in this manner.

2.2.4 Specification of the Main Theorem

For a collection $F = \{f_1, \dots, f_s\}$ of polynomials in $\mathbb{R}[x_1, y_1, \dots, x_m, y_m] = \mathbb{R}[\vec{x}\vec{y}]$, the **algebraic set** (or **algebraic variety**) defined by F is the set of common zeros,

$$Z(F) := \{\vec{x}\vec{y} \in \mathbb{R}^{2m} \mid f_1(\vec{x}\vec{y}) = \dots = f_s(\vec{x}\vec{y}) = 0\}.$$

The primary technical construction in this paper builds a globally noncrossing, partially rigidified linkage $\mathcal{L}(F)$ that draws precisely the algebraic set $Z(f_1, \dots, f_s) \subseteq \mathbb{R}^{2m}$, or at least a bounded piece thereof, up to a translation of \mathbb{R}^{2m} . Why is the translation necessary? Without it, some algebraic sets would require the drawing vertices in X to collocate at some or all of the linkage’s configuration space*, precluding the possibility of global noncrossing.

We are now prepared to precisely specify the properties of this construction, from which the results listed in Table 2.1 follow as corollaries. We thoroughly detail these properties here, so that the corollaries may be derived solely from Theorem 2.2.13’s statement without referring to the specifics of its proof (with one small exception, discussed in Section 2.7.7). This also allows for maximal reuse: the commonalities in our arguments for our three linkage contexts—unconstrained globally noncrossing linkages in Section 2.4, unit-distance linkage in Section 2.5, and matchstick linkages in Section 2.6—have been unified and generalized into Theorem 2.2.13, so only features unique to each context need be discussed in Sections 2.4–2.6.

The Main Theorem is divided into three parts because it must be used in subtly different ways by the four types of results we seek. Hardness of realizability requires a polynomial-time construction of an abstract linkage that draws $Z(f_1, \dots, f_s)$ *without knowing whether the resulting configuration space is empty*, whereas proving hardness of rigidity and global rigidity requires the polynomial-time construction of a linkage *together with a known configuration*. We thus separate these into different Parts of Theorem 2.2.13 with slightly different assumptions about the input polynomials f_i (Part II for realizability, Part III for rigidity and global rigidity). When proving universality, we must prove *existence* of a linkage to draw any compact semialgebraic set, but the coefficients of the polynomials defining this set may be non-rational or non-algebraic, as might the edge-lengths and coordinates of the resulting linkage, so we isolate this in Part I, away from algorithmic and efficiency concerns.

We measure the “size” of polynomials naïvely: a polynomial $f \in \mathbb{R}[x_1, y_1, \dots, x_m, y_m]$ with total degree d is specified by $\#\text{Coeffs}(f) := \binom{2m+d}{d} = \text{poly}(m^d, d^d)$ real coefficients, using dense representation.[†] If f ’s coefficients are integers with maximum magnitude M , we record its size as $\text{Size}(f) := M \cdot \#\text{Coeffs}(f) = \text{poly}(m^d, d^d, M)$ unary digits (not binary!). For a set $F = \{f_1, \dots, f_s\}$ of s polynomials in $\mathbb{R}[x_1, y_1, \dots, x_m, y_m]$ with maximum total degree d , we set $\#\text{Coeffs}(F) := s \cdot \binom{2m+d}{d} = \text{poly}(m^d, d^d, s)$, and if these coefficients are integers with maximum magnitude M , then $\text{Size}(F) := s \cdot M \cdot \binom{2m+d}{d} = \text{poly}(s, M, m^d, d^d)$.

Theorem 2.2.13. Part I. *Take as input a collection of polynomials $F = \{f_1, \dots, f_s\}$, each in $\mathbb{R}[x_1, y_1, \dots, x_m, y_m]$. Then we may construct a partially rigidified linkage $\mathcal{L} = \mathcal{L}(F)$ that*

*For example, two distinct vertices that draw the trace $\{((0,0), (0,0))\} \subset (\mathbb{R}^2)^2$ must meet.

[†]The measure $\#\text{Coeffs}(f)$ does *not* count the nonzero coefficients of f ; instead, it counts the total number of monomials that have total degree at most that of f .

draws, up to translation, a bounded portion of the algebraic set $Z(F)$: specifically, there is a translation T on \mathbb{R}^{2m} and a subset X of m vertices of \mathcal{L} such that

$$T(Z(F) \cap [-1, 1]^{2m}) \subseteq \pi_X(\text{Conf}(\mathcal{L})) \subseteq T(Z(F)).$$

Furthermore,

1. Vertices X draw this trace rigidly and continuously.
2. The number of vertices and edges in \mathcal{L} is $\text{poly}(\#\text{Coeffs}(F))$.
3. Each edge of \mathcal{L} has length $\Omega(1)$, and \mathcal{L} is globally noncrossing with global minimum feature size $\Omega(1)$.
4. For each constraint $\text{RigidCon}_{\mathcal{L}}(H, C_H)$ on \mathcal{L} , H is a tree that connects to $G \setminus H$ precisely at leaves of H , and configuration C_H has all edges parallel to the x - or y -axes. Each edge of G is contained in at most one rigidified subgraph (H, C_H) .
5. There is a combinatorial embedding σ of G such that every configuration $C \in \text{Conf}(\mathcal{L})$ agrees with σ . Furthermore, if v is not an internal vertex of any constrained tree H , then for each angle chain Λ at v , angle $\angle C(\Lambda)$ lies strictly between 60° and 240° .
6. Linkage \mathcal{L} has precisely $|P| = 3$ pinned vertices, which belong to one of the rigidified trees (H, C_H) and are not collinear in C_H .

Part II. If polynomials f_i have integer coefficients, we may bound the complexity of \mathcal{L} as follows:

7. All edge-lengths in \mathcal{L} are rational, with numerators bounded by $\text{poly}(\text{Size}(F))$ and denominators bounded by $O(1)$.
8. Constrained linkage \mathcal{L} , the set X of vertices, translation T , and combinatorial embedding σ may be constructed from F deterministically in time $\text{poly}(\text{Size}(F))$.

Part III. Finally, if the polynomials f_i each satisfy $f_i(\vec{0}) = 0$, we may additionally compute an initial configuration C_0 satisfying:

9. All coordinates of C_0 are rational numbers with magnitude bounded by $\text{poly}(\text{Size}(F))$ and with $O(1)$ denominators.
10. C_0 is the only configuration of \mathcal{L} that projects to $T(\vec{0}) \in \pi_X(\text{Conf}(\mathcal{L}))$.
11. C_0 may also be computed deterministically in time $\text{poly}(\text{Size}(F))$.

2.2.5 Roadmap

The rest of the paper is organized as follows: After gathering a few preliminary facts about real (semi)algebraic sets and the class $\exists\mathbb{R}$ in Section 2.3, we use the Main Theorem (Theorem 2.2.13) to prove that graph realizability, rigidity, and global rigidity are $\exists\mathbb{R}$ -complete or $\forall\mathbb{R}$ -complete, and that linkages are universal at drawing semialgebraic sets, in each of three separate contexts: for globally noncrossing graphs/linkages in Section 2.4, for unit-distance (or $\{1, 2\}$ -distance) graphs/linkages in Section 2.5, and for matchstick graphs/linkages in Section 2.6. After all of that, in Section 2.7, we finally provide the gory details of the Main Construction itself.

2.3 Preliminaries on Semialgebraic Sets and $\exists\mathbb{R}$

2.3.1 Semialgebraic Sets as Projections of Algebraic Sets

Here we prove some useful elementary facts about representing certain semialgebraic sets as projections of (semi)algebraic sets having specific forms, which will be used in later proofs of linkage universality. For a general introduction to real semialgebraic geometry, see [25].

A **basic semialgebraic set** in \mathbb{R}^k is a set of the form $\{\vec{x} \in \mathbb{R}^k \mid f_1(\vec{x}) = \dots = f_s(\vec{x}) = 0, g_1(\vec{x}) > 0, \dots, g_r(\vec{x}) > 0\}$ for polynomials f_i and g_j ; any **semialgebraic set** can be written as a finite union of basic semialgebraic sets [25, Prop. 2.1.8]; in fact, this is one of several equivalent ways to define semialgebraic sets.

Lemma 2.3.1. *Any bounded semialgebraic set $R \subset \mathbb{R}^k$ can be expressed as the projection of some bounded basic semialgebraic set onto the first k coordinates.*

Proof. We may assume R is nonempty. Any semialgebraic set can be written as a finite union of basic semialgebraic sets [25, Prop. 2.1.8], so write $R = \bigcup_{j=1}^t R_j$, where each R_j is a nonempty basic semialgebraic set. Let $\vec{x}, \vec{x}_1, \dots, \vec{x}_t$ each denote a variable point in \mathbb{R}^k , and define $R' \subset (\mathbb{R}^k)^{t+1}$ as the *basic* semialgebraic set defined by the conditions $x_j \in R_j$ for $1 \leq j \leq t$, as well as

$$f(\vec{x}, \vec{x}_1, \dots, \vec{x}_t) := \prod_{j=1}^t |\vec{x} - \vec{x}_j|^2 = 0.$$

This last equation exactly stipulates that $\vec{x} = \vec{x}_j$ for some $1 \leq j \leq t$, so $\pi(R') = \bigcup_{j=1}^t R_j = R$, where π denotes the projection onto the coordinates of \vec{x} . Set R' is bounded because each R_j is bounded. \square

Lemma 2.3.2. *Any bounded semialgebraic set $R \subset \mathbb{R}^k$ can be expressed as the projection of a bounded set of the form $\{\vec{x} \in \mathbb{R}^m \mid f_1(\vec{x}) = \dots = f_s(\vec{x}) = 0, g_1(\vec{x}) \neq 0, \dots, g_r(\vec{x}) \neq 0\}$, for polynomials f_i and g_j .*

Proof. By Lemma 2.3.1, we may assume R is a basic semialgebraic set, $R = \{\vec{x} \in \mathbb{R}^k \mid p_1(\vec{x}) = \dots = p_s(\vec{x}), q_1(\vec{x}) > 0, \dots, q_r(\vec{x}) > 0\}$. Because R is bounded, by scaling all of the polynomials p_i and q_j we may assume $|p_i(\vec{x})| < 1$ and $|q_j(\vec{x})| < 1$ for all $\vec{x} \in R$. Now introduce new real variables $\vec{a} = (a_1, \dots, a_r)$ and define

$$R' = \{(\vec{x}, \vec{a}) \in \mathbb{R}^{k+r} \mid p_1(\vec{x}) = \dots = p_s(\vec{x}) = 0, \\ q_1(\vec{x}) = a_1^2, a_1 \neq 0, \dots, q_s(\vec{x}) = a_s^2, a_s \neq 0\}.$$

This set is contained in $R \times [-1, 1]^r$ and is therefore bounded, and R is the projection of R' onto the coordinates of \vec{x} . \square

A **basic closed semialgebraic set** has the form $\{\vec{x} \in \mathbb{R}^k \mid f_1(\vec{x}) \geq 0, \dots, f_s(\vec{x}) \geq 0\}$; if this set is also bounded, we call it a **basic compact semialgebraic set**.

Lemma 2.3.3. *Any compact semialgebraic set $R \subset \mathbb{R}^k$ can be expressed as the projection of some compact algebraic set onto the first k coordinates.*

Proof. Compact semialgebraic set R can be written as a finite union of basic compact semialgebraic sets [25, Thm. 2.7.2], and the same proof used in Lemma 2.3.1 shows that R can be written as a coordinate projection of a basic compact semialgebraic set. (The condition

$f = 0$ can be expressed as $f \geq 0$ and $-f \geq 0$.) So it suffices to show that a compact *basic* semialgebraic set is a coordinate projection of a compact algebraic set.

So assume R is a compact basic semialgebraic set,

$$R = \{\vec{x} \in \mathbb{R}^n \mid f_1(\vec{x}) \geq 0, \dots, f_s(\vec{x}) \geq 0\}.$$

By scaling polynomials f_i we may assume $|f_i(\vec{x})| \leq 1$ for all $\vec{x} \in R$ (this uses R 's compactness). Choose new variables $\vec{y} \in \mathbb{R}^s$ and $\vec{z} \in \mathbb{R}^s$, and define $R' = \{(\vec{x}, \vec{y}, \vec{z}) \mid f_i(\vec{x}) = y_i^2 \text{ and } y_i^2 + z_i^2 = 1 \text{ for } 1 \leq i \leq s\}$. If π is the projection onto coordinates of \vec{x} , then one may check that $\pi(R') = R$. Furthermore, R' is bounded and therefore compact. This completes the proof. \square

2.3.2 Existential Theory of the Reals

The class $\exists\mathbb{R}$ is the complexity class consisting of all problems polynomially-reducible to the Existential Theory of the Reals (ETR), which is the language consisting of true formulas of the form $(\exists x_1, \dots, x_n \in \mathbb{R})\varphi(x_1, \dots, x_n)$, where φ is a quantifier-free formula over variables x_1, \dots, x_n using $\wedge, \vee, \neg, +, -, \times, 0, 1$, and $=$ as primitives. It is known that $\text{NP} \subseteq \exists\mathbb{R} \subseteq \text{PSPACE}$, though neither inclusion is known to be strict. The first inclusion is simple, as SAT may be encoded with polynomial constraints, so any problem complete for $\exists\mathbb{R}$ is automatically NP-hard. The second inclusion is a nontrivial theorem of Canny [27], and this is the tightest known upper bound on the hardness of $\exists\mathbb{R}$.

We now prove the $\exists\mathbb{R}$ hardness of two problems used in the forthcoming reductions. The problem H_2N of determining whether a collection of homogeneous polynomials $f_1, \dots, f_s \in \mathbb{Z}[x_1, \dots, x_k]$ have a nonzero common root in \mathbb{R}^k was introduced in [62] and was shown to be $\exists\mathbb{R}$ -complete in [74], even when all polynomials have degree 4. We wish to reduce from a stronger version of this problem where, additionally, all coefficients are in $\{0, \pm 1, \pm 2\}$. To prove hardness of this problem, only a slight modification of Schaefer's original argument is necessary. Our starting point is the following lemma:

Lemma 2.3.4. *The problem of determining whether a collection of polynomials $f_1, \dots, f_s \in \mathbb{Z}[x_1, \dots, x_n]$ of total degree at most 2 and with coefficients in $\{0, \pm 1, \pm 2\}$ have a common solution in \mathbb{R}^n is $\exists\mathbb{R}$ -complete.*

Proof. The problem is a subproblem of ETR, so it is certainly in $\exists\mathbb{R}$. For hardness, we reduce from the PointConfiguration problem: given a choice of orientation (clockwise, counterclockwise, or collinear) for each triple $1 \leq i < j < k \leq n$, determine whether there exists a configuration of points p_1, \dots, p_n in the plane so that each triangle $p_i p_j p_k$ has the specified configuration. This is equivalent (in fact, projectively dual) to the more well-known Stretchability problem, shown to be $\exists\mathbb{R}$ -complete by Mnëv [69]; see also [77].

If we write $p_i = (p_{i,1}, p_{i,2})$ in coordinates, then we may write

$$2 \cdot \text{area}(p_i p_j p_k) = \det(p_i, p_j, p_k) := \det \begin{pmatrix} p_{i,1} & p_{i,2} & 1 \\ p_{j,1} & p_{j,2} & 1 \\ p_{k,1} & p_{k,2} & 1 \end{pmatrix}$$

(using signed area), which is a homogeneous, degree 2 polynomial in the coordinate variables with coefficients ± 1 . For the triples with collinear configuration, we simply include the polynomial $\det(p_i, p_j, p_k) = 0$. For those with counterclockwise orientation, include the polynomials $\det(p_i, p_j, p_k) = a_{i,j,k}^2$ and $a_{i,j,k} b_{i,j,k} = 1$, for new variables $a_{i,j,k}$ and $b_{i,j,k}$.

Similarly, for clockwise triples, include $\det(p_i, p_j, p_k) = -a_{i,j,k}^2$ and $a_{i,j,k}b_{i,j,k} = 1$. It is clear that the point configuration is realizable if and only if the resulting system of $\{1, 0, -1\}$ -coefficient, degree ≤ 2 polynomials has a real solution. \square

Theorem 2.3.5. *The problem H_2N remains $\exists\mathbb{R}$ -complete even when the input polynomials have degree 4 and all coefficients lie in $\{0, \pm 1, \pm 2\}$.*

Proof. The proof by Schaefer [74, Lem. 3.9, Cor. 3.10] for the hardness of H_2N with degree 4 starts with a collection of degree- ≤ 2 polynomials and transforms them into a collection of homogeneous degree 4 polynomials that have a nontrivial solution if and only if the original collection has any solution. These transformations, when applied without modification to polynomials with coefficients in $\{0, \pm 1, \pm 2\}$, return a collection of homogeneous, degree 4 polynomials whose coefficients also lie in $\{0, \pm 1, \pm 2\}$. So this result follows from Schaefer's proof paired with Lemma 2.3.4. \square

We denote a related, well-known problem by **CommonZero**: determine whether a collection of (not necessarily homogeneous) polynomials $f_1, \dots, f_s \in \mathbb{Z}[x_1, \dots, x_m]$ has any common zero.

Theorem 2.3.6. *The CommonZero problem is $\exists\mathbb{R}$ -complete, even when the given polynomials $f_1, \dots, f_s \in \mathbb{Z}[x_1, \dots, x_m]$ have total degree at most 4, all coefficients are in $\{0, \pm 1, \pm 2\}$, and all common zeros, if any, are promised to lie in $[-1, 1]^m$.*

Proof. We reduce from H_2N , as strengthened in Lemma 2.3.5: we are given an instance $F = \{f_1, \dots, f_s\}$ of H_2N with degree 4 polynomials whose coefficients lie in $\{0, \pm 1, \pm 2\}$. Return the instance $F' = F \cup \{g\}$ of **CommonZero**, where $g(x_1, \dots, x_m) = x_1^2 + \dots + x_m^2 - 1$. Polynomial g guarantees that $Z(F') \subset \overline{B}(\vec{0}, 1) \subset [-1, 1]^m$, and all coefficients in F' are still in $\{0, \pm 1, \pm 2\}$, so F' has the required form.

If $Z(F)$ contains only $\vec{0}$, then $Z(F')$ is empty. On the other hand, if $\vec{a} \in Z(F) \setminus \{\vec{0}\}$, then $\vec{a} / \|\vec{a}\| \in Z(F')$. So the reduction is correct. \square

2.4 Globally Noncrossing Graphs and Linkages

Using the statement of the Main Theorem (Theorem 2.2.13) and the hardness results from Section 2.3.2, we show here that deciding realizability, rigidity, or global rigidity of a globally noncrossing graph/linkage is $\exists\mathbb{R}$ -complete or $\forall\mathbb{R}$ -complete, and that globally noncrossing linkages can draw any compact semialgebraic set in the plane.

2.4.1 Rigidifying Polygons

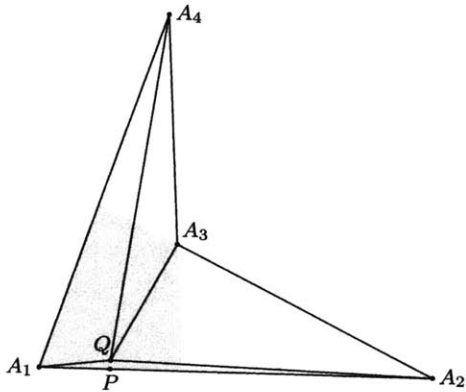
To simulate a rigidified subtree, we will construct a globally rigid graph in the shape of a slight thickening of the tree. To that end, we provide here a general method that constructs a globally rigid triangulation, with Steiner points, of any simple polygon.

Lemma 2.4.1. *Any simple quadrilateral $A = A_1A_2A_3A_4$ has a globally rigid triangulation with four triangles and one Steiner point as in Figure 2-1a.*

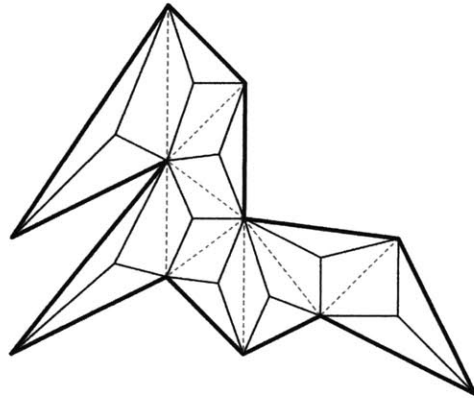
Proof. If any of the four angles of A is $\geq 180^\circ$, relabel so that this angle is at A_3 . This means any point in the interior of A within some distance $d > 0$ from A_1 is visible to all four of A 's vertices inside the quadrilateral.

Let S be the point on side A_1A_2 at distance $d/2$ from A_1 , so that the three angles $\angle A_2PA_3$, $\angle A_3PA_4$ and $\angle A_4PA_1$ lie strictly between 0° and 180° . By continuity, there is a point Q near P in the interior of A such that $\angle A_1QA_2$ is close to 180° and the three angles $\angle A_2QA_3$, $\angle A_3QA_4$ and $\angle A_4QA_1$ lie strictly between $180^\circ - \angle A_1QA_2$ and 180° . We may also assume Q is close to A_1 and is therefore visible to all four vertices of A .

We claim that the triangulation T using Steiner point Q is globally rigid. If not, then by Kawasaki's criterion [56] for flat-foldable single vertex crease patterns, it must be the case that some subset of the four angles at Q sum to 180° . But Q was chosen to ensure that this condition is false, since adding $\angle A_1QA_2$ to any of the other three angles results in more than 180° . \square



(a) All points in the gray region are visible to the four vertices of quadrilateral A . Any point Q close enough to P renders this 4-triangle triangulation of A globally rigid.



(b) Rigidifying each quadrilateral in this decomposition using Lemma 2.4.1 rigidifies the entire graph.

Lemma 2.4.2. *If A is a simple n -sided polygon, we may construct a triangulation of A with $2n - 5$ Steiner points that is globally rigid as a configured graph.*

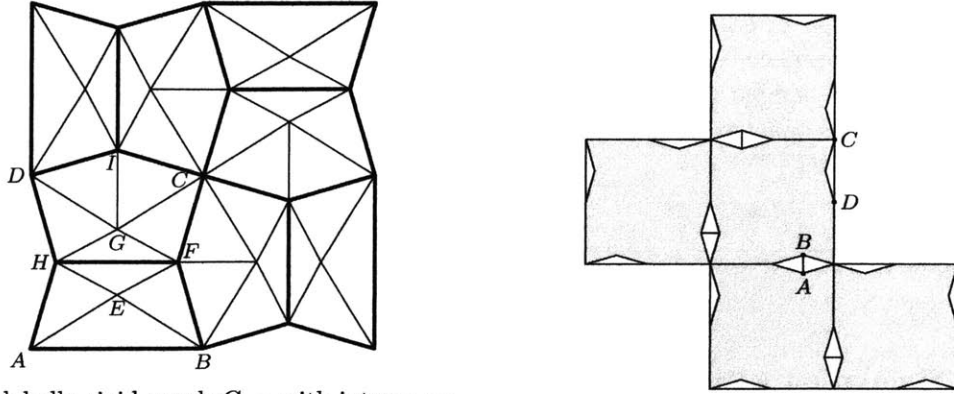
Proof. Compute any triangulation T_1 of A by $n - 3$ interior diagonals, and subdivide each triangle at its centroid to obtain triangulation T_2 . Delete the original $n - 2$ interior diagonals from T_2 to obtain a subdivision T_3 of polygon A into triangles and quadrilaterals, as illustrated in Figure 2-1b. Finally, apply Lemma 2.4.1 to each quadrilateral in T_3 to obtain the final triangulation T . The three triangles or quadrilaterals meeting at each centroid are individually globally rigid, so the union of these three pieces is globally rigid again by Kawasaki's criterion: three creases around a vertex, no two collinear, are insufficient for a nontrivial single-vertex flat folding. Applying this reasoning around each centroid shows that the entire triangulation T is globally rigid, as desired. \square

This Lemma will be used when simulating rigidified subtrees in Section 2.4.3.

2.4.2 Rigidifying Polyominoes

When the partially rigidified tree (H, C_H) that we wish to simulate has *integer* coordinates, we will use a more refined method that uses only rational coordinates and rational edge lengths whose numerators and denominators have constant size. To accomplish this, we show in this section that any polyomino, after scaling up by a factor of 1440, can be rigidified

with only integer coordinates and constant-sized integer edge-lengths. (A **polyomino** is a polygon formed by an edge-connected set of squares in the standard unit-square tiling of \mathbb{R}^2 .) We begin by rigidifying a single square:



(a) A globally rigid graph G_{cell} with integer coordinates and integer edge-lengths in the shape of a 1440×1440 square with indents.

(b) Multiple copies of G_{cell} can be joined into a globally rigid polyomino.

Figure 2-2: Any connected polyomino of 1440×1440 squares can be turned into a globally rigid graph with integer coordinates and integer edge lengths.

Lemma 2.4.3. *The configured graph G_{cell} shown in Figure 2-2a, which has the shape of a 1440×1440 square with small **indents** on the edges, is globally rigid. The vertex coordinates and edge lengths are all integers.*

Proof. To specify G_{cell} graph in more detail, the labeled vertices have coordinates

$$\begin{aligned} A = (0, 0), \quad B = (720, 0), \quad C = (720, 720), \quad D = (0, 720), \quad E = (360, 224), \\ F = (615, 360), \quad G = (360, 496), \quad H = (105, 360), \quad I = (360, 825), \end{aligned}$$

(in particular, pentagon $HFCID$ and quadrilateral $ABFH$ have a vertical line of symmetry), and the edge-lengths are

$$AB = 720, \quad AE = 424, \quad EF = 289, \quad GI = 329, \quad HF = 510.$$

The rest of the coordinates may be computed by the 90° -degree rotational symmetry of G_{cell} , and all distinct edge-lengths are listed above. Note that A, E, F are not quite collinear, and similarly for H, G, C .

To show that G_{cell} is globally rigid, we again make repeated use Kawasaki's criterion. First, the five triangles forming pentagon $HFCID$ (with Steiner point G) form a globally rigid subgraph: indeed it may be checked that $\angle HGF + \angle FGC = 180^\circ + \arcsin \frac{60}{901} > 180^\circ$ and $\angle DGF + \angle CGI = 180^\circ + \arcsin \frac{525}{15317} > 180^\circ$, so any subset of the five angles at G that includes $\angle HGF$ cannot add to exactly 180° . The four triangles forming quadrilateral $ABFH$ form a globally rigid subgraph as well, because the crease pattern around E is congruent to a subset of the crease pattern around G . The rotationally symmetric copies of pentagon $HFCID$ and $ABFG$ are likewise globally rigid. Finally, the quadrilateral and two pentagons meeting at F are globally rigid together because degree-3 crease patterns have no nontrivial flat foldings, and applying this reasoning four times around the square shows that all of G_{cell} is globally rigid. \square

Lemma 2.4.4. *Any connected polyomino P made of 1440×1440 squares can be triangulated (allowing Steiner points and edge subdivision) into a globally rigid triangulation that has integer coordinates and constant-sized integer edge lengths.*

Proof. Place a copy of G_{cell} in each cell of polyomino P , swapping orientation for every other cell so that adjacent cells have aligned indents. A short edge of length 210 within the indents—such as edge AB in Figure 2-2b—renders each pair of adjacent cells globally rigid. Because P is an edge-connected monomino, the whole assembly is thus globally rigid. Finally, indents along P 's boundary can be covered with edges of length 360, as in Figure 2-2b. \square

We show in the next section how to use this method to simulate integer-length rigidified subtrees.

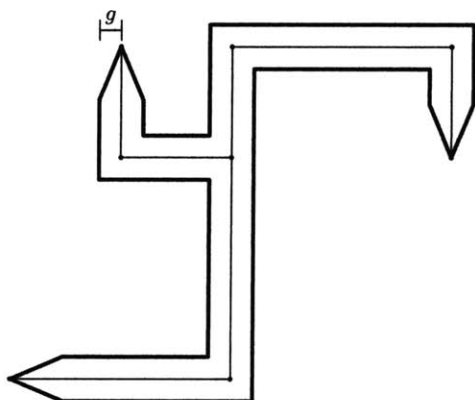
2.4.3 Simulation with Globally Noncrossing Linkages

Construction 2.4.5. *Use notation as in Theorem 2.2.13.*

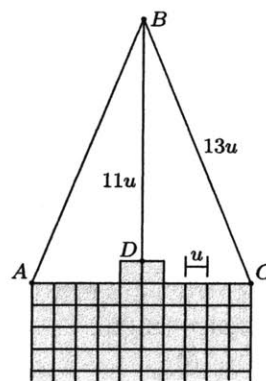
If the hypotheses of Part I hold, we may construct a globally noncrossing linkage $\mathcal{M} = \mathcal{M}(F)$ (without constraints) that perfectly simulates the constrained linkage $\mathcal{L}(F)$.

If Part II also holds, then \mathcal{M} may be constructed from \mathcal{L} in time $\text{poly}(\text{Size}(F))$, and furthermore, each edge of \mathcal{M} has rational length with numerators and denominators of size $O(1)$.

Finally, if Part III additionally holds, the configuration of \mathcal{M} corresponding to $C_0 \in \text{Conf}(\mathcal{L})$ has rational coordinates bounded in magnitude by $\text{poly}(\text{Size}(F))$ and with denominators of size $O(1)$.



(a) Thickening a partially rigidified subtree (H, C_H) by radius g , with wedges at each leaf.



(b) Rigidifying the thickened tree using a polyomino and 5-12-13 triangles at the leaves.

Figure 2-3: Thickening a rigidified subtree (H, C_H) into a globally rigid polygon.

Proof. Suppose all edges of \mathcal{L} have length at least e and that \mathcal{L} has global minimum feature size f for positive numbers $e, f = \Omega(1)$ (by Property 3). Consider one of the rigid constraints $\text{RigidCon}(H, C_H)$ on \mathcal{L} , and draw a polygon P that thickens (H, C_H) by distance $g = \min(e/4, f/4)$ in each direction, with angled wedges smaller than 60° at each leaf vertex, as shown in Figure 2-3a. Now apply the construction of Lemma 2.4.2 to this polygon P to

obtain a triangulation T , and replace subgraph H of \mathcal{L} with this triangulation T . Let \mathcal{M} be the linkage that results after performing this replacement for each rigid constraint on \mathcal{L} .

To see that \mathcal{M} is globally noncrossing, note that property 5 of Theorem 2.2.13 guarantees that the 60° wedges at leaf nodes do not intersect locally around their shared vertices, and because g is less than half the minimum feature size of any configuration of \mathcal{L} , the thickened trees do not intersect elsewhere.

If Part II holds, we construct globally rigid graph T from rigidified tree (H, C_H) more carefully, by building a rigidified polyomino with Lemma 2.4.4. Begin by choosing the thickening amount $g \leq \min(e/4, f/4)$ such that each edge length in (H, C_H) is an integer multiple of g ; this g can be chosen to be rational with constant-sized numerator and denominator by Property 7. Thicken (H, C_H) by distance g into a polygon P as above, and at each leaf of (H, C_H) , form a wedge with angle $2 \arcsin \frac{5}{13} < 60^\circ$. Now we can fill polygon P with a polyomino made from cells of side-length $u = g/5$, where the wedge at each leaf is attached with three bars of length $13u$, $11u$, and $13u$ respectively, as shown in Figure 2-2b. By rigidifying the polyomino with Theorem 2.4.4, the resulting graph T is a globally rigid thickening of tree (H, C_H) that has constant-sized rational edge-lengths and rational coordinates, all with constant-sized denominators. This proves both the Part II and Part III claims. \square

2.4.4 Hardness and Universality of Globally Noncrossing Linkages

We may finally prove the desired hardness and universality results about globally noncrossing graphs and linkages.

Theorem 2.4.6 (Hardness of Globally Noncrossing Realizability). *Deciding whether a given abstract weighted graph \mathcal{G} is realizable, even when \mathcal{G} is promised to be globally noncrossing and has constant-sized integer edge-lengths, is strongly $\exists\mathbb{R}$ -complete.*

Proof. We reduce from **CommonZero** (Theorem 2.3.6). Given an instance $F = \{f_1, \dots, f_s\}$, apply construction 2.4.5 (Part II) to obtain a globally noncrossing linkage \mathcal{L} such that $\text{Conf}(\mathcal{L})$ is isomorphic to $Z(F) \cap [-1, 1]^{2m} = Z(F)$ and the scaled linkage $t \cdot \mathcal{L}$ has Pythagorean edge-lengths of polynomial magnitude. In particular, F has a common root if and only if $\text{Conf}(\mathcal{L})$ is nonempty, i.e., if and only if $t \cdot \mathcal{L}$ is realizable. Finally, by Property 6, if we remove the three pins of linkage $t \cdot \mathcal{L}$ to form a weighted graph \mathcal{G} , then all configurations of \mathcal{G} are congruent to configurations of $t \cdot \mathcal{L}$, i.e., $t \cdot \mathcal{L}$ is realizable if and only if \mathcal{G} is realizable. \square

Theorem 2.4.7 (Hardness of Noncrossing Rigidity and Global Rigidity). *Deciding whether a given configured weighted graph (\mathcal{G}, C_0) is rigid, when \mathcal{G} is promised to be globally noncrossing (so in particular, C_0 is noncrossing) and C_0 has integer coordinates and constant-sized integer edge-lengths, is strongly $\forall\mathbb{R}$ -complete. It remains $\forall\mathbb{R}$ -complete if “rigid” is replaced by “globally rigid”.*

Proof. We reduce from $\mathbf{H}_2\mathbf{N}$ as strengthened in Lemma 2.3.5 in Section 2.3.2, so suppose we’re given a family of homogeneous polynomials $F = \{f_1, \dots, f_s\}$ in $\mathbb{Z}[x_1, y_1, \dots, x_m, y_m]$ with constant-sized integer coefficients and homogeneous degree 4 (and in particular no constant term). We may use Construction 2.4.5 (Part III) to build a globally noncrossing configured linkage (\mathcal{M}, C_0) that continuously and rigidly draws a trace $\pi_X(\text{Conf}(\mathcal{M}))$ satisfying $Z(F) \cap [-1, 1]^{2m} \subseteq T(\pi_X(\text{Conf}(\mathcal{M})) \subseteq Z(F)$ for some translation T , together with an integer $t \leq \text{poly}(\text{Size}(F))$ where all coordinates of C_0 are rationals with magnitude bounded by $\text{poly}(\text{Size}(F))$ and with denominators dividing t . The result of this reduction will be the

configured graph \mathcal{G} formed by scaling (\mathcal{M}, C_0) by t and removing the three pins. Configured graph \mathcal{G} indeed has polynomially-bounded integer coordinates.

To verify the validity of this reduction, suppose first that $Z(F)$ contains some nonzero point \vec{a} . Then $Z(F)$ contains the entire path $t \mapsto t \cdot \vec{a}$ starting at $\vec{0}$, so $\vec{0}$ is not isolated in $Z(F)$, i.e., $T(\vec{0})$ is not isolated in $\pi_X(\text{Conf}(\mathcal{M}))$. Because \mathcal{M} draws continuously, (\mathcal{M}, C_0) is not rigid (as a linkage), and therefore \mathcal{G} is not rigid (as a graph). On the other hand, if $Z(F) = \{\vec{0}\}$ then $\pi_X(\text{Conf}(\mathcal{M}))$ contains only the single point $\pi_X(C_0) = T(\vec{0})$, and by property 10 of Theorem 2.2.13 (uniqueness), it follows that $\text{Conf}(\mathcal{M}) = \{C_0\}$, i.e., \mathcal{M} is both rigid and globally rigid (as a linkage). By property 6, all configurations of \mathcal{G} are Euclidean transformations of configurations of $t \cdot \mathcal{M}$, and so \mathcal{G} is rigid and globally rigid (as a graph). \square

Theorem 2.4.8 (Universality of Globally Noncrossing Linkages). *For any compact semialgebraic set $R \subset \mathbb{R}^2$, there is a globally noncrossing linkage \mathcal{L} that draws R .*

Proof. By Lemma 2.3.3, R may be written as the projection of some basic compact set $R' = Z(f_1, \dots, f_s) \subset \mathbb{R}^{2m}$ onto the first two variables, so it suffices to show that some translation of R' may be drawn by a globally noncrossing linkage (by ignoring all drawing vertices except the first). By scaling as necessary, we may further assume that the compact set R' lies in the box $[-1, 1]^{2m}$. But this now follows directly from Construction 2.4.5, Part I. \square

2.5 Unit-Distance and $\{1, 2\}$ -Distance Graphs and Linkages

Definition 2.5.1 (Unit Distance Graphs/Linkages). Define an **abstract unit-distance graph (or linkage)** as an abstract weighted graph (or abstract linkage) where all edge-lengths equal 1; a **configured unit-distance graph (or linkage)** additionally comes with such a configuration.

With our terminology, an **abstract unit distance graph/linkage** does not necessarily have any valid configurations, in contrast to another common usage of the term “unit-distance graph” which implies a nonempty configuration space. To mitigate confusion with this overloaded term, we will always refer to a unit-distance graph/linkage as “abstract” or “configured”.

In this section we prove the strong $\exists\mathbb{R}$ -completeness or $\forall\mathbb{R}$ -completeness of realizability, rigidity, and global rigidity for unit-distance (or in the case of global rigidity, $\{1, 2\}$ -distance) linkages that allow crossings. We also show universality: any compact semialgebraic set in \mathbb{R}^2 can be drawn by a unit-distance linkage.

There are two noteworthy obstacles in these arguments that were not present in the previous section. First, the universality proof involves a new complication, namely, non-algebraic numbers. To illustrate, the circle $C = \{(x, y) \mid x^2 + y^2 = r^2\}$ (where r is any uncomputable number, such as Chaitin’s constant) can be drawn easily by a linkage (using an edge of length $|r|$), but simulating such an edge with a unit-distance graph is impossible. As a workaround, we instead rely on *pins* to introduce non-algebraic values. Indeed, we may slightly generalize curve C by introducing new variables (a, b) and considering the modified curve

$$C' = \{((x, y), (a, b)) \in \mathbb{R}^4 \mid x^2 + y^2 = a^2\}.$$

As C' is now defined by polynomials with *integer* coefficients, the Main Theorem (Part II) applies and may be simulated by a unit distance linkage as above. Finally, with one pin, we may fix the values $a = r$ and $b = 0$, which recovers the desired circle C . Suitably generalized, this argument can be made to work for arbitrary compact semialgebraic sets; see Theorem 2.5.9 for details.

For the second obstacle, we were not able to prove $\forall\mathbb{R}$ -completeness of detecting global rigidity of unit-distance graphs. Indeed, we are not aware of the existence of *any* globally rigid unit-distance graphs larger than a triangle!

Question. *Are there any simple, globally rigid unit-distance graphs with more than 3 edges?*

As a consolation prize, we demonstrate $\forall\mathbb{R}$ -completeness of global rigidity for graphs with edge-lengths in $\{1, 2\}$, an appropriate strengthening of Saxe's result [73] that global rigidity is coNP-hard for graphs with edge-lengths in $\{1, 2\}$.

2.5.1 Simulation with Unit- and $\{1, 2\}$ -Distance Linkages

Here we show how to simulate Theorem 2.2.13 using $\{1, 2\}$ -distance linkages.

Lemma 2.5.2. *A single edge of integer length n is perfectly simulated by a **reinforced bar** graph formed by adjoining $n - 1$ degenerate $\{1, 1, 2\}$ -sided triangles along unit edges.*

Proof. This is a simple extension of a tool used in [73, Cor. 4.3]. □

To rigidify orthogonal trees with $\{1, 2\}$ -graphs, it suffices to rigidify entire lattice grids:

Lemma 2.5.3. *Let (G, C) be the configured graph whose vertices lie at all integer points in $[0, n] \times [0, n]$ and whose unit-length edges connect vertically and horizontally adjacent vertices in this grid. Then the rigidified graph $\mathcal{G} := ((G, C), \text{RigidCon}_{\mathcal{G}}(G, C))$, where configuration C is rigidified in its entirety, can be perfectly simulated by an unconstrained $\{1, 2\}$ -distance graph, called a **reinforced grid**.*

Proof. We may assume $n \geq 5$. For each $0 \leq j, k \leq n$, let $v_{j,k}$ be the vertex of G with $C(v_{j,k}) = (j, k)$. By Lemma 2.5.2, we may add length-2 edges $(v_{j,k}, v_{j+2,k})$ and $(v_{j,k}, v_{j,k+2})$ to force each row and column of vertices in G to remain straight. Now add one more rigidified bar of length 5 connecting $v_{4,0}$ and $v_{0,3}$, which constrains row $k = 0$ and row $j = 0$ to remain at 90° from each other. In fact, this resulting graph G' is the desired graph. Indeed, suppose we have a configuration of G' ; by a Euclidean motion, we may assume $v_{0,0}$, $v_{n,0}$, and $v_{0,n}$ are configured at $(0, 0)$, $(n, 0)$, and $(0, n)$ respectively. Because $|v_{n,0} - v_{n,n}| = n$ and $|v_{0,n}, v_{n,n}| = n$ in any configuration, $v_{n,n}$ must rest at (n, n) or $(0, 0)$. In the latter case, $v_{n,1}$ rests at $(n - 1, 0)$, which is not distance n away from $v_{0,1}$ at $(0, 1)$, contradicting Lemma 2.5.2. So $v_{n,n}$ must indeed lie at (n, n) , and the rest of the vertices' locations are then fixed. □

Construction 2.5.4. *Use notation as in Theorem 2.2.13.*

If the hypotheses of Part II hold, we may construct, in $\text{poly}(\text{Size}(F))$ time, an abstract linkage $\mathcal{M} = \mathcal{M}(F)$ with edge-lengths in $\{1, 2\}$ that perfectly simulates the scaled linkage $c \cdot \mathcal{L}(F)$ for some constant integer $c = O(1)$.

If \mathcal{L} also satisfies Part III, the configuration of \mathcal{M} corresponding to $C_0 \in \text{Conf}(\mathcal{L})$ has rational coordinates whose numerators and denominators are bounded by $\text{poly}(\text{Size}(F))$.

Proof. Choose some $c = O(1)$ such that all edge-lengths of $c \cdot \mathcal{L}$ are integers, and replace \mathcal{L} with $c \cdot \mathcal{L}$ for the remainder of this proof. By Properties 2, 4 and 7, there is some bound $b = \text{poly}(\text{Size}(F))$ such that each rigidified subgraph (H, C_H) in \mathcal{L} has (after translation if necessary) integer coordinates bounded in magnitude by b .

For each such (H, C_H) , build a reinforced grid of unit squares as in Lemma 2.5.3 large enough to include the coordinates of the leaf vertices of (H, C_H) , and replace (H, C_H) by this grid. Each edge of \mathcal{L} not belonging to any rigidified tree gets replaced with a reinforced segment of appropriate length as in Lemma 2.5.2. Call the resulting linkage $\mathcal{M} = \mathcal{M}(F)$; by Lemmas 2.5.3 and 2.5.2, \mathcal{M} perfectly simulates \mathcal{L} .

If Part III holds, then c as above may be chosen such that, additionally, configuration C_0 (after scaling by c) has integer coordinates bounded by $\text{poly}(\text{Size}(F))$ in magnitude. It is now straightforward to check that the corresponding configuration of \mathcal{M} has rational coordinates whose magnitudes and denominators are bounded by $\text{poly}(\text{Size}(F))$, as claimed. \square

We rely on the full strength of *perfect* simulation in the proof of global rigidity below, but for the other hardness results, continuous and rigid simulation is sufficient. This may be achieved with only unit-length edges:

Lemma 2.5.5 ([74, Lemma 3.4]). *A single edge of length 2 can be continuously and rigidly simulated by a unit-distance graph with 19 edges, formed by joining two copies of Moser’s Spindle along a common equilateral triangle.*

2.5.2 Hardness and Universality of Unit-Distance and $\{1, 2\}$ -Distance Graphs

Theorem 2.5.6. *Deciding whether a given abstract unit-distance graph is realizable is $\exists\mathbb{R}$ -complete.*

Note: this was shown by Schaefer with a simpler, specialized construction, but we reprove it here for completeness.

Proof. This follows by reduction from **CommonZero** exactly as in the proof of Theorem 2.4.6, using Construction 2.5.4 in place of Construction 2.4.5. \square

Theorem 2.5.7. *The problem of determining whether a configured unit-distance graph with coordinates in $\mathbb{Q}[\sqrt{3}, \sqrt{5}]$ is rigid is strongly $\forall\mathbb{R}$ -complete.*

Proof. This follows by reduction from the complement of $\mathbf{H}_2\mathbf{N}$: given an instance $F = \{f_1, \dots, f_s\}$ of this problem (which we may assume consists of polynomials of degree 4 with constant-sized coefficients by Lemma 2.3.5), use Construction 2.5.4 and Lemma 2.5.5 to build a configured unit-distance linkage \mathcal{M} that continuously and rigidly draws a translation of some scaling $c \cdot (Z(F) \cap [-1, 1]^{2m})$. Then this linkage is rigid if and only if $\vec{0}$ is the only common zero of F , i.e., F is a “no” instance of $\mathbf{H}_2\mathbf{N}$. As in the proof of Theorem 2.4.7, removing the three pins of \mathcal{M} results in a unit-distance graph that is rigid if and only if F is a “no” instance of $\mathbf{H}_2\mathbf{N}$. \square

Theorem 2.5.8. *The problem of deciding whether a given configured $\{1, 2\}$ -distance graph with coordinates in $\mathbb{Q}[\sqrt{3}, \sqrt{5}]$ is globally rigid is strongly $\forall\mathbb{R}$ -complete.*

Proof. This follows by reduction from the complement of H_2N just as in the proof of Theorem 2.4.7, using Construction 2.5.4 instead of Construction 2.4.5. As in the proof of Theorem 2.4.7, this makes essential use of the fact that \mathcal{M} from Construction 2.5.4 perfectly simulates the linkage \mathcal{L} from the Main Theorem. \square

Proving or disproving the analogous statement for unit-distance graphs instead of $\{1, 2\}$ -distance graphs remains an interesting open problem. Even identifying a globally rigid unit-distance graph with more than 3 edges appears open! If one is found, it is likely that the methods of this paper can turn it into a proof of hardness.

As discussed in Section 2.5, Construction 2.5.4 requires the input polynomials to have integer coefficients, but some compact semialgebraic sets cannot be expressed in this way. We now formalize the workaround described there to prove universality of unit-distance linkages.

Theorem 2.5.9. *Any compact semialgebraic set $R \subset \mathbb{R}^2$ may be drawn by a unit-distance linkage.*

Proof. As in the proof of Theorem 2.4.8 we may write R as the projection onto coordinates x_1, y_1 of some compact basic algebraic set

$$R' = Z(f_1, \dots, f_s), f_i \in \mathbb{R}[x_1, y_1, \dots, x_m, y_m],$$

and it suffices to show that some translation of R' may be drawn with a unit-distance linkage. In fact, it suffices to show that the scaled set $\frac{1}{n} \cdot R'$ may be drawn (up to translation) by a unit-distance linkage \mathcal{L} , for some $n \in \mathbb{N}$: indeed, if a unit-distance linkage \mathcal{L} draws a translation of $\frac{1}{n} \cdot R'$, then $n \cdot \mathcal{L}$ draws a translation of R' and has integer edge-lengths, so by Lemmas 2.5.2 and 2.5.5 it may be simulated by a unit-distance linkage, as required. By this reasoning, we may replace the compact set R' by some small-enough $\frac{1}{n} \cdot R'$ and thereby assume that R' lies in the box $[-1, 1]^{2m}$.

Let c be the constant integer implied in Theorem 2.5.4, and write the scaled set $\frac{1}{c} \cdot R'$ as $Z(g_1, \dots, g_s)$, where $g_j(\vec{x}\vec{y}) = f_j(c \cdot \vec{x}y)$. We wish to apply Theorem 2.5.4 to polynomials g_1, \dots, g_s so the resulting linkage draws precisely R' (up to translation), but these coefficients may not be integers (or even algebraic numbers, as described above), so we will temporarily replace these coefficients with variables, as follows. By scaling, we may assume polynomials g_i have coefficients in $[-1, 1]$. For each nonzero monomial $c_{i,J} \vec{x}^J$ in each g_i , create new variables $a_{i,J}$ and $b_{i,J}$, gather all of these new variables into a vector \vec{ab} with length $2r$, and define the new polynomials

$$h_i(\vec{x}\vec{y}, \vec{ab}) := \sum_{i,J \text{ such that } c_{i,J} \neq 0} a_{i,J} \vec{x}^J,$$

for $1 \leq i \leq s$. Polynomials h_i now have *integer* coefficients, so by Construction 2.5.4 and Lemma 2.5.5 we may construct some unit-distance linkage \mathcal{L} , a set of drawing vertices

$$X = \{v_i \mid 1 \leq i \leq s\} \cup \{v_{i,J} \mid 1 \leq i \leq s \text{ and } c_{i,J} \neq 0\}$$

corresponding to variables (x_i, y_i) and $(a_{i,J}, b_{i,J})$ respectively, and some translation T on \mathbb{R}^{2m+2r} such that X draws a set between $T(c \cdot Z(g_1, \dots, g_s) \cap [-1, 1]^{2m+2r})$ and $T(c \cdot Z(h_1, \dots, h_s))$. Finally, we pin all the vertices $v_{i,J}$ in the plane to force variables $(a_{i,J}, b_{i,J})$

to take the values $(c_{i,J}, 0)$: specifically, if $T_{i,J}$ denotes the translation T restricted to coordinates $(a_{i,J}, b_{i,J})$, we pin $v_{i,J}$ to the point $T_{i,J}(c_{i,J}, 0)$. It may be checked that the trace of vertices v_1, \dots, v_s in this pinned linkage is a translation of $c \cdot Z(g_1, \dots, g_s) = R'$, as required. \square

2.6 Matchstick Graphs and Linkages

Definition 2.6.1 (NX-Constrained Linkages and Matchstick Linkages). If \mathcal{L} is a linkage, we define the **NX-constraint** on \mathcal{L} , denoted $\text{NXCon}_{\mathcal{L}}$, as the set of configurations of \mathcal{L} that do not cross. In other words, if $\mathcal{L}' := (\mathcal{L}, \text{NXCon}_{\mathcal{L}})$ is an **NX-constrained linkage**, then its configuration space is defined by $\text{Conf}(\mathcal{L}') := \text{NConf}(\mathcal{L})$.

If \mathcal{L} is an unconstrained linkage all of whose edges have length 1, then the NX-constrained linkage $\mathcal{L}' := (\mathcal{L}, \text{NXCon}_{\mathcal{L}})$ is called a **matchstick linkage**.

In this section we prove results about matchstick linkages and graphs analogous to those of the prior sections: matchstick graph realization is $\exists\mathbb{R}$ -complete, matchstick graph rigidity and global rigidity are $\forall\mathbb{R}$ -complete, and any compact semialgebraic set can be drawn by a matchstick linkage. However, it is unfair to refer to this last result as “universality”, because matchstick linkages (more generally, NX-constrained linkages) can draw more than just compact semialgebraic sets! As a simple example, the NX-constrained two-bar linkage \mathcal{A} of Figure 2-4 draws the half-open annulus $\{(x, y) \in \mathbb{R}^2 \mid 1 < x^2 + y^2 \leq 9\}$, because configurations of (the unconstrained linkage underlying) \mathcal{A} with v on circle $x^2 + y^2 = 1$ are crossing and are thus considered invalid by the constraint. In general, the configuration spaces of NX-constrained linkages are *bounded and semialgebraic* sets. And NX-constrained linkages—in fact, matchstick linkages—can indeed draw all such sets, but our proof of this stronger result subtly breaks the abstraction barrier set up by the Main Theorem, so we postpone this proof until the end of Section 2.2.

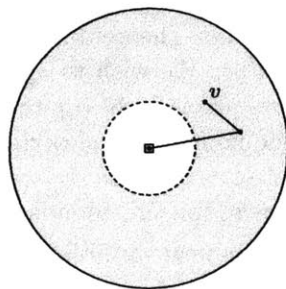


Figure 2-4: The trace of an NX-constrained linkage need not be closed.

2.6.1 Simulation with Matchstick Linkages

For an integer $a \geq 5$, we will simulate length- a edges by **edge polyiamonds** as shown in Figure 2-5. Note that any polyiamond, considered as a matchstick graph, is globally rigid. As proven in the following lemma, two simple but effective **wing edges** suffice to fix the relative orientation and position of two adjacent edge polyiamonds (Figure 2-5). We may more easily describe the *relative* orientations of edge polyiamonds P and Q by temporarily pinning P in place:

Lemma 2.6.2 (Wing Edges). *Consider the matchstick linkage \mathcal{L} drawn in Figure 2-5, with two edge polyiamonds P and Q sharing a vertex, three pins in P , and two extra **wing edges** attached at vertices a and b as shown. Then in every configuration of this linkage, Q has the same orientation as P , and its central axis is rotated from P 's axis by an angle $60^\circ < \theta < 240^\circ$. Every such θ corresponds to a unique configuration of \mathcal{L} .*

Proof. Polyiamond Q must be drawn with the same orientation as P , or else the wing edges will create crossings. Likewise, a , o , b , and c must form a rhombus to prevent these four edges from intersecting each other. So Q 's and c 's location are determined by the single angle θ , and it may be seen that no crossings occur precisely when $60^\circ < \theta < 240^\circ$. \square

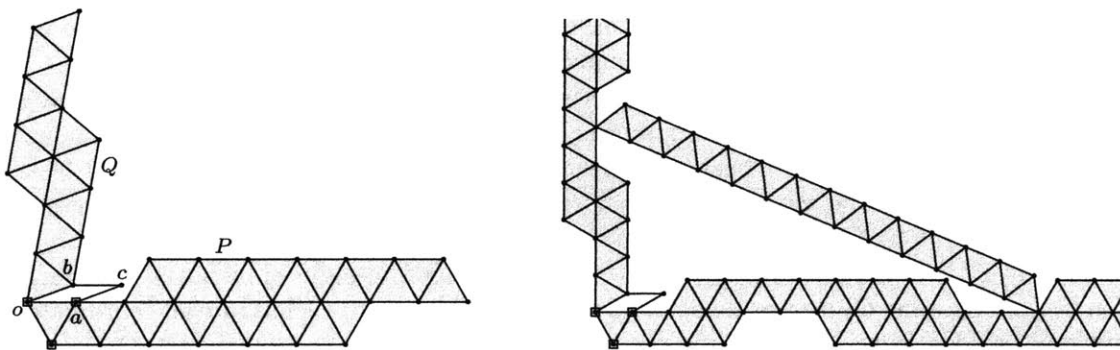


Figure 2-5: Left: Edge polyiamonds used to simulate edges of integer length. Right: Edge polyiamonds braced at 90° .

Lemma 2.6.3 (Orthogonal Braces). *By modifying the edge polyiamonds and adding a **hypotenuse polyiamond** as shown in Figure 2-5, the resulting matchstick linkage \mathcal{M} is globally rigid, and its unique configuration has $\theta = 90^\circ$.*

Proof. The previous lemma and the 5-12-13 right triangle force Q to be drawn at a $+90^\circ$ angle from P and with the same orientation. The hypotenuse polyiamond then has only one crossing-free position. \square

Construction 2.6.4. *Use notation as in Theorem 2.2.13.*

If the hypotheses of Part II hold, we may construct, in $\text{poly}(\text{Size}(F))$ time, an abstract matchstick linkage \mathcal{M} that perfectly simulates the scaled linkage $n \cdot \mathcal{L}$ for a constant positive integer n .

If \mathcal{L} also satisfies Part III, the configuration of \mathcal{M} corresponding to $n \cdot C_0 \in n \cdot \text{Conf}(\mathcal{L})$ has coordinates of the form $(a + b\sqrt{3})/c$ for integers a, b, c with magnitude at most $\text{poly}()$.

Proof. We may assume that in each rigidified subtree (H, C_H) of \mathcal{L} , no internal node has degree exactly 2 with edges at 180° from each other in C_H , because such nodes can be erased by merging these two edges together. There is a constant scale factor $n \geq 40$ so that in $n \cdot \mathcal{L}$, each rigidified subtree is guaranteed by Theorem 2.2.13 to have integer coordinates, and $n \cdot \mathcal{L}$ is guaranteed to have feature size at least 40; for the remainder of this proof, replace \mathcal{L} with $n \cdot \mathcal{L}$. Every edge of \mathcal{L} has length at least 40, and every configuration of \mathcal{L} has feature size at least 40.

We create a matchstick linkage \mathcal{M} simulating \mathcal{L} using Lemmas 2.6.2 and 2.6.3, as follows. First, \mathcal{M} has a vertex corresponding to each vertex of \mathcal{L} , with the same three pins. Each

edge of \mathcal{L} is replaced by an edge polyiamond of the appropriate length connecting the corresponding vertices. At any vertex v internal to some (necessarily unique) rigidified tree (H, C_H) , brace any right angles at v in (H, C_H) as in Lemma 2.6.3, by modifying the corresponding edge polyiamonds and adding wings and a hypotenuse. Any 180° or 270° angles at v in (H, C_H) can be left alone—there are no straight degree-2 vertices in (H, C_H) , so bracing only the right angles suffices to rigidify v 's neighborhood in (H, C_H) . Thus, all edge polyiamonds corresponding to edges in H are braced to perfectly simulate the rigid constraint $\text{RigidCon}_{\mathcal{L}}(H, C_H)$.

At any vertex $v \in V(G)$ that is *not* internal to any rigidified tree, simply attach wings (as in Lemma 2.6.2) according to the cyclic order σ_v at v . Because all configurations $C \in \text{Conf}(\mathcal{L})$ agree with σ and have all angles at v strictly between 60° and 240° , Lemma 2.6.2 guarantees that C 's placement of the edges adjacent to v corresponds to exactly one configuration of the edge polyiamonds and wings surrounding v in \mathcal{M} .

To show that this matchstick linkage \mathcal{M} computed from \mathcal{L} indeed perfectly simulates \mathcal{L} , it remains to show that no configuration of \mathcal{L} creates crossings in the corresponding placement of \mathcal{M} . This follows because every configuration C of \mathcal{L} has feature size at least 40, while the linkage \mathcal{M} extends less than 20 units away from the vertices and edges it simulates.

Finally, if Part III applies, let D_0 be the configuration of \mathcal{M} induced by initial configuration C_0 (after scaling C_0 by c as above), and note that (the scaled) C_0 has integer coordinates and integer edge lengths, both of polynomial magnitude. We must show that D_0 's coordinates have the required form $(a + b\sqrt{3})/c$ for polynomially-bounded integers a, b, c . For any edge of C_0 with integer endpoints (a_1, b_1) and (a_2, b_2) and integer length r , the vertices along the central axis of the corresponding edge polyiamond have rational coordinates of the form $(1 - \frac{h}{r})(a_1, b_1) + \frac{h}{r}(a_2, b_2)$ for integers $0 \leq h \leq r$, and the rest of the vertices in this polyiamond are offset from these coordinates by the vector $(1 - \frac{1}{r})(a_1, b_1) + \frac{1}{r}(a_2, b_2)$ rotated by some multiple of 60° ; all of these coordinates have the required form. The same computation holds for hypotenuse polyiamonds. The only vertices not yet accounted for are the wing vertices, and these have the form $A + B - O$ where A, B , and O are rational-coordinate points along edge-polyiamond axes as described above. \square

2.6.2 Hardness and (Weak) Universality of Matchstick Linkages

With the tools above in place, the proofs of the following four theorems are perfectly analogous to those in Section 2.5, using Construction 2.6.4 instead of Construction 2.5.4 and Lemma 2.5.5, so we omit their proofs.

Theorem 2.6.5 (Realizability of Matchstick Graphs). *Deciding whether a given abstract matchstick graph is realizable is $\forall\mathbb{R}$ -complete.*

Theorem 2.6.6 (Rigidity of Matchstick Graphs). *Deciding whether a given configured matchstick graph (whose coordinates are in $Q[\sqrt{3}]$) is rigid is strongly $\forall\mathbb{R}$ -complete.*

Theorem 2.6.7 (Global Rigidity of Matchstick Graphs). *Deciding whether a given configured matchstick graph (whose coordinates are in $Q[\sqrt{3}]$) is globally rigid is strongly $\forall\mathbb{R}$ -complete.*

Theorem 2.6.8 (Partial Universality of Matchstick Linkages). *For any compact semialgebraic set $R \subset \mathbb{R}^2$, there is a matchstick linkage that draws exactly R .*

Recall that Theorem 2.6.8 is incomplete. We prove full universality—that matchstick linkages can draw all bounded semialgebraic sets—in Section 2.7.7.

2.7 Extended Linkages and the Main Construction

In this section, we prove the Main Theorem in full detail, following the plan outlined in Section 2.7.2.

2.7.1 Defining Extended Linkages

For convenience and clarity, we define and use **extended linkages**, which are constrained linkages whose constraints are tailored for the specifics of our construction. The specific constraints used in the definition of extended linkages below are highly specialized to the constructions in this paper; more general constraints are possible but not necessary here. The first of these constraints, the cyclic constraint, specifies a preferred arrangement of edges around each vertex.

Definition 2.7.1 (Cyclic Constraint). For an abstract linkage \mathcal{L} with combinatorial embedding* σ , a configuration C of \mathcal{L} satisfies the **cyclic constraint** $\text{CyclicCon}_{\mathcal{L}}(\sigma)$ precisely when, for each vertex v (having $\sigma_v = [e_1, \dots, e_{\deg(v)}]$), segments $C(e_1), \dots, C(e_{\deg(v)})$ intersect only at $C(v)$ and are arranged counterclockwise around $C(v)$ in this order.

Note that this is a purely local condition around each vertex and does not force C to **agree with** σ as in Definition 2.2.5, because C may have crossings while still satisfying $\text{CyclicCon}_{\mathcal{L}}(\sigma)$.

Definition 2.7.2 (Sliceform Constraint). For a constrained abstract linkage \mathcal{L} possessing a cyclic constraint $\text{CyclicCon}_{\mathcal{L}}(\sigma)$, a **sliceform constraint**, $\text{SliceCon}_{\mathcal{L}}(S)$, is specified by a subset $S \subset V(G)$ of (some or all of the) vertices of degree 4. A configuration $C \in \text{Conf}(\mathcal{L})$ (necessarily satisfying $\text{CyclicCon}_{\mathcal{L}}(\sigma)$) satisfies the sliceform constraint $\text{SliceCon}_{\mathcal{L}}(S)$ if, for each **sliceform vertex** $v \in S$, segments $C(e_1)$ and $C(e_3)$ are collinear and $C(e_2)$ and $C(e_4)$ are collinear, where $\sigma_v = [e_1, e_2, e_3, e_4]$.

Sliceforms allow a limited form of “nonplanar” interaction while still being simulatable without crossings (cf. Figure 2-16), so they are our primary tool in circumventing the difficulties of planarity.

An angle constraint specifies that angles should not be allowed to change very much.

Definition 2.7.3 (Angle Constraint). If linkage \mathcal{L} has a cyclic constraint $\text{CyclicCon}_{\mathcal{L}}(\sigma)$, an **angle constraint**, $\text{AngleCon}_{\mathcal{L}}(A, \Delta)$, is specified by an assignment of an angle $0 \leq A(\Lambda) \leq 2\pi$ and an angle tolerance $\Delta(\Lambda) \geq 0$ to each angle chain Λ of \mathcal{L} , with the condition that A assigns a total of 2π to the angle chains around each vertex.

A configuration $C \in \text{Conf}(\mathcal{L})$ (necessarily satisfying $\text{CyclicCon}_{\mathcal{L}}(\sigma)$) satisfies the angle constraint $\text{AngleCon}_{\mathcal{L}}(A, \Delta)$ if, for each angle chain Λ , angle $\angle C(\Lambda)$ lies in the closed interval

$$[A(\Lambda) - \Delta(\Lambda), A(\Lambda) + \Delta(\Lambda)].$$

In particular, any angle chain with $\Delta(\Lambda) = 0$ is rigid: its angle in C must be exactly $A(\Lambda)$.

*Recall that a **combinatorial embedding** consists of a preferred cyclic counterclockwise ordering σ_v of the edges incident to each vertex v (Definition 2.2.5).

An extended linkage is simply a linkage with each type of constraint listed above, with a few convenient restrictions:

Definition 2.7.4 (Extended Linkage). An (ε, δ) -extended linkage where $0 < \delta < \varepsilon < \pi/4$ is defined as a constrained linkage \mathcal{L} whose constraints K have the form

$$K = \{\text{CyclicCon}_{\mathcal{L}}(\sigma), \text{SliceCon}_{\mathcal{L}}(S), \text{AngleCon}_{\mathcal{L}}(A, \Delta)\},$$

where at each angle chain Λ of \mathcal{L} , $A(\Lambda) \in \{90^\circ, 180^\circ, 270^\circ, 360^\circ\}$ and $\Delta(\Lambda) \in \{0, \delta, \varepsilon\}$. We will call \mathcal{L} simply an **extended linkage** when ε and δ are clear from context.

2.7.2 Detailed Overview of Strategy

Our Main Construction is primarily concerned with showing, for a finite set F of polynomials in $\mathbb{R}[x_1, y_1, \dots, x_m, y_m]$, how to construct an extended linkage that draws a bounded portion of the common zero set $Z(F)$, i.e., something between $Z(F) \cap [-1, 1]^{2m}$ and $Z(F)$, up to a translation. This subsection outlines the key points of our approach, with full details to follow in subsequent subsections.

Our construction uses a transformation to polar coordinates similar to the one used in Kempe's original (flawed) construction and the corrected construction of Abbott et al. [2]. To that end, we define

$$\text{Rect}(\alpha, \beta) := (\cos \alpha, \sin \alpha) + (-\sin \beta, \cos \beta) - (1, 1) = e^{i\alpha} + i \cdot e^{i\beta} - (1 + i); \quad (2.1)$$

note that we use coordinate or complex number representation interchangeably when convenient. Note also that $\text{Rect}(0, 0) = (0, 0)$.

Lemma 2.7.5. *For any angle $0 < \theta < \pi/4$, the function Rect is a homeomorphism from the region $[-\theta, \theta]^2$ in the (α, β) -plane onto its image in the (x, y) -plane, which is a compact region (specifically, the Minkowski sum of two circular arcs) containing the box $[-\theta/2, \theta/2]^2$. This region is star-shaped around $(0, 0)$.*

Proof. This image region can be formed from a square of side-length $2 \sin \theta$ by transplanting two segments of unit circles as illustrated in Figure 2-6. The region contains $[-\theta/2, \theta/2]^2$ because $\frac{\theta}{2} + (1 - \cos \theta) \leq \sin \theta$, which may be verified. The rest of the claims follow straightforwardly. \square

In place of rectangular coordinates (x_j, y_j) , we will use angles (α_j, β_j) related by $(x_j, y_j) = 2r \cdot \text{Rect}(\alpha_j, \beta_j)$, where the radius $2r$ will be carefully chosen later. We may write this equivalently as

$$\begin{aligned} x_j &= r \left(e^{i\alpha_j} + e^{-i\alpha_j} + ie^{i\beta_j} - ie^{-i\beta_j} - 2 \right), \\ y_j &= r \left(-ie^{i\alpha_j} + ie^{-i\alpha_j} + e^{i\beta_j} + e^{-i\beta_j} - 2 \right). \end{aligned} \quad (2.2)$$

By making this latter substitution into each polynomial $f \in F$, we eventually arrive at a representation of the form

$$f(\vec{x}, \vec{y}(\vec{\alpha}, \vec{\beta})) = f(0) + \sum_{u=0}^3 \sum_{I \in \text{Coeffs}(2m, d)} i^u \cdot d_{u, I} \cdot \left(e^{i \cdot (I \cdot \vec{\alpha}, \vec{\beta})} - 1 \right), \quad (2.3)$$

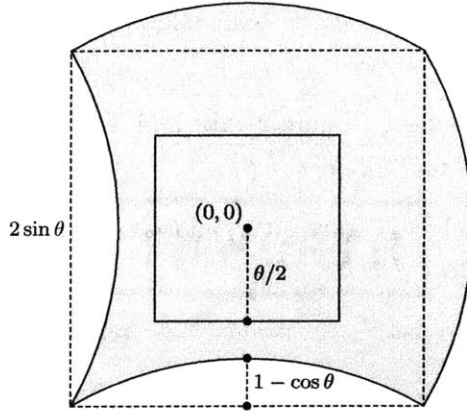


Figure 2-6: The region described in Lemma 2.7.5 (gray) is bounded by four circular arcs and contains the square $[-\theta/2, \theta/2]^2$.

where $\vec{\alpha\beta} := (\alpha_1, \beta_1, \dots, \alpha_m, \beta_m)$, each $d_{u,I}$ is a nonnegative real number, and

$$\text{Coeffs}(2m, d) := \{(a_1, \dots, a_{2m}) \in \mathbb{Z}^{2m} \mid |a_1| + \dots + |a_{2m}| \leq d\}.$$

Only $d_{u,I} - d_{u+2,I}$ (with indices taken modulo 4) affects the total, so by further cancellation we may assume one or both of these coefficients is 0 (for each u, I pair). Even though $f(\vec{x}\vec{y}(\vec{\alpha\beta}))$ is real, this complex representation proves more useful for computations below.

Lemma 2.7.6. *If f has total degree at most d and integer coefficients with magnitude at most M , and r is a positive integer, then in representation (2.3) there are at most $2(2m)^d(2d+1)^d$ nonzero coefficients $d_{u,I}$, each coefficient $d_{u,I}$ is an integer, and the absolute values of all of these coefficients add to at most $6^d \cdot r^d \cdot M \cdot \binom{2m+d}{d}$. All of these coefficients may be computed from f in deterministic time $\text{poly}(m^d, d^d, r^d, M)$.*

Proof. We may bound the size of $\text{Coeffs}(2m, d)$ by overcounting as follows: construct an integer vector $I \in \mathbb{Z}^{2m}$ by choosing d entries (or choosing all of them if $d > 2m$), assigning each chosen entry a value from $\{-d, -d+1, \dots, d\}$, and assigning 0 for the remaining entries. The number of such I is $\max(\binom{2m}{d}, 1) \cdot (2d+1)^d \leq (2m)^d(2d+1)^d$. There are therefore at most $2(2m)^d(2d+1)^d$ nonzero coefficients $d_{u,I}$.

For each monomial $c \cdot \vec{x}\vec{y}^J$ in f , substitute each x_j and y_j as in Equation 2.2 and expand fully *without collecting like terms*. There are at most 6^d terms in this expansion (interpreting 2 as $1+1$), each of the form $c \cdot r^d \cdot i^u \cdot e^{I \cdot \vec{\alpha\beta}}$, so the sum of the magnitudes of these coefficients is at most $6^d \cdot r^d \cdot c \leq 6^d \cdot r^d \cdot M$. Adding this quantity across all monomials in f , we obtain a sum no larger than $6^d \cdot r^d \cdot M \cdot \binom{2m+d}{d}$. Finally, collecting like terms (including cancelling $d_{u,I}$ and $d_{u+2,I}$ pairs as much as possible) can only decrease the sum of the magnitudes of the coefficients by the triangle inequality. This proves the claim. \square

We will use this polar representation as a template to compute each polynomial f in the linkage. Indeed, much like in the strategies referenced above, we provide gadgets for the following tasks.

- The Start Gadget (Figure 2-12a) converts from rectangular position (x_j, y_j) to polar angles (α_j, β_j) .

- The Angle Average Gadget (Figure 2-9c) allows adding and subtracting angles to construct all of the $I \cdot \overrightarrow{\alpha\beta}$ values.
- The Vector Creation Gadget (Figure 2-12b) and Vector Rotation Gadget (Figure 2-12d) compute the vectors $i^u \cdot d_{u,I} \cdot e^{i(I \cdot \overrightarrow{\alpha\beta})}$.
- The Vector Average Gadget (Figure 2-12c) allows adding vectors to compute the values $f(\overrightarrow{xy}(\overrightarrow{\alpha\beta})) - f(0)$ for each $f \in F$.
- The End Gadget (Figure 2-12e) constrains these values to equal $-f(0)$.

We employ several new ideas to ensure the resulting extended linkage $\mathcal{E}(F)$ is noncrossing. First, we construct a rigid grid of large square cells. Each gadget is isolated in one or $O(1)$ of these cells, and information is passed between gadgets/cells only using sliceform vertices along grid edges. In this way, these modular gadgets may be analyzed individually, as there is no possibility for distinct gadgets to intersect each other. We therefore rely on the Copy Gadget (Figure 2-9a) to copy angles and propagate them along paths of cells to distant gadgets in the grid. The Crossover Gadget (Figure 2-9b) allows these paths to cross, so we are not restricted to planar communication between gadgets. These gadgets make frequent use of the Parallel Gadget in Figure 2-8, which (with pins removed) keeps segments parallel without otherwise restricting motion. Figure 2-15 shows an example of the gadgets working together.

This linkage $\mathcal{E}(F)$ is an (ε, δ) -extended linkage, where ε and δ (the angle tolerances in $\text{AngleCon}_{\mathcal{L}}(A, \Delta)$) should be interpreted in the following manner. The parameter ε constrains bar movement enough to protect against crossings and to ensure uniqueness. By contrast, δ serves (morally) as a lower bound: in each gadget we construct, we ensure that every angle chain with tolerance δ can in fact realize any offset in the entire interval $[-\delta, \delta]$ —this is how we ensure we can draw a large enough portion of $Z(F)$.

Finally, we simulate linkage $\mathcal{E}(F)$ with a partially rigidified linkage $\mathcal{L}(F)$, in two steps. First, by replacing a vicinity of each sliceform vertex in $\mathcal{E}(F)$ with the Sliceform Gadget (Figure 2-16), we construct an extended linkage $\mathcal{E}'(F)$ that perfectly simulates $\mathcal{E}(F)$ but has no sliceforms. Then, we replace each edge of $\mathcal{E}'(F)$ with a rigidified orthogonal tree, connected to neighboring edges with the Angle Restrictor Gadget (Figure 2-17), which exactly enforces the cyclic constraint and the angle constraints.

2.7.3 The Gadgets

It will be useful to describe the gadgets below not by position of vertices as with π_X , but by the angles of a chosen set of angle chains. We therefore define a function Diff that measures how these angles differ from their “neutral” values given by A in the angle constraint $\text{AngleCon}(A, \Delta)$:

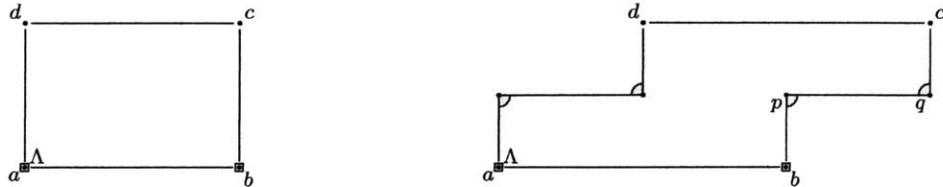
Definition 2.7.7 (The Diff Function). If \mathcal{L} is an extended linkage and $Y = (\Lambda_1, \dots, \Lambda_k)$ is a tuple of (some or all of) the angle chains of \mathcal{L} , define the function $\text{Diff}_Y : \text{Conf}(\mathcal{L}) \rightarrow \mathbb{R}^{|X|}$ by

$$\text{Diff}_Y(C) = (\angle C(\Lambda_1) - A(\Lambda_1), \dots, \angle C(\Lambda_k) - A(\Lambda_k)).$$

In drawings of extended linkages, we use a few conventions. All angle chains Λ are drawn at their “resting” angle, $A(\Lambda)$. Angle chains marked with a solid gray sector have $\Delta(\Lambda) = 0$,

and the rest have $\Delta(\Lambda) = \varepsilon$ unless otherwise specified. Vertices surrounded by squares are pinned, and those marked with an “x” are sliceform vertices.

We rely heavily on parallelograms to force edges to remain parallel. This directly relates to the pivotal flaw in Kempe’s original argument, which did not account for the possibility of parallelograms becoming contraparallelograms, and vice versa. The parallelogram case was resolved by Abbott et al. [2] using an additional *brace* along the midline. We offer another resolution via angle constraints: a parallelogram whose angles are constrained narrowly enough cannot “flip” into an contraparallelogram.*



(a) Parallelogram linkage \mathcal{P}_1 , initially with opposite corners at $a = (0, 0)$ and $c = (x, y)$ where $x, y \geq 1$.

(b) Parallelogram linkage \mathcal{P}_2 , with corners initially at $a = (0, 0)$, $b = (4, 0)$, $c = (6, 2)$, and $d = (2, 2)$.

Figure 2-7: Creating parallelograms with extended linkages.

Lemma 2.7.8 (Parallelogram Linkage). *For a small enough constant ε , extended linkages \mathcal{P}_1 and \mathcal{P}_2 of Figures 2-7a and 2-7b are globally noncrossing with global minimum feature size $\geq 1/2$, and every configuration of either linkage has $\vec{dc} = \vec{ab}$. Furthermore, the configuration spaces $\text{Conf}(\mathcal{P}_1)$ and $\text{Conf}(\mathcal{P}_2)$ are perfectly described by the map Diff_Λ , which provides homeomorphisms*

$$\text{Conf}(\mathcal{P}_1) \simeq [-\varepsilon, \varepsilon] \quad \text{and} \quad \text{Conf}(\mathcal{P}_2) \simeq [-\varepsilon, \varepsilon].$$

Proof. Begin with linkage \mathcal{P}_1 in Figure 2-7a, where all angle chains have tolerance ε and which has initial configuration $a = (0, 0)$ and $c = (x, y)$. In any configuration C , $d = y(\cos(\pi/2 + \theta), \sin(\pi/2 + \theta))$ where $\theta = \text{Diff}_\Lambda(C) \in [-\varepsilon, \varepsilon]$. Vertex c has two potential positions: $b + d - a$ (forming a parallelogram), or the reflection of this point across diagonal bd (forming an contraparallelogram). The former case indeed satisfies all of its constraints and is noncrossing with feature size at least $\min(x, y)/2$. In the latter case, angle chain dcb would have angle $3\pi/2 - \theta$, which is well outside the allowable range of $[\pi/2 - \varepsilon, \pi/2 + \varepsilon]$. So only the parallelogram configuration exists.

In linkage \mathcal{P}_2 , the vertices have initial positions $a = (0, 0)$, $b = (4, 0)$, $c = (6, 2)$, and $d = (2, 2)$. The same argument applies directly to \mathcal{P}_2 , showing that each configuration C is uniquely determined by $\theta = \text{Diff}_\Lambda(C)$, and that \mathcal{P}_2 is globally noncrossing with global minimum feature size $\geq 1/2$. \square

Parallelograms are especially useful in pairs, as with the linkage in Figure 2-8, which forces bars ab and ef to remain parallel while letting them move freely relative to each other (within some neighborhood). The classical counterpart of this linkage suffers from nonuniqueness: When positions for bars ab and ef are chosen (necessarily parallel), usually

*Angle constraints *per se* only apply to *constrained* linkages. It is more correct to say that we resolve parallelogram flipping via angle constraints *combined* with a method of *simulating* these constraints using classical linkages.

there are two possible locations for bar cd . Angle constraints again improve the situation: one of these two positions violates the angle constraints and is therefore invalid.

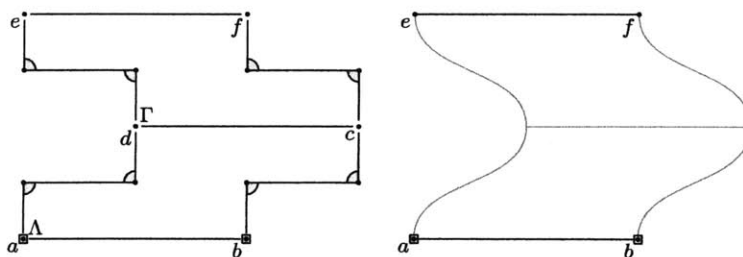


Figure 2-8: Left: The Parallel Gadget allows e to move freely in a neighborhood of its initial position while forcing ef to remain parallel to ab . Right: a schematic representation of the same gadget.

Lemma 2.7.9 (Parallel Gadget). *For extended linkage $\mathcal{L}_{\text{parallel}}$ drawn in Figure 2-8, vertex e is free to move in some region U containing a neighborhood of $(0, 4)$, but the entire configuration is uniquely determined by e 's position, and vector ef is always parallel to ab . Specifically, the map $\pi_{(e,f)}$ provides a homeomorphism*

$$\text{Conf}(\mathcal{L}_{\text{parallel}}) \simeq \{(p, q) \mid p \in U \text{ and } q - p = (x, 0)\}.$$

Proof. If C is a configuration of $\mathcal{L}_{\text{parallel}}$, let $(\alpha, \beta) = \text{Diff}_{\Lambda, \Gamma}(C)$. Lemma 2.7.8 shows that α and β together are enough to uniquely determine C , so long as this configuration satisfies its angle tolerances. By this Lemma, all tolerances are automatically satisfied except possibly the angle chains to the left of d and to the right of c , which have offsets of $\alpha - \beta$ and $\beta - \alpha$ respectively. So C is a valid configuration precisely when $\alpha - \beta \in [-\varepsilon, \varepsilon]$. By the same Lemma, this configuration is indeed noncrossing with minimum feature size at least $1/2$.

The region U of valid positions for e is the region

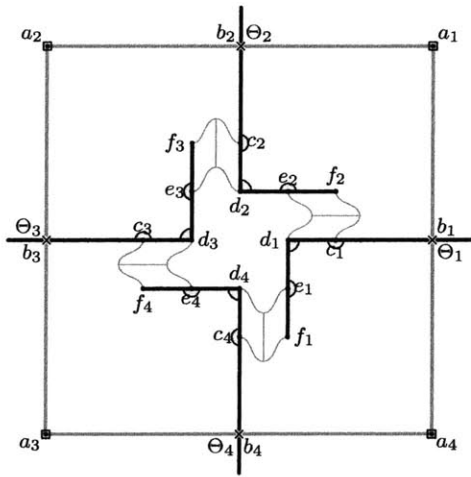
$$S = \{\text{Rect}(\alpha, \beta) \mid \alpha, \beta \in [-\varepsilon, \varepsilon] \text{ and } \alpha - \beta \in [-\varepsilon, \varepsilon]\}$$

which has been dilated by $2\sqrt{2}$, rotated by 45° , and finally translated by $(0, 4)$. Because S contains a neighborhood of 0 (indeed, it contains $\text{Rect}([-\varepsilon/2, \varepsilon/2]^2)$), U indeed contains a neighborhood of $(0, 4)$. \square

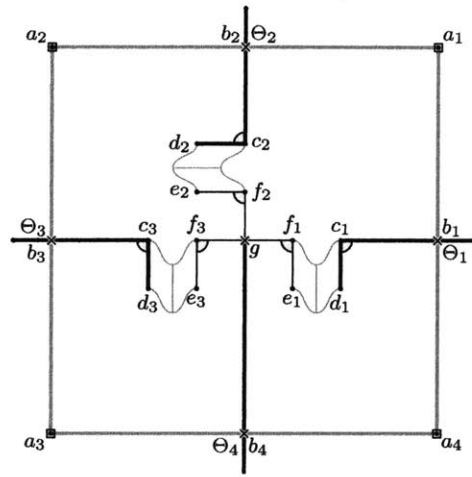
We now present the modular cell gadgets themselves. Each will inhabit one or more $Q \times Q$ square cells, where each cell has pairs of **transmission edges** attached at **transmission vertices**, which are sliceform vertices at the midpoints of cell edges. The angle chains around each transmission vertex will have tolerance 0 or δ —never ε —as required by the gadget. An angle chain (u, v, w) where (u, v) is a transmission edge is the corresponding **transmission angle chain**; the angles at these angle chains are the gadgets' only means of communicating with each other.

We begin with the gadgets that manipulate angles directly.

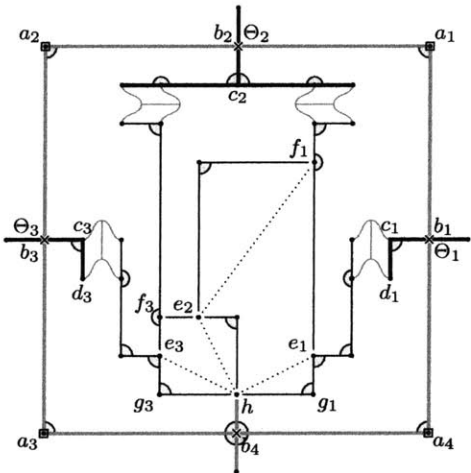
Lemma 2.7.10 (Gadgets for Angle Manipulation). *For any small enough constant $\varepsilon > 0$ and small enough $\delta < \varepsilon$, whenever $Q \geq 40$, the abstract (ε, δ) -extended linkages below are globally noncrossing with global minimum feature size at least $\frac{1}{2}$, and their configuration spaces are exactly parameterized as described below. Each gadget is described by an initial configuration where all coordinates are integer multiples of $Q/40$.*



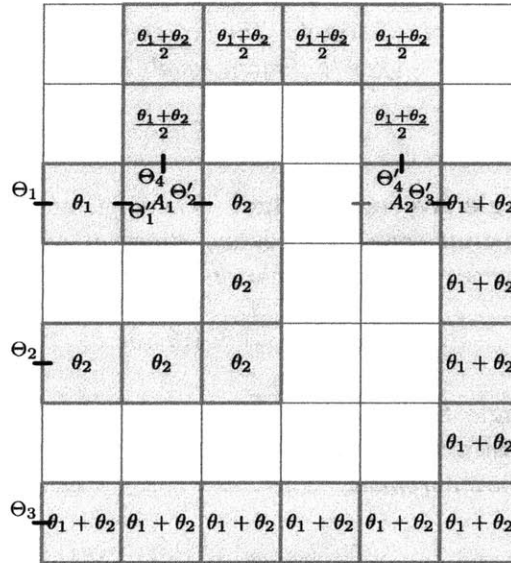
(a) The Copy Gadget forces $\theta_1 = \theta_2 = \theta_3 = \theta_4$. See this gadget's movement in Figure 2-10.



(b) The Crossover Gadget forces $\theta_1 = \theta_3$ and $\theta_2 = \theta_4$.



(c) The Angle Average Gadget forces $\theta_2 = (\theta_1 + \theta_3)/2$. See this gadget's movement in Figure 2-11.



(d) Angle Sum Gadget

Figure 2-9: Gadgets for manipulating angles in the Main Construction. In these figures we use the convention that $\text{Diff}_\Theta(C) = \theta$, $\text{Diff}_\Lambda(C) = \alpha$, and $\text{Diff}_\Gamma(C) = \beta$, possibly with subscripts. Angle chains marked with a solid gray sector are frozen, i.e., have tolerance 0. Unfrozen angle chains at cell edge midpoints are assigned tolerance δ . All remaining angle chains have tolerance ε . Vertices surrounded by squares are pinned, and those marked with an “x” are sliceform vertices. The pins shown here at the vertices a_i are for clarification only; in the overall construction, these nodes are forbidden from moving by other means, so these explicit pins are unnecessary.

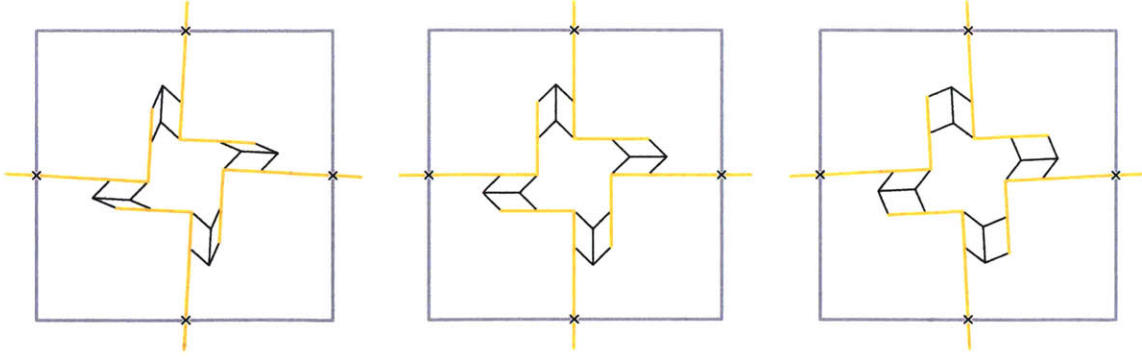


Figure 2-10: Snapshots of the configuration space of the Copy Gadget (cf. Figure 2-9a, Lemma 2.7.10), shown for angle offsets of 3° , 0° and -3° respectively.

Copy Gadget. The copy gadget $\mathcal{L}_{\text{copy}}$ drawn in Figures 2-9a and 2-10 constrains its four sliceform angles to remain equal. Specifically, $\text{Diff}_{(\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4)}$ provides a homeomorphism

$$\text{Conf}(\mathcal{L}_{\text{copy}}) \simeq \{(\theta, \theta, \theta, \theta) \mid \theta \in [-\delta, \delta]\}.$$

Crossover Gadget. The crossover gadget $\mathcal{L}_{\text{cross}}$ drawn in Figure 2-9b constrains opposite sliceform angles to remain equal. Specifically, $\text{Diff}_{(\Lambda_1, \Lambda_2, \Lambda_3, \Lambda_4)}$ provides a homeomorphism

$$\text{Conf}(\mathcal{L}_{\text{cross}}) \simeq \{(\theta_1, \theta_2, \theta_1, \theta_2) \mid \theta_1, \theta_2 \in [-\delta, \delta]\}.$$

Angle Average Gadget. The angle average gadget $\mathcal{L}_{\angle \text{avg}}$ drawn in Figures 2-9c and 2-11 constrains one sliceform angle to equal the average of the other two. Specifically, $\text{Diff}_{(\Lambda_1, \Lambda_2, \Lambda_3)}$ provides a homeomorphism

$$\text{Conf}(\mathcal{L}_{\angle \text{avg}}) \simeq \{(\theta_1, \theta_2, \theta_3) \mid \theta_1, \theta_3 \in [-\delta, \delta] \text{ and } \theta_2 = (\theta_1 + \theta_3)/2\}.$$

Angle Summation Gadget. The angle summation gadget $\mathcal{L}_{\angle \text{sum}}$ drawn in Figure 2-9d constrains one sliceform angle to equal the sum of the other two: $\text{Diff}_{(\Lambda_1, \Lambda_2, \Lambda_3)}$ provides a homeomorphism

$$\text{Conf}(\mathcal{L}_{\angle \text{sum}}) \simeq \{(\theta_1, \theta_2, \theta_3) \mid \theta_1, \theta_2, \theta_3 \in [-\delta, \delta] \text{ and } \theta_3 = \theta_1 + \theta_2\}.$$

Proof. Throughout this proof, we use the convention that $\text{Diff}_{\Theta_i}(C) = \theta_i$.

Copy Gadget Consider first the copy gadget $\mathcal{L}_{\text{copy}}$ of Figure 2-9a, whose movement is shown in Figure 2-10. Although the idea of the proof is quite simple, we explain it in full detail so that we may reuse the proof techniques for many future gadgets. We may assume $Q = 40$, because a larger Q results only in a larger scaled version of $\mathcal{L}_{\text{copy}}$, which therefore has a larger global minimum feature size.

By Lemma 2.7.9, the parallel gadgets force $d_j c_j \parallel e_{j+1} f_{j+1}$ for $1 \leq j \leq 4$ (with indices taken modulo 4) in any configuration of $\mathcal{L}_{\text{copy}}$.^{*} It follows that $\theta_1 = \theta_2 = \theta_3 = \theta_4$ in any configuration, i.e., $\varphi(\text{Conf}(\mathcal{L}_{\text{copy}})) \subset U$, where $\varphi = \text{Diff}_{(\Theta_1, \Theta_2, \Theta_3, \Theta_4)}$ and $U = \{(\theta, \theta, \theta, \theta) \mid \theta \in [-\delta, \delta]\}$.

^{*}Recall that we write v instead of $C(v)$ when the configuration C is understood from context.

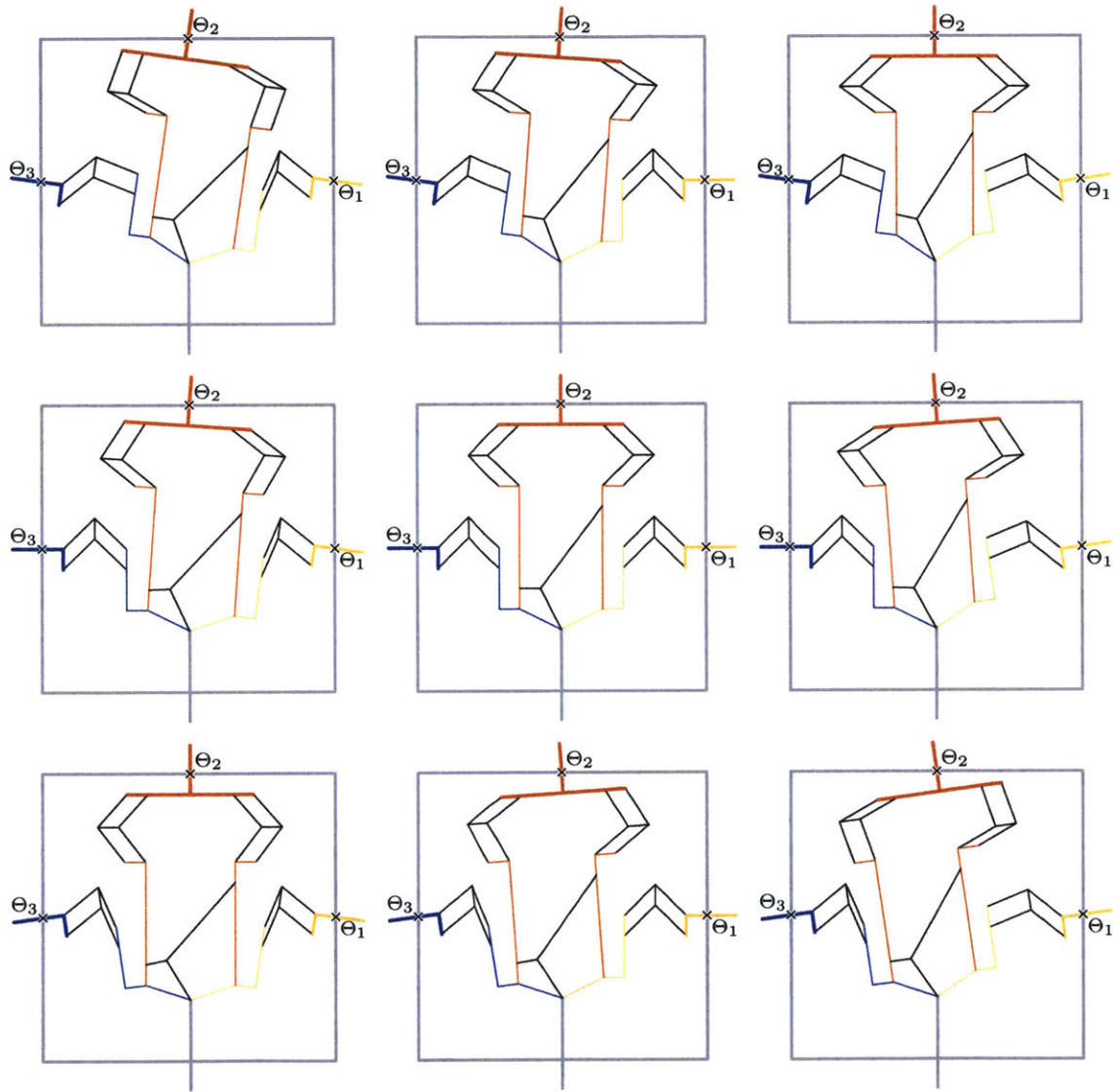


Figure 2-11: Snapshots of the Angle Average Gadget's configuration space (cf. Lemma 2.7.10, Figure 2-9c), shown for each pair of values $\theta_1, \theta_3 \in \{-8^\circ, 0^\circ, 8^\circ\}$ (where $\theta_i = \text{Diff}_{\theta_i}$). In each configuration, it may be seen that $\theta_2 = (\theta_1 + \theta_3)/2$. Edge lengths have been altered from those in Figure 2-9c to exaggerate the gadget's movement.

We next show that φ is a bijection onto U for small enough constants ε and δ , i.e., that there is a unique configuration C with $\varphi(C) = (\theta, \theta, \theta, \theta)$ for each $\theta \in [-\delta, \delta]$. Fix any constant $\varepsilon < \pi/4$. If $\varphi(C) = (\theta, \theta, \theta, \theta)$, then the position of each hook $b_j c_j d_j e_j f_j$ is uniquely and continuously determined by θ (this hook must be rotated by angle θ from its position in Figure 2-9a), and then by Lemma 2.7.9, the locations of the vertices in the parallel gadgets are uniquely and continuously determined, for θ in some neighborhood I around 0. By continuity, for θ in some (possibly smaller) neighborhood $I' \ni 0$, all angle chains will satisfy their ε -tolerances and the configuration will have minimum feature size $\geq \frac{1}{2}$. Choose δ small enough that $[-\delta, \delta] \subseteq I \cap I'$ (so that the δ -tolerances at the transmission angle chains Λ_j force $\theta \in [-\delta, \delta]$), which indeed guarantees that $\varphi : \text{Conf}(\mathcal{L}_{\text{copy}}) \rightarrow U$ is a bijection.

The above paragraph also shows that $\varphi^{-1} : U \rightarrow \text{Conf}(\mathcal{L}_{\text{copy}})$ is continuous, and because φ itself is continuous, φ is a homeomorphism, as claimed. (Alternatively, continuous bijection φ is a homeomorphism because $\text{Conf}(\mathcal{L}_{\text{copy}})$ and U are compact.)

Crossover Gadget The proof proceeds similarly to that of $\mathcal{L}_{\text{copy}}$ above. By scale-invariance of the crossover gadget $\mathcal{L}_{\text{cross}}$ of Figure 2-9b, we may assume $Q = 40$. In any configuration of $\mathcal{L}_{\text{cross}}$, the parallel gadgets and the sliceform vertex g enforce $b_3 c_3 \parallel f_3 g \parallel g f_1 \parallel c_1 b_1$ and likewise $b_4 g \parallel g f_2 \parallel c_2 b_2$. This shows that $\theta_1 = \theta_3$ and $\theta_2 = \theta_4$ throughout $\text{Conf}(\mathcal{L}_{\text{cross}})$. Conversely, it may be seen that the locations of all vertices of $\mathcal{L}_{\text{cross}}$ are uniquely and continuously determined by θ_1 and θ_2 , with each chosen independently in a small enough interval $[-\delta, \delta]$.

Angle Average Gadget In the gadget \mathcal{L}_{avg} depicted in Figure 2-9c (drawn to scale), the coordinates of all marked points are integer multiples of $Q/10$. It may be seen that kites $f_3 e_3 h e_2$ and $h e_1 f_1 e_2$ are similar.

The two upper parallel-gadgets guarantee that $e_3 f_3$ and $e_1 f_1$ (and $c_2 b_2$) remain parallel in any configuration, and in [5] it is shown that this is sufficient for the mechanism depicted to keep kites $f_3 e_3 h e_2$ and $h e_1 f_1 e_2$ similar in any configuration, which in turn guarantees that $c_2 b_2$ remains parallel to the angle bisector of $e_3 h e_1$ in any configuration. Thus, $\theta_2 = (\theta_1 + \theta_3)/2$ in any configuration. Conversely, any such triple of angles (in a small enough neighborhood of 0) gives rise to a unique configuration of \mathcal{L}_{avg} : indeed, θ_1 and θ_3 determine the rotation of $h e_1$ and $h e_3$ around (stationary) point h , and $\theta_2 = (\theta_1 + \theta_3)/2$ determines the rotations of $e_1 f_1$ and $e_3 f_3$. From here, e_2 's location is uniquely determined, and Lemma 2.7.9 shows that the parallel gadgets are also uniquely configured.

The proof may conclude with a continuity argument as used in the earlier gadgets. The movement of this gadget has been illustrated in Figure 2-11 (not to scale).

Angle Sum Gadget The Angle Sum gadget in Figure 2-9d is constructed in a 6×7 grid of $Q \times Q$ cells, with a transmission edge and sliceform at the midpoint of each edge (most not depicted). The cells labeled A_1 and A_2 are angle average gadgets (\mathcal{L}_{avg}), where one of A_2 's transmission sliceforms (drawn in gray) has been modified with a 0-tolerance instead of a δ -tolerance, fixing that input angle to 0. The other gray cells are copy gadgets, $\mathcal{L}_{\text{copy}}$. Transmission sliceforms between two white (unused) cells are frozen with 0 tolerance.

Suppose ε and δ are chosen small enough for the above claims regarding \mathcal{L}_{avg} and $\mathcal{L}_{\text{copy}}$ to hold. Then for any configuration C of \mathcal{L}_{sum} with $\theta_j = \text{Diff}_{\Theta_j}(C)$ and $\theta'_j = \text{Diff}_{\Theta'_j}(C)$ for $1 \leq j \leq 4$, the copy gadgets ensure that $\theta_j = \theta'_j$ for $1 \leq j \leq 4$; angle average gadget A_1 ensures that $\theta_4 = (\theta_1 + \theta_2)/2$; and angle average gadget A_2 ensures θ_4 is the average of

0 and θ_3 , i.e., $\theta_3 = 2 \cdot \theta_4$. In other words, $\theta_3 = \theta_1 + \theta_2$, as required. Conversely, for any choice of $\theta_1, \theta_2, \theta_3$, each in $[-\delta, \delta]$ and satisfying $\theta_3 = \theta_1 + \theta_2$, the angle $\theta_3/2$ also lies in $[-\delta, \delta]$, so by the earlier parts of this Theorem, there is a unique configuration C of $\mathcal{L}_{\angle\text{sum}}$ with $\text{Diff}_{(\Lambda_1, \Lambda_2, \Lambda_3)}(C) = (\theta_1, \theta_2, \theta_3)$. This concludes the proof. \square

As described above, we also require gadgets to manipulate *vectors* that are specified in polar coordinates, either as $2r \cdot \text{Rect}(\alpha, \beta)$ or $R \cdot \text{Rect}(\alpha, \beta)$. We demonstrate those gadgets here.

Lemma 2.7.11 (Gadgets for Vector Manipulation). *For any small enough constant $\varepsilon > 0$ and small enough $\delta < \varepsilon$, whenever $r \geq 1$ and $Q \geq 40r$, the (ε, δ) -extended linkages below are globally noncrossing with global minimum feature size at least $\frac{1}{2}$, and their configuration spaces are exactly parameterized as described below. Unless otherwise specified, each gadget is described by an initial configuration where all coordinates are integer multiples of $Q/40$.*

Start Gadget. *The start gadget $\mathcal{L}_{\text{start}}$ drawn in Figure 2-12a converts the rectangular coordinates of vertex v to the angles α, β in its polar representation. Specifically, the function $\pi_v \times \text{Diff}_{\Lambda, \Gamma}$ provides a homeomorphism of $\text{Conf}(\mathcal{L}_{\text{start}})$ with the set*

$$\{((x, y), (\alpha, \beta)) \mid \alpha, \beta \in [-\delta, \delta] \text{ and } (x, y) = (Q/5, Q/5) + (2r, 2r) + 2r \text{Rect}(\alpha, \beta)\}.$$

Vector Average Gadget. *The vector average gadget \mathcal{L}_{avg} (Figure 2-12c) constrains one vector to be the average of two others. Specifically, the function $\text{Diff}_{(\Lambda_1, \Gamma_1, \Lambda_2, \Gamma_2, \Lambda_3, \Gamma_3)}$ provides a homeomorphism of $\text{Conf}(\mathcal{L}_{\text{avg}})$ with*

$$\{(\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3) \in [-\delta, \delta]^6 \mid 2 \text{Rect}(\alpha_2, \beta_2) = \text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_3, \beta_3)\}.$$

Vector Sum Gadget. *The vector sum gadget \mathcal{L}_{sum} (Figure 2-13) constrains one vector to be the sum of two others. Specifically, $\text{Diff}_{(\Lambda_1, \Gamma_1, \Lambda_2, \Gamma_2, \Lambda_3, \Gamma_3)}$ provides a homeomorphism of $\text{Conf}(\mathcal{L}_{\text{sum}})$ with*

$$\{(\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3) \in [-\delta, \delta]^6 \mid \text{Rect}(\alpha_3, \beta_3) = \text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_2, \beta_2)\}.$$

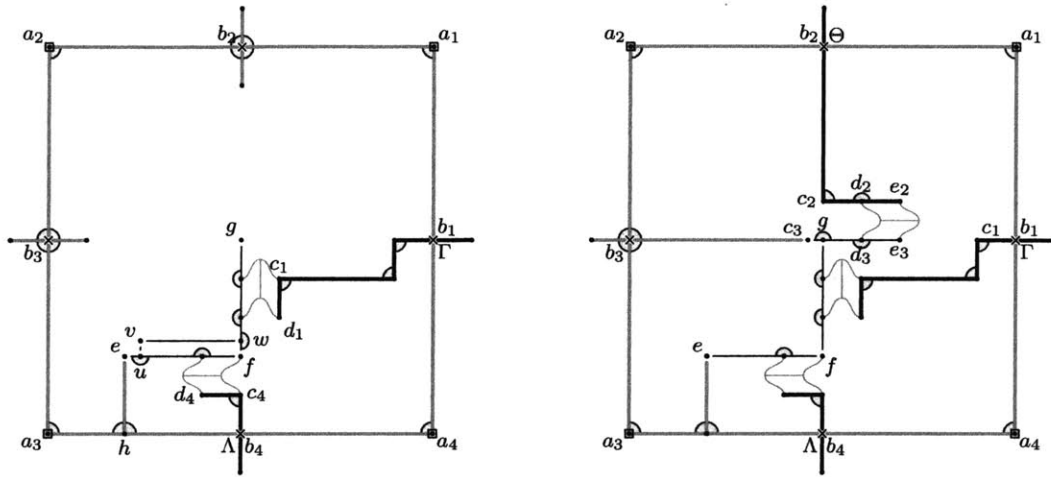
Vector Creation Gadget. *For a positive number $1 \leq w \leq 3\delta Q/40$, the vector creation gadget $\mathcal{L}_{\text{create}}(w)$ (Figure 2-12b) computes polar coordinates for a vector with complex representation $w \cdot (e^{i\theta} - 1)$. Specifically, $\text{Diff}_{(\Theta, \Lambda, \Gamma)}$ provides a homeomorphism of $\text{Conf}(\mathcal{L}_{\text{create}}(c))$ with*

$$\{(\theta, \alpha, \beta) \in [-\delta, \delta]^2 \mid w \cdot (\cos \theta, \sin \theta) - (w, 0) = R \cdot \text{Rect}(\alpha, \beta)\}.$$

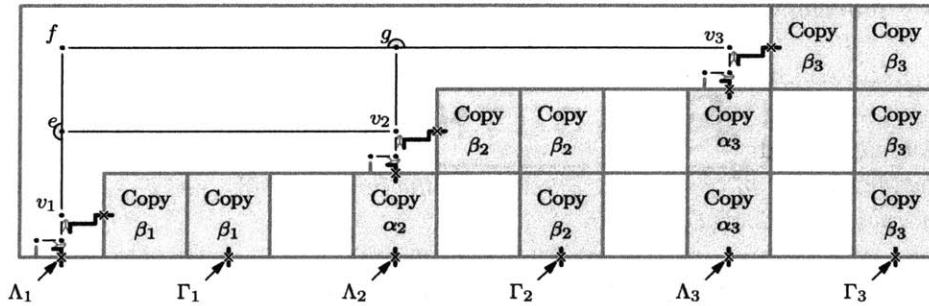
Vector Rotation Gadget. *The vector rotation gadget \mathcal{L}_{rot} (Figure 2-12d) computes the 90° rotation of a given vector. Specifically, $\text{Diff}_{(\Lambda_1, \Gamma_1, \Lambda_2, \Gamma_2)}$ is a homeomorphism of $\text{Conf}(\mathcal{L}_{\text{rot}})$ with*

$$\{(\alpha_1, \beta_1, \alpha_2, \beta_2) \in [-\delta, \delta]^4 \mid \text{Rect}(\alpha_1, \beta_1) = i \cdot \text{Rect}(\alpha_2, \beta_2)\}.$$

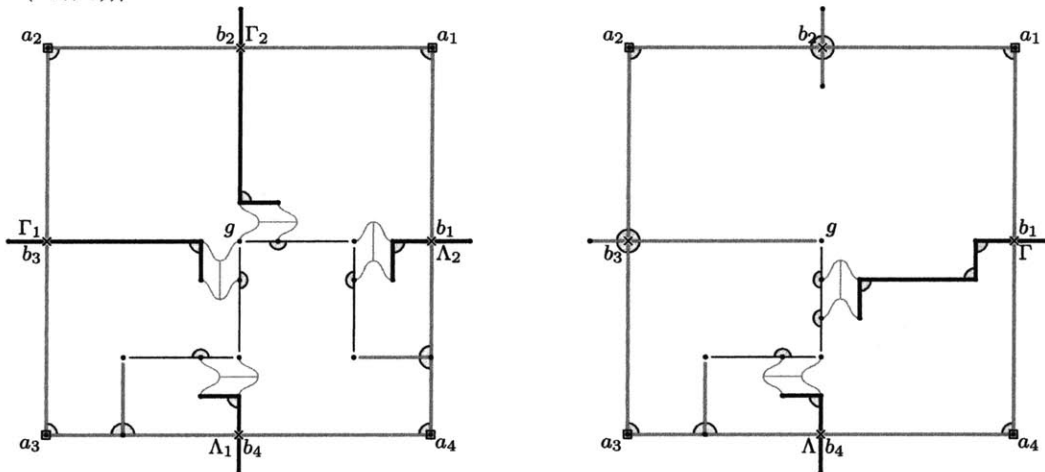
End Gadget. *For any chosen w with magnitude at most $3Q\delta/20$, the end gadget $\mathcal{L}_{\text{end}} = \mathcal{L}_{\text{end}}(w)$ (Figure 2-12e) constrains a vector to the position $(Q/2 + w, Q/2)$. Specifically, $\text{Conf}(\mathcal{L})$ has exactly one configuration C , and this configuration has $R \text{Rect}(\alpha, \beta) = (w, 0)$, where $(\alpha, \beta) = \text{Diff}_{(\Lambda, \Gamma)}(C)$. When $c = 0$, the coordinates of all vertices in $\mathcal{L}_{\text{end}}(0)$ are integer multiples of $Q/40$.*



(a) The Start Gadget forces $v = 2r \cdot \text{Rect}(\alpha, \beta)$. (b) The Vector Creation Gadget forces $R \cdot \text{Rect}(\alpha, \beta) = w \cdot (e^{i\theta} - 1)$.



(c) The Vector Average Gadget forces $v_2 = (v_1 + v_3)/2$, i.e., $\text{Rect}(\alpha_2, \beta_2) = (\text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_3, \beta_3))/2$.



(d) The Vector Rotation Gadget forces $\text{Rect}(\alpha_1, \beta_1) = i \cdot \text{Rect}(\alpha_2, \beta_2)$. (e) The End Gadget forces $R \cdot \text{Rect}(\alpha, \beta) = (w, 0)$.

Figure 2-12: Gadgets for manipulating vectors in the Main Construction. In these figures we use the conventions that $\text{Diff}_\Lambda(C) = \alpha$ and $\text{Diff}_\Gamma(C) = \beta$, possibly with subscripts, in addition to the conventions of Figure 2-9.

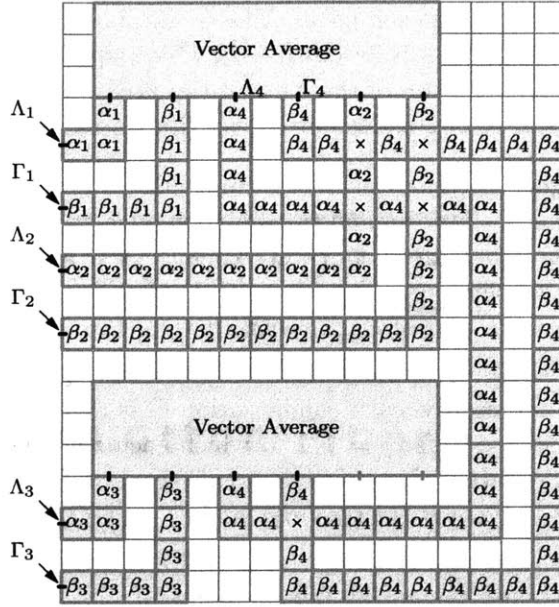


Figure 2-13: Vector Sum Gadget

Figure 2-14: The Vector Sum Gadget is built out of more primitive gadgets from Lemmas 2.7.10 and 2.7.11.

Proof. Throughout this proof, we follow the convention that $\text{Diff}_\Theta(C) = \theta$, $\text{Diff}_\Lambda(C) = \alpha$, $\text{Diff}_\Gamma(C) = \beta$, and similarly with subscripts. We consider each gadget in turn.

Start Gadget. This gadget does *not* scale with Q , because the parameters Q and r are not coupled. Indeed, in our application, our choice of r will be $o(Q)$, so we cannot directly appeal to scale invariance as with previous gadgets.

We sidestep this difficulty as follows: let $\mathcal{L}'_{\text{start}}$ be identical to $\mathcal{L}_{\text{start}}$ except that edges uv and vw are removed, and the two edges meeting at u are merged, as are the two edges meeting at w . We first prove that $\text{Conf}(\mathcal{L}'_{\text{start}})$ is homeomorphic to $[-\delta, \delta]^2$ via the map $\varphi' = \text{Diff}_{\Lambda, \Gamma}$, all of its configurations are noncrossing with minimum feature size at least $1/2$, and no vertex moves more than $Q/60$ from its position in Figure 2-12a. Note that the modified linkage $\mathcal{L}'_{\text{start}}$ *does* scale linearly with Q , so it suffices to prove this only for $Q = 40$. Indeed, the parallel gadgets guarantee that the angles $\alpha = \text{Diff}_\Lambda(C)$ and $\beta = \text{Diff}_\Gamma(C)$ uniquely and continuously determine the angles of ef and fg respectively and thus the location of every vertex of $\mathcal{L}'_{\text{start}}$. By continuity arguments as before, we may choose $\delta < \varepsilon < \pi/8$ small enough for the desired conclusion about $\mathcal{L}'_{\text{start}}$ (when $Q = 40$, and therefore for any $Q \geq 40$). Note that vertex g has location

$$e + R \cdot (\cos \alpha, \sin \alpha) + R \cdot (\sin \beta, -\cos \beta) = (Q/2, Q/2) + R \cdot \text{Rect}(\alpha, \beta).$$

Now back to $\mathcal{L}_{\text{start}}$. Given angles $\alpha, \beta \in [-\delta, \delta]$, let C' be the unique configuration of $\mathcal{L}'_{\text{start}}$ they determine. By Lemma 2.7.8 applied to parallelogram $vu'fw$, C' extends to a unique configuration C of $\mathcal{L}_{\text{start}}$, which has $u = e + 2r \cdot (\cos \alpha, \sin \alpha)$, $w = f + 2r \cdot (\sin \beta, -\cos \beta)$, and v at the fourth vertex of the parallelogram, namely $v = e + 2r \cdot \text{Rect}(\alpha, \beta)$ (as claimed). The minimum feature size of C either equals that of C' (which is $\geq 1/2$) or it

involves u, v, w , or any of their adjacent edges. The smallest such feature is either $2r \cdot \cos \alpha$ (this is $\text{dist}(u, eh)$ when $\alpha \geq 0$), or it is $2r \cdot \cos(\alpha - \beta)$, which equals $\text{dist}(f, vw)$ or $\text{dist}(u, vw)$ depending on the sign of $\alpha - \beta$. Both of these quantities are at least $2r \cdot \cos \pi/4 > r > 1/2$, so the minimum feature size of C is indeed at least $1/2$, as desired.

Vector Average Gadget. As with the Angle Sum gadget of Lemma 2.7.10, each gray $Q \times Q$ cell has a transmission sliceform and transmission edge at the midpoint of each edge, but the transmission edges that are not drawn explicitly and are not adjacent to a copy gadget are frozen with a 0 tolerance. The cells with lower left corners $(0, 0)$, $(4Q, Q)$, and $(8Q, 2Q)$ are populated with copies of $\mathcal{L}'_{\text{start}}$ (without their top and left edges).

This gadget scales with Q , so we may prove only the case $Q = 40$. Ignoring the edges adjacent to e, f , and g temporarily, each configuration C is determined uniquely and continuously by the six angles α_j, β_j for $j = 1, 2, 3$, which each vary independently in $[-\delta, \delta]$. Furthermore, in this configuration C , vertices v_1, v_2, v_3 are located at $v_1 = (Q/2, Q/2) + R \cdot \text{Rect}(\alpha_1, \beta_1)$, $v_2 = (9Q/2, 3Q/2) + R \cdot \text{Rect}(\alpha_2, \beta_2)$, and $v_3 = (17Q/2, 5Q/2) + R \cdot \text{Rect}(\alpha_3, \beta_3)$.

Now recall e, f, g , and their edges. By Lemma 2.7.8, fev_2g must remain a parallelogram, and therefore v_2 must remain at the midpoint of v_1v_3 , proving $2 \text{Rect}(\alpha_2, \beta_2) = \text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_3, \beta_3)$, as claimed. Furthermore, for δ small enough (by continuity), the ε angle tolerances at e, f, g, v_1, v_2, v_3 will be satisfied for all sextuples of angles in $[-\delta, \delta]^6$ and the minimum feature size of the entire gadget will remain greater than $1/2$, as required.

Vector Sum Gadget In this gadget illustrated in Figure 2-13, each $Q \times Q$ cell labeled with an angle α_j or β_j is a copy gadget, and each cell with an “x” is a crossover gadget. As with prior multi-cell gadgets, all unused transmission edges are frozen with tolerance 0, as are the two transmission sliceforms illustrated in gray instead of black at the right of the bottom-most vector average gadget.

For this gadget, there is no need for subtle continuity arguments. Instead, we may appeal directly to the characterizations of the copy, crossover, and vector average gadgets. The copy and crossover gadgets faithfully duplicate angles α_j and β_j (for $j = 1, \dots, 4$) as indicated in the Figure. The vector average gadgets then constrain $2 \text{Rect}(\alpha_4, \beta_4) = \text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_2, \beta_2)$ and $2 \text{Rect}(\alpha_4, \beta_4) = \text{Rect}(\alpha_3, \beta_3) + \text{Rect}(0, 0)$, i.e.,

$$\text{Rect}(\alpha_3, \beta_3) = \text{Rect}(\alpha_1, \beta_1) + \text{Rect}(\alpha_2, \beta_2) \quad (2.4)$$

Conversely, we must show that there exists a unique configuration C corresponding to any sextuple of angles in $[-\delta, \delta]$ satisfying equation (2.4). To that end, let $Y = \text{Rect}([-\delta, \delta]^2)$, which (by Lemma 2.7.5) is star-shaped with center $(0, 0)$. Because $p = \text{Rect}(\alpha_3, \beta_3) \in Y$, it follows that $p/2 \in R$, and thus there exist a unique pair of angles $(\alpha_4, \beta_4) \in [-\delta, \delta]^2$ with $\text{Rect}(\alpha_4, \beta_4) = p/2$. By our knowledge of $\text{Conf}(\mathcal{L}_{\text{copy}})$, $\text{Conf}(\mathcal{L}_{\text{cross}})$, and $\text{Conf}(\mathcal{L}_{\text{avg}})$, a unique configuration of \mathcal{L}_{sum} can be reconstructed from angles $\alpha_j, \beta_j, j = 1, 2, 3, 4$. And because each individual gadget comprising \mathcal{L}_{sum} has global minimum feature size at least $1/2$, so does \mathcal{L}_{sum} itself.

Vector Creation Gadget In $\mathcal{L}_{\text{create}}$ of Figure 2-12b, stationary point c_3 is located at $(Q/2 - w, Q/2)$, but all other marked vertices have initial coordinates that are integer multiples of $Q/10$ as shown. This gadget does not scale with Q : we cannot apply straightforward continuity arguments to $(40/Q) \cdot \mathcal{L}_{\text{create}}$ as we have done with earlier gadgets because w/Q is not bounded below by a constant.

The bottom half is a copy of $\mathcal{L}'_{\text{start}}$, so vertex g is located at $(Q/2, Q/2) + R \cdot \text{Rect}(\alpha, \beta)$. The other parallel gadget forces $c_3 \vec{g} = w \cdot (\cos \theta, \sin \theta)$, and therefore $R \cdot \text{Rect}(\alpha, \beta) = w \cdot (\cos \theta, \sin \theta) - (w, 0)$, as claimed.

Conversely, suppose an angle $\theta \in [-\delta, \delta]$ is given; we must show there exists a unique configuration C with $\text{Diff}_\Theta(C) = \theta$, and that this configuration is noncrossing with minimum feature size at least $1/2$. The position of g is uniquely determined by θ , and furthermore, $|g - (Q/2, Q/2)| \leq |g - c_3| + |c_3 - (Q/2, Q/2)| = 2w < R\delta/2$, so by Lemma 2.7.5, there is a unique pair of angles (α, β) , each in $[-\delta, \delta]$, with $R \cdot \text{Rect}(\alpha, \beta) = g - (Q/2, Q/2)$. This uniquely determines the rest of the configuration, but we must verify (1) δ may be chosen small enough (but still constant!) for Lemma 2.7.9 to apply to d_3, e_3, d_2, e_2 ; (2) that the ε -tolerances are obeyed; and (3) the feature size stays above $1/2$.

For part (1), it may be computed that

$$|d_2 - d_3 - (0, Q/10)| = \frac{2}{5} \cdot |\sin(\theta/2)| \cdot \sqrt{4Q^2 + 25w^2} \leq \frac{2}{5} \sin(\delta/2) \cdot 3Q,$$

which uses the weak bound $w < 3\delta Q/40 < Q/\sqrt{5}$. By decreasing δ , this bound may be reduced to any desired constant multiple of $Q/10$ (the latter is the initial length of $d_2 d_3$), so Lemma 2.7.9 indeed applies.

For part (2), the only angle chains where the ε -tolerance might break are at g ; indeed, all other angle chains have been argued already via either $\mathcal{L}'_{\text{start}}$ or by $\mathcal{L}_{\text{parallel}}$. We have $\angle f g d_3 = \pi/2 + \theta - \beta$, whose deviation from $\pi/2$ is at most $|\theta| + |\beta| \leq 2\delta < \varepsilon$.

Finally, for part (3), for small enough constant δ , the minimum feature size of any configuration of $\mathcal{L}_{\text{create}}$ is either $\text{dist}(c_3, fg) = w \cos(\beta - \theta) > w/2 \geq 1/2$ (if $\theta \geq \beta$) or simply $|c_3 g| = w > 1/2$ (if $\theta \leq \beta$).

Vector Rotation Gadget Vertex g is located at both $(Q/2)(1 + i) + R \cdot \text{Rect}(\alpha_1, \beta_1)$ and $(Q/2)(1 + i) + R \cdot i \cdot \text{Rect}(\alpha_2, \beta_2)$, which must therefore be equal. The rest of the proof by continuity is analogous to previous gadgets.

End Gadget The end gadget $\mathcal{L}_{\text{end}}(0)$ is depicted in Figure 2-12e; in general, $\mathcal{L}_{\text{end}}(w)$ consists of $\mathcal{L}'_{\text{start}}$ with an additional edge $b_3 g$ of length $Q/2 + w$ that freezes g to the point $(Q/2 + w, Q/2)$. Because $(w, 0) \in [-R\delta/2, R\delta/2]^2 \subseteq R \text{Rect}([-\delta, \delta]^2)$ by Lemma 2.7.5, the claimed configuration indeed exists (and is unique). When $w = 0$, we have $\alpha = \beta = 0$, and the coordinates of all nodes in the unique configuration of $\mathcal{L}_{\text{end}}(0)$ are integer multiples of $Q/40$. (When $w \neq 0$ we make no promises about the niceness of these angles and coordinates.) \square

2.7.4 Combining the Gadgets

It's time to use these gadgets to construct an extended linkage that computes a piece of $Z(F)$, up to a translation. Recall that $F = \{f_1, \dots, f_s\}$ is a family of polynomials in $\mathbb{R}[x_1, y_1, \dots, x_m, y_m]$, each with total degree at most d . We write each polynomial f_r , for $1 \leq r \leq s$, in polar form as in Theorem 2.7.6:

$$f_r(\vec{x}\vec{y}(\vec{\alpha}\vec{\beta})) = f_r(0) + \sum_{u=0}^3 \sum_{I \in \text{Coeffs}(2m, d)} i^u \cdot d_{r,u,I} \cdot \left(\exp(i \cdot (I \cdot \vec{\alpha}\vec{\beta})) - 1 \right), \quad (2.5)$$

where the numbers $d_{r,u,I}$ are nonnegative. If necessary, scale the polynomials f_j up until all nonzero coefficients $d_{r,u,I}$ are at least 1. (If the polynomials f_j have integer coefficients, then the $d_{r,u,I}$ are already integers by Theorem 2.7.6, so this scaling is unnecessary.) Suppose now that all coefficients of f_1, \dots, f_s (as standard polynomials in \overline{xy}) have magnitude at most M .

Suppose real numbers ε and $\delta < \varepsilon/2$ are given that are small enough according to Theorems 2.7.9, 2.7.10 and 2.7.11, and fix these for the remainder of this subsection. We choose the other parameters as follows: $r = \lceil d/\delta \rceil$, $Q = 40 \lceil 56dr^d M \binom{2m+d}{d} / (3\delta) \rceil$, and $R = 3Q/10$.

The extended linkage $\mathcal{E} = \mathcal{E}(F)$ built in this section will draw a translation of $Z(F) \cap (2r \cdot \text{Rect}([- \delta/d, \delta/d]^2))^m$, which by Lemma 2.7.5 contains $Z(F) \cap [-1, 1]^2$.

Step 1: Grid. The extended linkage \mathcal{E} is built within a grid of $Q \times Q$ cells with tolerance-0 corners and **transmission edges** attached at cell edge midpoints as described above. The transmission vertices (which are sliceform vertices) initially have tolerance 0 on their angle chains. Add three pins to three noncollinear vertices of this grid, so \mathcal{E} is globally rigid. There will be no other pins anywhere in \mathcal{E} ; pins illustrated in the gadgets above are removed before use here.

To build the requisite computation circuit, gadgets will be added to this grid, with their transmission vertices' angle chains upgraded to tolerance- δ as required by the gadgets.

Step 2: Start Gadgets to set up $\alpha_1, \beta_1, \dots, \alpha_m, \beta_m$. Add m start gadgets $\mathcal{L}_{\text{start}}(j)$ (for $1 \leq j \leq m$) to \mathcal{E} as illustrated in Figure 2-15, and let v_j be the vertex in $\mathcal{L}_{\text{start}}(j)$ corresponding to vertex v in Figure 2-12a; these vertices $X = \{v_1, \dots, v_m\}$ will be the drawing vertices of \mathcal{E} . Define angle chains Λ_j and Γ_j analogously, with $(\alpha_j, \beta_j) = (\alpha_j(C), \beta_j(C)) = \text{Diff}_{\Lambda(j), \Gamma(j)}(C)$ for a configuration C . By Lemma 2.7.11, $\text{Conf}(\mathcal{E})$ is perfectly described by $Y = (\Lambda_1, \Gamma_1, \dots, \Lambda_m, \Gamma_m)$; the map Diff_Y is a homeomorphism of $\text{Conf}(\mathcal{E})$ with $[-\delta, \delta]^{2m}$.

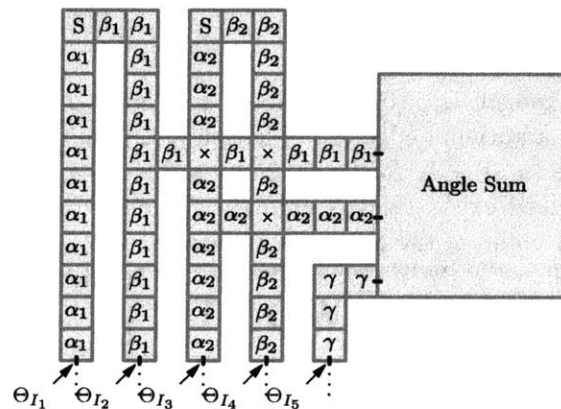


Figure 2-15: Computing the sum $\gamma = \beta_1 + \alpha_2$. Cells with “S” are start gadgets; those with “x” are crossover gadgets; and those with α_j , β_j , or γ are copy gadgets.

Step 3: Compute all linear combinations θ_I . In the next step, we modify \mathcal{E} further to compute all the angles $\theta_I := I \cdot \overrightarrow{\alpha\beta}$, where $\overrightarrow{\alpha\beta} = (\alpha_1, \beta_1, \dots, \alpha_m, \beta_m)$ and I ranges over the nonzero entries of $\text{Coeffs}(2m, d)$.

As base cases, the angles α_j and β_j (corresponding to $I = (0, \dots, 1, \dots, 0)$) are already computed by the start gadgets. Use copy gadgets to build **wires** that copy these angle values vertically along columns, as shown in Figure 2-15. We call this collection of wires the **wire column**, and it will accumulate other wires computing other angles as the construction continues.

By ordering all the remaining vectors $I \in \text{Coeffs}(2m, d)$ according to the sum of the absolute values of I 's entries, each successive θ_I may then be computed as $\theta_I = \theta_{I'} \pm \alpha_j$ or $\theta_I = \theta_{I'} \pm \beta_j$ using a single addition gadget. To do this, as illustrated in Figure 2-15, we use crossover gadgets to transmit the desired input angles to the addition gadget, and then the computed θ_I joins the wire column. Repeat this for each remaining $I \in \text{Coeffs}(2m, d)$.

For each I , let Θ_I be a(ny) transmission angle chain in the wire coming from the computation of θ_I above. We claim that $\text{Conf}(\mathcal{E})$ is still perfectly described by the angle chains $Y = (\Lambda_1(1), \Lambda_2(1), \dots, \Lambda_1(m), \Lambda_2(m))$, but this time on a smaller range: we'll show that Diff_Y is a homeomorphism of $\text{Conf}(\mathcal{E})$ with $[-\delta/d, \delta/d]^{2m}$, and furthermore, that the angle chains Θ_I indeed compute $\text{Diff}_{\Theta_I}(C) = \theta_I(C)$. The latter claim follows from induction on I . To see that the α_j and β_j are each precisely (and independently) constrained to the interval $[-\delta/d, \delta/d]$, first consider the angle chain Θ_I where $I = (d, 0, 0, \dots, 0)$. We have $\theta_I = d \cdot \alpha_1$, and the δ -tolerance at angle chain Θ_I requires this value to remain in $[-\delta, \delta]$, so α_1 must remain in $[-\delta/d, \delta/d]$. The other coordinates in $\vec{\alpha\beta}$ are likewise constrained by analogous arguments. In the other direction, for any vector $\vec{\alpha\beta} \in [-\delta/d, \delta/d]^{2m}$ and any $I \in \text{Coeffs}(2m, d)$, it may be seen that $I \cdot \vec{\alpha\beta}$ lies in the interval $[-\delta, \delta]$. By Lemma 2.7.10, then, any such vector $\vec{\alpha\beta}$ may be realized by a unique configuration of \mathcal{E} .

Step 4: Compute the terms in the polar expansions of f_1, \dots, f_s . For each $1 \leq r \leq s$ and each nonzero coefficient $d_{r,u,I}$ in the polar representation of $f_r(\vec{x}\vec{y}(\vec{\alpha\beta}))$, insert a vector creation gadget using angle θ_I and coefficient $d_{r,u,I}$, and if $u \neq 0$, chain this with u vector rotation gadgets. The result computes the desired vector $i^u \cdot d_{r,u,I} \cdot (\exp(i \cdot \theta_I) - 1)$: specifically, the result is a pair of wires with transmission angle chains $\Gamma_{r,u,I,1}$ and $\Gamma_{r,u,I,2}$ whose angles $\gamma_{r,u,I,1}(C) = \text{Diff}_{\Gamma_{r,u,I,1}}(C)$ and $\gamma_{r,u,I,2}(C) = \text{Diff}_{\Gamma_{r,u,I,2}}(C)$ satisfy

$$R \cdot \text{Rect}(\gamma_{r,u,I,1}, \gamma_{r,u,I,2}) = i^u \cdot d_{r,u,I} \cdot (\exp(i \cdot \theta_I) - 1).$$

The vector creation gadgets are safe to use because $d_{r,u,I} \leq 6^{d_r d} M \binom{2m+d}{d} < 3\delta Q/40$, as required in Lemma 2.7.11. Likewise, the vector being sent through each vector rotate gadget has magnitude at most $2d_{r,u,I} \leq 3\delta Q/20 = R\delta/2$, so the vector rotation gadget offers no obstruction by Lemma 2.7.5, so the map Diff_Y is still a homeomorphism of $\text{Conf}(\mathcal{E})$ with $[-\delta/d, \delta/d]^{2m}$.

Step 5: Add the vectors to compute $f_r(\vec{\alpha\beta}) - f_r(0)$. For each $1 \leq r \leq s$, use vector addition gadgets to successively compute the sum of the nonzero vectors among $R \cdot \text{Rect}(\gamma_{r,u,I,1}, \gamma_{r,u,I,2})$, resulting in a single pair wires with transmission angle chains $\Gamma_{r,1}, \Gamma_{r,2}$ whose angle offsets $\gamma_{r,1}, \gamma_{r,2}$ compute

$$R \cdot \text{Rect}(\gamma_{r,1}, \gamma_{r,2}) = (f_r(\vec{\alpha\beta}) - f_r(0), 0).$$

(Join these wires to the wire column, as usual.) Each vector computed in this fashion is a partial sum of equation 2.5, which by Theorem 2.7.6 has magnitude at most

$2 \cdot 6^d r^d M \binom{2m+d}{d} < R\delta/2$, which fits comfortably within the working regions of the vector addition gadgets. Thus, $\text{Diff}_Y : \text{Conf}(\mathcal{E}) \rightarrow [-\delta/d, \delta/d]^{2m}$ is still a homeomorphism.

Step 6: Conclude with End Gadgets Finally, for each $1 \leq r \leq s$, feed the $\gamma_{r,1}$ and $\gamma_{r,2}$ wires into an end gadget that enforces the constraint $R \cdot \text{Rect}(\gamma_{r,1}, \gamma_{r,2}) = (-f_r(0), 0)$. (Note that the vector $R \cdot \text{Rect}(\gamma_{r,1}, \gamma_{r,2})$ has y -coordinate 0, so as used in this construction, the end gadget simply constraints its vector to a specified point on this line.) The configuration space of the resulting linkage \mathcal{E} is then homeomorphic to the subset of $\vec{\alpha\beta} \in [-\delta/d, \delta/d]^{2m}$ that satisfies the constraints of these end gadgets, i.e.,

$$\text{Conf}(\mathcal{E}) = \left\{ \vec{\alpha\beta} \in [-\delta/d, \delta/d]^{2m} \mid f_r(\vec{\alpha\beta}) = 0 \right\}.$$

And because the offset of each vertex v_j from the lower-left corner of its cell is precisely $(Q/5, Q/5) + (2r, 2r) + 2r \cdot \text{Rect}(\alpha_j, \beta_j) = (Q/5, Q/5) + (2r, 2r) + (x_j, y_j)$, these vertices $\{v_1, \dots, v_m\}$ indeed draw a translation of

$$Z(F) \cap (2r \cdot \text{Rect}([-\delta/d, \delta/d], [-\delta/d, \delta/d]))^m,$$

as required.

2.7.5 Sliceform and Angle Constraint Gadgets

We must now modify the extended linkage $\mathcal{E}(F)$ constructed above into a partially constrained linkage (with no other constraints) and verify that it satisfies the various and sundry requirements of the Main Theorem. We proceed with the help of two gadgets: the **sliceform gadget** obviates the need for sliceform vertices, and the **angle restrictor gadgets** enforce \mathcal{E} 's rotation constraint and angle constraint after replacing each edge with a rigidified tree.

Theorem 2.7.12 (Sliceform Gadget). *Whenever $\delta < \pi/4$, the (ε, δ) -extended linkage that has four edges meeting at a sliceform vertex with tolerance δ is perfectly simulated by the (ε, δ) -extended linkage $\mathcal{L}_{\text{slice}}$, all of whose angle chains have tolerance δ (see Figure 2-16).*

Proof. By Lemma 2.7.8 applied to the preponderance of parallelograms, all parallelograms in the Figure remain parallelograms in all configurations. This, paired with the 0-tolerance angles keeping collinear edges collinear, ensures the desired collinearity. \square

The last gadget is built from *partially rigidified* linkages instead of *extended* linkages. Recall that a **partially rigidified** linkage is a constrained linkage where all constraints are **rigid constraints** $\text{RigidCon}(H, C_H)$, each forcing a certain subgraph H to maintain a chosen, rigid shape given by configuration C_H .

Theorem 2.7.13 (Angle Restrictor Gadget). *For a positive integer n , let $\mathcal{L}_{\text{restrict}} = \mathcal{L}_{\text{restrict}}(n)$ be the partially constrained linkage shown in Figure 2-17, with two rigid constraints fixing orthogonal trees A and B in the configurations shown (but not governing their relative position or orientation). Tree B is also pinned rigidly to the plane, for convenience.*

All of the following hold for sufficiently large integers n : $\mathcal{L}_{\text{restrict}}$ is globally noncrossing, and vertex a_1 continuously and rigidly draws the locus

$$\{(\ell \sin \theta, -\ell \cos \theta) \mid \pi/2 - \gamma \leq \theta \leq \pi/2 + \gamma\},$$

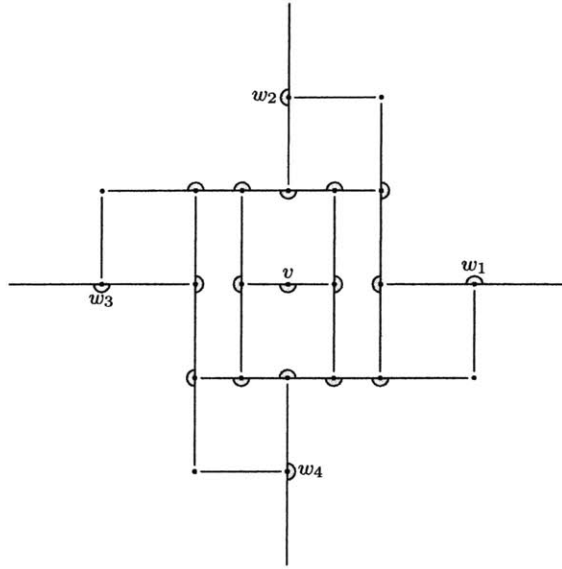


Figure 2-16: The Sliceform gadget keeps w_1, v, w_3 and w_2, v, w_4 collinear.

where the angle $\gamma = \gamma(n)$ tends toward 0 as n goes to infinity. In each configuration of $\mathcal{L}_{\angle\text{restrict}}$, tree A is configured with the same orientation as in Figure 2-17, and every pair of consecutive edges around each vertex forms an angle strictly between 60° and 240° . Furthermore, there is a unique configuration of $\mathcal{L}_{\angle\text{restrict}}$ having $\angle b_1 o a_1 = \pi/2$.

The limiting case $\mathcal{L}_{\angle\text{restrict}}(\infty)$ is globally rigid.

Proof. First we specify the initial configuration (drawn solid in Figure 2-17) more precisely. Let $\kappa = \sin^{-1}(5/13)$, which is a **Pythagorean angle**, meaning $\sin(\kappa) = 5/13$ and $\cos(\kappa) = 12/13$ are both rational, and define another Pythagorean angle $\varphi = \sin^{-1}(2n/(n^2 + 1))$ (its cosine is $(n^2 - 1)/(n^2 + 1) \in \mathbb{Q}$). In the initial configuration, $a_2 = (\cos \kappa, \sin \kappa)$, $c_1 = 2a_2$, $\angle b_3 c_1 o = \angle o c_1 b_2 = \pi/2 - \varphi$, and d_1 is the midpoint of $c_1 b_2$. Edges $c_1 d_1$, $d_1 b_2$, and $c_1 b_3$ have length t , t , $2t$ respectively, where $t = 3/20$ is some small constant. The other assembly, anchored at a_3, b_4, b_5 , is (initially) the reflection through the line $y = x$. Because we chose Pythagorean angles, all coordinates are initially rational.

We now investigate a general configuration C of $\mathcal{L}_{\angle\text{restrict}}$. Bar $b_3 c_1$ restricts c_1 to the dashed circle with radius $2t$, and bars $b_2 d_1$ and $d_1 c_1$ further restrict vertex c_1 to the disk with radius $2t$, as shown. As a result, c_1 must lie on the small circular arc $g_1 g_2$ (see Figure 2-17), centered at b_3 with angle 2φ . (Note that g_1 and g_2 are fixed points in the plane; they are not linkage vertices.) In turn, because $|o a_2| = |a_2 c_1| = 1$, vertex a_2 is confined to the circular arc $h_1 h_2$ centered at o with radius 1, where the endpoints of confinement, h_1 and h_2 , are determined by rhombus $o h_1 g_1 h_2$. Because o, g_1 , and g_2 are collinear, we may compute that $|o g_1| = 2 - 4t \sin \varphi$, and so $\gamma := \frac{1}{2} \angle h_1 o h_2 = \cos^{-1}(1 - 2t \sin \varphi)$. Because $\gamma \rightarrow 0$ as $n \rightarrow \infty$, we may choose n large enough so that arcs $h_1 h_2$ and $h_3 h_4$ are disjoint. This forces $\angle a_2 o a_3$ to have counterclockwise orientation, and so tree A must be oriented as claimed, and $\theta = \angle b_1 o a_1$ must remain in the closed interval between $\pi/2 \pm \gamma$.

Conversely, given an angle θ in this range, we'll argue there are either 1 or 4 configurations C having $\angle b_1 o a_1 = \theta$. This angle determines the position of a_2 along arc $h_1 h_2$, and then there is a unique point c_1 on arc $g_1 g_2$ with $|a_2 c_1| = 1$. If c_1 lies at g_1 (i.e., $\theta = \pi/2$), then because $|b_2 c_1| = 2t$, vertex d_1 must be configured at the midpoint of $b_2 c_1$. The case is similar

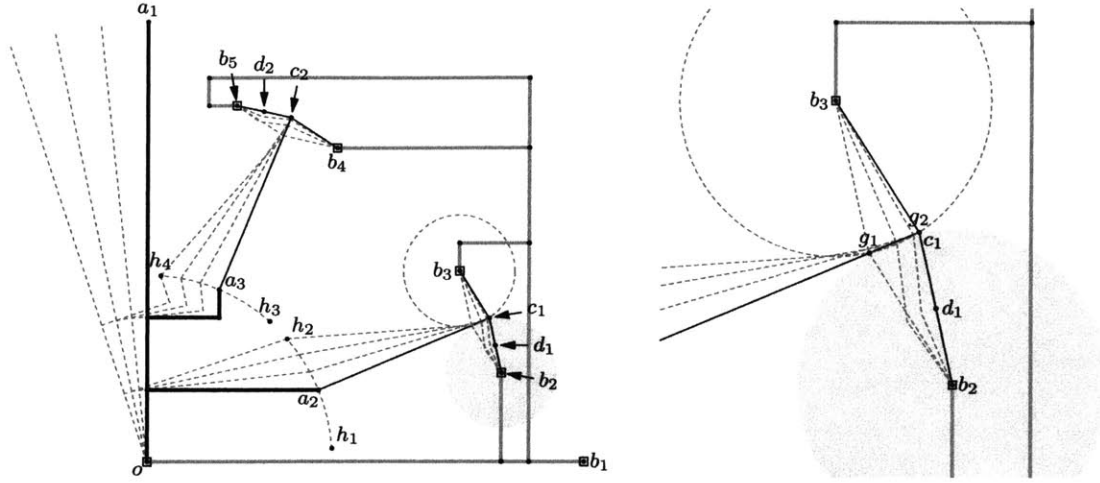


Figure 2-17: Angle Restrictor Gadget, $\mathcal{L}_{\angle restrict}$, shown in full (left) and closeup (right).

if c_1 is at g_1 , i.e., $\theta = \pi/2 \pm \gamma$. Otherwise, d_1 lies in the interior of the solid gray disk, so there are two possible choices for the location of d_1 . The same is true for the other assembly anchored at a_3, b_4 , and b_5 , and furthermore, vertices d_1 and d_2 may vary continuously with θ . This proves that vertex a_1 draws continuously and rigidly.

The initial configuration C_0 is noncrossing, and in this initial configuration, the angles between consecutive edges around each vertex lie strictly between 60° and 240° . For large enough n (meaning smaller φ and smaller γ), these properties may be ensured for *all* configurations of $\mathcal{L}_{\angle restrict}$, by continuity.

In the limiting case $n = \infty$, the angles κ and φ are 0, so the resulting linkage $\mathcal{L}_{\angle restrict}(\infty)$ is globally rigid. \square

2.7.6 Implementing Extended Linkages with Partially Rigidified Linkages

Choose positive integers n_ε and n_δ large enough such that ε and $\delta < \varepsilon/2$ are small enough global constants according to Theorems 2.7.9, 2.7.10 and 2.7.11, where

$$\varepsilon := \cos^{-1} \left(1 - \frac{3}{10} \cdot \frac{2n_\varepsilon}{n_\varepsilon^2 + 1} \right) \quad \text{and} \quad \delta := \cos^{-1} \left(1 - \frac{3}{10} \cdot \frac{2n_\delta}{n_\delta^2 + 1} \right).$$

Also ensure n_ε and n_δ are large enough for Theorem 2.7.13, so the angle restrictor gadgets $\mathcal{L}_{\angle restrict}(n_\varepsilon)$ and $\mathcal{L}_{\angle restrict}(n_\delta)$ are globally noncrossing and implement angle tolerances $\gamma = \varepsilon$ and $\gamma = \delta$ respectively. Use these values of ε and δ to build linkage $\mathcal{E}(F)$ from polynomials F as described in Section 2.7.4.

We now transform this extended linkage \mathcal{E} into a partially rigidified linkage $\mathcal{L} = \mathcal{L}(F)$ that simulates \mathcal{E} , not perfectly, but at least continuously and rigidly.

Extended linkage \mathcal{E} has all edge-lengths at least 1 and has global minimal feature size at least $1/2$, because each gadget individually has these properties. Therefore, if a new linkage $\mathcal{E}' = \mathcal{E}'(F)$ is obtained by replacing each δ -tolerance sliceform vertex v in \mathcal{E} with a sliceform gadget $\mathcal{L}_{\angle slice}$ (scaled so that $|vw| = 1/8$), then \mathcal{E}' perfectly simulates \mathcal{E} , is globally noncrossing, and can be checked to have global minimum feature size at least $\frac{1}{64}$. All other sliceform vertices in \mathcal{E} have tolerance 0, i.e., are rigid, so removing them from the sliceform set S in \mathcal{E}' does not affect $\text{Conf}(\mathcal{E}')$. So \mathcal{E}' has no sliceform vertices at all.

Finally, we transform \mathcal{E}' into a partially rigidified linkage $\mathcal{L} = \mathcal{L}(F)$ as follows: first, consider each edge (u, v) of \mathcal{E}' as a rigidified tree $T_{(u,v)}$ initially containing just one edge. For each angle chain (u, v, w) with $A(u, v, w) = \pi/2$, we modify trees $T_{(u,v)}$ and $T_{(v,w)}$ in a $\frac{1}{256}$ neighborhood of v by attaching a scaled angle restrictor gadget, using $\mathcal{L}_{\angle\text{restrict}}(n\delta)$, $\mathcal{L}_{\angle\text{restrict}}(n\varepsilon)$, or $\mathcal{L}_{\angle\text{restrict}}(\infty)$ depending on whether $\Delta(u, v, w)$ equals δ , ε , or 0 respectively. If $A(u, v, w)$ is π or $3\pi/2$, attach an edge to $T_{(u,v)}$ at right angle to $T_{(v,w)}$ before inserting the appropriate angle restrictor gadget. And if $A(u, v, w) = 2\pi$ (i.e., if $u = w$ and $\deg(v) = 1$), nothing need be done. Finally, remove the three pins in \mathcal{L} coming from the pins in \mathcal{E}' (that serve only to fix the background grid in place), and replace them with three noncollinear pins in one of the rigidified trees built from the background grid (which equally fix the background grid in place).

This concludes the construction of the desired linkage $\mathcal{L}(F)$; below we show that this linkage satisfies all requirements of the Main Theorem.

Theorem 2.7.14. *This linkage $\mathcal{L}(F)$, built from $F = \{f_1, \dots, f_s\}$ as in Section 2.7.4, satisfies all of the requirements of Theorem 2.2.13.*

Proof. Define $X \subset V(\mathcal{L})$ as the set of vertices of \mathcal{L} that simulate vertices v_1, \dots, v_m of \mathcal{E} , and define translation $T(\vec{x}\vec{y}) = \vec{x}\vec{y} + (a_1, b_1, \dots, a_m, b_m)$ where (a_j, b_j) are the coordinates of the center of the cells containing start gadget $\mathcal{L}_{\text{start}}(j)$.

The fact that \mathcal{L} continuously and rigidly simulates \mathcal{E}' (and thus simulates \mathcal{E}) follows from Lemma 2.7.13, and therefore \mathcal{L} indeed draws $T(Z(F) \cap R)$ continuously and rigidly, where $R = (2r \cdot \text{Rect}([-d/\delta, d/\delta]))^m$. We now verify the many conditions of the Main Theorem:

Part I. 1. Already discussed.

2. The total number of gadgets used in \mathcal{E} is bounded by $\text{poly}(\#\text{Coeffs}(F))$: indeed, following the computation steps, there are m start gadgets; at most $O(\text{Coeffs}(2m, d)) = O(\text{poly}(\#\text{Coeffs}(F)))$ angle addition, vector creation, vector rotation, and vector addition gadgets; and s end gadgets. Each of these gadgets occupies $O(1)$ cells of the grid, So the overall size of the grid used is $O(\text{poly}(\#\text{Coeffs}(F)) \times O(\text{poly}(\#\text{Coeffs}(F))))$ cells. Each cell is either empty or contains some gadget with $O(1)$ vertices and edges, proving the result.
3. We already concluded that \mathcal{E}' is globally noncrossing with global minimum feature size at least $\frac{1}{256}$. In the insertion of angle restrictor gadgets to transform \mathcal{E}' to \mathcal{L} , scale each gadget to fit within a $\frac{1}{1024}$ neighborhood of its vertex, which guarantees that \mathcal{L} is globally noncrossing with global minimum feature size at least $\frac{1}{2048}$.
4. Each rigidified tree $T_{(u,v)}$ in \mathcal{L} is constructed by attaching (a portion of) an angle restrictor, possibly with one extra edge, to one or both endpoints of edge (u, v) . This configured tree is indeed orthogonal and contains no unattached leaves.
5. The combinatorial embedding σ of \mathcal{L} follows from the combinatorial embedding of extended linkage \mathcal{E}' , augmented with the unique combinatorial embedding of angle restrictor gadgets. Any vertex of \mathcal{L} that is not internal to any rigidified tree either corresponds to a degree-4 vertex of \mathcal{E}' , or belongs to an angle restrictor gadget. In the former case, all four angle chains at v have values in $[\pi/2 - \varepsilon, \pi/2 + \varepsilon]$. The latter case follows from Theorem 2.7.13.
6. This is true by direct construction.

- Part II.** 7. When the coefficients of f_1, \dots, f_s are integers, all coefficients $d_{r,u,I}$ are also integers bounded by $\text{poly}(\text{Size}(F))$ (by Theorem 2.7.6). Parameters Q, r, R are also integers bounded by $\text{poly}(\text{Size}(F))$, so it may be checked all edge lengths of \mathcal{E} are integers bounded by Q . The scaled sliceform gadgets $\mathcal{L}_{\angle\text{slice}}$ inserted to form \mathcal{E}' have rational edge lengths with $O(1)$ denominators, so \mathcal{E}' shares this property. Finally, the scaled angle restrictor gadgets $\mathcal{L}_{\angle\text{restrict}}(n_\varepsilon)$, $\mathcal{L}_{\angle\text{restrict}}(n_\delta)$, an $\mathcal{L}_{\angle\text{restrict}}(\infty)$ have rational edge lengths whose denominators have constant size, so \mathcal{L} shares this property.
8. Computing polar forms $g_r(\overrightarrow{\alpha\beta})$ from polynomials $f_r(\overrightarrow{x\hat{y}})$ may be done in deterministic polynomial time by straightforward polynomial multiplication algorithms, and the coefficients never exceed $\text{poly}(\text{Size}(F))$ in magnitude. Our transformations from f_1, \dots, f_m to $\mathcal{L}(F)$ are explicit and deterministic.

Part III. 9. When the given polynomials f_j satisfy $f_j(\overrightarrow{0}) = 0$, our construction of $\mathcal{E}(F)$ uses end gadgets with input $w = 0$, which come with an initial configuration with coordinates that are integer multiples of $Q/40$ and therefore integers. All other gadgets have this property unconditionally. Together, these specify an initial configuration of $\mathcal{E}(F)$ corresponding to $\overrightarrow{\alpha\beta} = \overrightarrow{0}$, i.e., $\overrightarrow{x\hat{y}} = \overrightarrow{0}$, which indeed has integer coordinates.

The initial configurations of the sliceform gadgets and angle restrictor gadgets were also illustrated in Figures 2-16 and 2-17 and shown to have rational coordinates with constant-sized denominators. These induce the desired initial configuration C_0 of $\mathcal{L}(F)$.

10. Linkage \mathcal{E}' perfectly draws $T(Z(F) \cap R)$, meaning there is only one configuration of \mathcal{E}' mapping to $T(\overrightarrow{0})$. In this unique configuration, each angle restrictor gadget must simulate an angle of $\theta = \pi/2$, and by Theorem 2.7.13, there is only one such configuration for each gadget. So C_0 is indeed unique.
11. As described above, our construction of C_0 is explicit and deterministic.

All properties have been verified, so this concludes the proof. \square

2.7.7 Modifications for Strong Matchstick Universality

We may subtly modify the above proof of Theorem 2.2.13 to prove that the nontrivial subsets of \mathbb{R}^2 drawn by matchstick linkages are *exactly* the bounded semialgebraic sets. We use one extra cell gadget when constructing extended linkage $\mathcal{E}(F)$, the Crossing End gadget (Figure 2-18), which is used to create a crossing precisely when $g(\overrightarrow{x\hat{y}}) = 0$ for a given polynomial g . When linkage $\mathcal{E}(F)$ is simulated by a matchstick linkage $\mathcal{M}(F)$ as described in Section 2.6, all of $\mathcal{E}(F)$'s noncrossing configurations transfer to $\mathcal{M}(F)$, i.e., thickening does not introduce unintended crossings. This allows us to draw semialgebraic sets of the form

$$\{\overrightarrow{x} \in \mathbb{R}^k \in \mathbb{R}^2 \mid f_1(\overrightarrow{x}) = \dots = f_s(\overrightarrow{x}) = 0, g_1(\overrightarrow{x}) \neq 0, \dots, g_r(\overrightarrow{x}) \neq 0\},$$

as well as coordinate projections thereof. This is sufficient to draw any bounded semialgebraic set in the plane.

Theorem 2.7.15. *The nontrivial subsets $R \subseteq \mathbb{R}^2$ that are drawable by a matchstick linkage are exactly the bounded semialgebraic sets.*

Proof. Let $\mathcal{M} = (\mathcal{L}, \text{NXCon}_{\mathcal{L}})$ be a matchstick linkage with k vertices that is connected and has at least one pin. Then $\text{Conf}(\mathcal{L})$ is described by closed polynomial conditions and

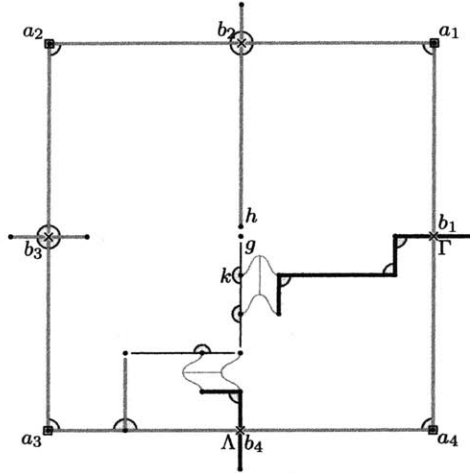


Figure 2-18: The Crossing End Gadget creates a crossing at $g = h$ precisely when $\alpha = \beta = 0$, assuming g remains on the line $y = Q/2$.

is therefore compact and semialgebraic. Noncrossing can be described by the nonvanishing of polynomials, so $\text{Conf}(\mathcal{M}) = \text{NConf}(\mathcal{L}) \subset \text{Conf}(\mathcal{L})$ is bounded and semialgebraic (but not necessarily closed). Then, for any vertex v , the trace $\pi_v(\text{Conf}(\mathcal{M}))$ is also bounded and semialgebraic.

Conversely, suppose $R \subset \mathbb{R}^2$ is a bounded, semialgebraic set. By Lemma 2.3.2, R may be written as the projection onto the first two coordinates of some bounded, basic semialgebraic set

$$R' = \{\vec{x} \in \mathbb{R}^{2m} \mid f_1(\vec{x}) = \dots = f_s(\vec{x}) = 0, g_1(\vec{x}) \neq 0, \dots, g_r(\vec{x}) \neq 0\}.$$

We may assume R' does not contain the origin (by translating if necessary), so by replacing each $g_j(\vec{x})$ by $g_j(\vec{x}) \cdot |\vec{x}|^2$ (which does not modify R'), we may further assume that each g_j satisfies $g_j(\vec{0}) = 0$. Finally, as in the proof of Theorem 2.6.8, we may assume $R' \subset [-1, 1]^{2m}$ and the coefficients of each f_j and g_j are integers.

We create a custom extended linkage cell gadget, the **crossing end gadget** $\mathcal{L}_{X\text{-end}}$, that is *not* globally noncrossing, but instead induces a crossing precisely when fed the zero vector:

Lemma 2.7.16 (Crossing End Gadget). *The crossing end gadget, $\mathcal{L}_{X\text{-end}}$, is a copy of $\mathcal{L}'_{\text{start}}$ with one additional edge of length $Q/2$. (It is drawn short to differentiate g from h , but in the initial configuration depicted, g and h actually coincide.) Let $H \subset \text{Conf}(\mathcal{L}_{X\text{-end}})$ consist of those configurations in which g has y coordinate equal to $Q/2$. Then the only configuration in H with a crossing has $g = (Q/2, Q/2)$. All configurations in H have minimum feature size at least $1/2$ if the distance between edges b_2h and kg is ignored.*

Proof. This follows from the analysis of $\mathcal{L}'_{\text{start}}$ in the proof of Theorem 2.7.11. \square

We may now proceed as in the proof of the Main Theorem (Theorem 2.2.13) using polynomials f_1, \dots, f_s and g_1, \dots, g_r , and concluding each g_r with an End Crossing gadget instead of an End gadget. (As in the proof of the Main Theorem, the pair of angles θ_1, θ_2 fed into a crossing end gadget will always satisfy $R \cdot \text{Rect}(\theta_1, \theta_2) = (g(\vec{x}\vec{y}(\vec{\alpha}\vec{\beta})) - g(0), 0)$.) For the resulting extended linkage \mathcal{E} with drawing vertices X , the map π_X is a homeomorphism

between $\text{NConf}(\mathcal{E})$ and a translation of R' , and furthermore, the only crossings in any of \mathcal{E} 's configurations come from the End Crossing gadgets.

As in the proof of Theorem 2.2.13, eliminate sliceforms in \mathcal{E} to form \mathcal{E}' and process this further (as before) into partially rigidified linkage \mathcal{L} . Finally, modify \mathcal{L} into a matchstick linkage \mathcal{M} as in Construction 2.6.4. We claim that each noncrossing configuration C of \mathcal{E} gives rise to a noncrossing configuration of \mathcal{M} , i.e., no *unintended* crossings arise in \mathcal{M} 's underlying linkage. Supposing the contrary, configuration C must have minimum feature size less than $1/2$, so by Lemma 2.7.16, the crossing must happen in the vicinity of a crossing end gadget: the edge polyiamonds ultimately built from edges kg and b_2h of $\mathcal{L}_{X\text{-end}}$ must intersect. But the former edge polyiamond lies entirely in the lower half of the gadget, and the latter edge polyiamond lies in the upper half, so they could intersect only at h . But then C is a crossing configuration of \mathcal{E} , contrary to assumption. \square

Chapter 3

Self-Touching Linkages and the Expansive Self Touching Carpenter’s Rule Theorem

Chapter Summary. We propose an alternative paradigm for working with *self-touching* planar linkages, which allow bars to butt up against each other so long as the linkage does not cross through itself. Such overlapping bars are often informally described as being “infinitesimally” offset from each other to one side or the other; we propose simply making this intuition formal. The resulting model remains geometrically intuitive and simple to work with, and furthermore, thanks to the powerful Tarski-Seidenberg principle and some new tools we develop in the theory of real fields with infinitesimals, much existing work with *nontouching* linkages easily transfers to the self-touching case.

In particular, we focus on the Expansive Carpenter’s Rule Theorem, that any nontouching configuration of a chain linkage can be unfolded without overlapping itself and, more strongly, so that each pair of nodes monotonically move apart during the motion. We show that the same holds for self-touching linkages, and furthermore, we show how this follows as a *formal consequence* of its truth for non-touching linkages. We relate our self-touching linkage model to earlier models, concluding that the Expansive Carpenter’s Rule Theorem holds for them as well. We then explore consequences of this result to self-touching chains adorned with polygons or other regions attached to the bars.

3.1 Introduction

In this chapter we revisit the Carpenter’s Rule Theorem: given any chain of line segments (which may be imagined as a “carpenter’s rule”) drawn in the plane without self-touching, it is possible to continuously straighten or convexify the chain while keeping edge lengths constant and while avoiding self touching (cf. Figure 1-2 on page 19). In fact, more is true: it is possible to find an *expansive* motion, one where every pair of vertices of the chain moves monotonically farther apart throughout the motion. We will call this the Expansive Carpenter’s Rule Theorem.

The Expansive Carpenter’s Rule Theorem was demonstrated twice at the 2000 Symposium on the Foundations of Computer Science, by Connelly, Demaine, and Rote [32] (full version [34]) and independently by Streinu [78] (full version [79]), using very different techniques: the former relies on equilibrium stresses and energy minimization (see also the

purely energy-based approach by Cantarella et al. [28]), while the latter is a combinatorial argument using pointed pseudotriangulations.

This Expansive Carpenter’s Rule Theorem assumes that the initial chain configuration is strictly simple, i.e., it does not touch itself, even though it may come very close to intersecting. In this chapter we study a natural follow-up question: does the result still hold if the initial chain configuration is *self-touching*, i.e., is allowed to rest right up against itself (without passing through)? This is a lower-dimensional analog of a folded piece of idealized zero-thickness origami paper, which can fold onto itself—perhaps many times over or in intricate arrangements—but cannot pass through itself. More recently, Abbott, Demaine, and Gassend proved that the Carpenter’s Rule Theorem is indeed true for self-touching configurations [1], but they left the existence of an *expansive* motion open. They argued in [1], as we shall vehemently reiterate in this chapter, that the primary challenge in proving the Self-Touching Carpenter’s Rule Theorem is not the *intrinsic difficulty* of the theorem but instead the lack of effective *language* to describe and discuss self-touching linkage configurations and their motions. Indeed, it has proven surprisingly challenging to express the intuitive idea of “touching without passing through” in a way that is formally precise yet simple enough to use in practice; this remains an important foundational pursuit in the theory of linkages (discussed below) as well as origami (discussed in detail in Chapter 4).

3.1.1 Our Results

We propose a radically new approach to modelling self-touching, grounded firmly in the intuitive language of infinitesimal offsets, with many advantages over prior models. Indeed, in Section 3.3, we show that this model quickly settles the Expansive Self-Touching Carpenter’s Rule Theorem, revealing it to be a *formal consequence* of its non-touching analog via the Tarski-Seidenberg “transfer” principle for real closed fields. In Section 3.4 we introduce *gradual functions* as a generally useful tool for projecting semialgebraic sets over $\mathbb{R}\langle\varepsilon\rangle$ down to \mathbb{R} by “forgetting” infinitesimals, which allows us to prove in Section 3.5 that the Expansive Self-Touching Carpenter’s Rule Theorem holds not only in our infinitesimal model but also in the prior “combinatorial” and “limiting” models from [33] and [1], respectively. (In the process, we also locate and fix a small error in this original combinatorial model.) Finally, we note that the results of [31] easily transfer to the self-touching setting: specifically, if a (possibly self-touching) linkage has “regions” or **adornments** attached to its bars, and if these adornments initially do not overlap (in a manner to be made precise) and have a specially defined *slender* shape, then they cannot intersect during an expansive motion.

But before all of that, in the remainder of this Introduction we carefully motivate our new *infinitesimal* model of self-touching linkages and briefly contrast it with the two prior models.

3.1.2 Motivation for Our Infinitesimal Model

Self-touching linkage configurations are usually formalized in two different pieces: the ambiguous base *geometry* (positions of all the nodes in the plane) and then, separately, the *layering* information to annotate and disambiguate the geometry.

Our approach is able to specifying both at once instead, in a geometrically intuitive way: proceeding informally for a moment, imagine giving each vertex a position (in \mathbb{R}^2) and possibly also a *tiny* offset, in terms of a small positive number ε . In the example in Figure 3-1, the location of vertex A is $(0 + \varepsilon, -1 + 2\varepsilon)$, meaning A geometrically lies at

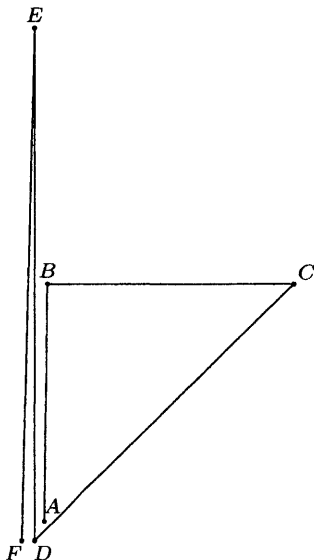


Figure 3-1: A visual representation of a self-touching linkage in our infinitesimal model. The precise vertex locations are listed in equation (3.1). Spaces between touching bars are exaggerated; they are actually infinitesimal.

$(0, -1)$ but, for layering purposes, we imagine it has been shifted by a tiny vector in the direction $(1, 2)$. The full coordinate specification for the linkage in Figure 3-1 is as follows:

$$\begin{aligned}
 A &= (0 + \varepsilon, -1 + 2\varepsilon) & B &= (0, 0 + \varepsilon) & C &= (2, 0) & (3.1) \\
 D &= (0 + \varepsilon, -1) & E &= (0 - \varepsilon, 1) & F &= (0, -1).
 \end{aligned}$$

We may compute that triangle EDB has signed area

$$\frac{1}{2} \det \begin{pmatrix} 0 - \varepsilon & 1 & 1 \\ 0 + \varepsilon & -1 & 1 \\ 0 & 0 + \varepsilon & 1 \end{pmatrix} = \varepsilon^2,$$

and because this “number” is (small but) positive, this should indicate that B is located “to the right” of segment ED . So long as the offsets (in terms of ε) are “compatible” in some way so as to avoid overlaps, these coordinates would hopefully specify the geometry *and* the layering in one go.

This fruitful line of thought may be understood more formally in terms of limits, as was done in [1]: formula (3.1) describes the configuration obtained by taking the limit as ε decreases toward 0. This makes sense as long as the bars don’t intersect each other when ε is small enough. Such a “compatibility condition” ensures that the result is a continuous limit of nontouching linkages, a fairly intuitive geometric view of what a “self-touching linkage” should mean.

So far, this line of reasoning is very similar to the model of self-touching linkages defined in [1]. Our model diverges with this next idea, which keeps our model easy to work with while preserving, and perhaps amplifying, the simple geometric intuition captured above. We avoid the need to work directly with limits by considering ε *not* as a positive real number tending toward 0, but as a **formal infinitesimal**: a transcendental value smaller than any

positive real number. So instead of working with a limit of *many* configurations (with coordinates in \mathbb{R}), we may instead consider a *single* linkage configuration whose coordinates are chosen from a larger ordered field, $\mathbb{R}(\varepsilon)$. (This field $\mathbb{R}(\varepsilon)$ is the fraction field of the polynomial ring $\mathbb{R}[\varepsilon]$ and the smallest field containing the transcendental variable ε .) The “compatibility” condition simplifies as well: this single linkage configuration (over $\mathbb{R}(\varepsilon)$) must be nontouching (over $\mathbb{R}(\varepsilon)$).

Just one more idea brings us to the exact model proposed in this chapter: use a slightly large field in place of $\mathbb{R}(\varepsilon)$. We choose the real closure of $\mathbb{R}(\varepsilon)$, the field $\mathbb{R}\langle\varepsilon\rangle$ of algebraic Puiseux series, which is another ordered field and is a $\sqrt{-1}$ -extension away from being algebraically closed, like \mathbb{R} itself. This choice gives us access to the powerful Tarski-Seidenberg principle or transfer principle, which says that any first-order statement (in the language of real ordered fields) that holds for \mathbb{R} is also true for $\mathbb{R}\langle\varepsilon\rangle$ or any other real-closed field containing \mathbb{R} . For example, the Expansive Carpenter’s Rule Theorem for nontouching linkages (over \mathbb{R}) may be expressed in this way, so the Tarski-Seidenberg principle guarantees it holds for nontouching linkages over $\mathbb{R}\langle\varepsilon\rangle$ as well. Since our model of self-touching linkages is precisely defined by nontouching linkages over $\mathbb{R}\langle\varepsilon\rangle$, this suffices to prove the Expansive Self-Touching Carpenter’s Rule Theorem in our proposed model! (This argument will be more carefully executed in Sections 3.2 and 3.3.)

3.1.3 Two Prior Models of Self-Touching Linkages

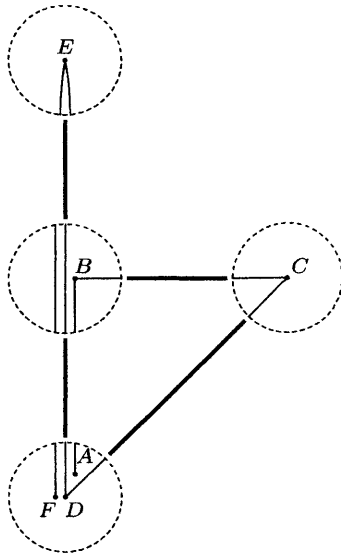
For comparison, we briefly review the primary pre-existing strategies for modeling self-touching linkages. There are two main approaches: describing layer ordering *combinatorially* on top of the underlying (self-intersecting) geometry, or describing the configuration as a *limit* of nontouching geometry.

On the combinatorial side, the authors of [33] define a **combinatorial self-touching configuration** of a linkage G via the geometry $C : V(G) \rightarrow \mathbb{R}^2$ (whose bars may overlap but not properly cross)* where additionally, at each node location $p \in \mathbb{R}^2$, we must specify a **magnified view** M_p around p , which is a plane graph drawn within a disk that represents the combinatorial “stacking order” of all edges and nodes that pass through or lie at p . This graph can be imagined intuitively as a schematic view of an “infinitesimal neighborhood” of p , and in fact we will make this comparison precise in Section 3.5.2. To illustrate, the self-touching linkage from Figure 3-1 has been redrawn in Figure 3-2a according to this combinatorial model.

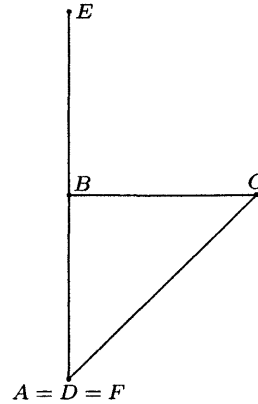
This model also specifies which *motions* are valid by forbidding three types of violations: a vertex penetrating an edge, a vertex penetrating a chain of two edges, or an edge crossing over another edge. Each violation must be avoided by *both* the macroscopic geometry *and* the microscopic magnified views. We will review these constraints in more detail as they are needed in the proof of Theorem 3.5.1 below.

While straightforward, this definition has proven too complicated and cumbersome to be easily used in practice. In our opinion, this is primarily due to its inherent discontinuous nature: the validity of a *continuous* motion is defined by its behavior during certain *discontinuous* changes in the magnified views, such as when two nodes move together and their magnified views instantaneously merge. This inconvenience is illustrated more concretely in the necessarily long and repetitive proof of Theorem 3.5.1 below.

*Two segments in \mathbb{R}^2 are said to **properly cross** if they are not parallel and their relative interiors intersect, necessarily at a single point.



(a) A self-touching configuration visualized according to the combinatorial model. The self-intersecting base geometry is shown in bold. Each magnified view gives a *combinatorial* view of the edge layering in a neighborhood of every vertex, in the form of a plane graph.



$e_1 \backslash e_2$	AB	BC	CD	DE	EF
AB	0	0	1	-1	-1
BC	0	0	1	0	0
CD	$1/\sqrt{2}$	$1/\sqrt{2}$	0	$\sqrt{2}$	$\sqrt{2}$
DE	1	0	1	0	-2
EF	-1	0	-1	-2	0

(b) A self-touching configuration visualized according to the limiting model: the base geometry together with a matrix of pairwise edge annotations describing edge sidedness.

Figure 3-2: A self-touching linkage configuration as represented by the combinatorial and limiting models.

By contrast, a limit-based alternative model has been proposed and utilized by Abbott, Demaine, and Gassend [1]. In this model, which we call the **limiting self-touching linkage** model throughout this chapter, a self-touching configuration takes the form (c, A) , where c describes the (self-intersecting) geometry as above, and $A : E(G) \times E(G) \rightarrow \mathbb{R}$ provides a real number “annotation” for each pair of edges that disambiguates layer ordering for overlapping pairs. (These annotations will be explained in detail in Section 3.5.2.) A configuration (c, A) is declared valid if it is the limit of *nontouching* linkage configurations that are likewise annotated. A self-touching linkage illustrated according to this model is shown in Figure 3-2b.

While equivalent to the combinatorial model above in the linkage configurations and motions they can represent (as proven in [1]), the limiting model has a number of clear advantages over the combinatorial model. It is intrinsically *continuous*, i.e., the representation (c, A) changes continuously as the corresponding linkage moves continuously. Furthermore, it concisely represents the entire space of self-touching configurations as a closed semialgebraic set, instead of defining this configuration space indirectly through constraints on paths. This concision crucially facilitates their proof of (a non-expansive form of) the Self-Touching Carpenter’s Rule Theorem, as well as the brevity of our Section 3.5.1 relating this model to our infinitesimal one.

But there are some drawbacks as well. It is based on the solid geometric intuition of nearby nontouching perturbations, but this intuition gets clouded by the details of defining and interpreting the annotations: indeed, Figure 3-2b, which shows a literal representation of a self-touching configuration in this model, is not easily interpreted, even after learning

the meanings of the annotations (these are described in Section 3.5.2). Relatedly, these annotations are highly specialized, relying on particular properties of linkages in the plane (specifically, their Ord function) without addressing fundamental challenges of more general self-touching situations, such as linkages in higher ambient dimensions.

3.2 Background on $\mathbb{R}\langle\varepsilon\rangle$ and Real Closed Fields

We summarize definitions and facts that we will need about real closed fields, the field of real algebraic Puiseux series $\mathbb{R}\langle\varepsilon\rangle$, and the Tarski-Seidenberg principle here. For more information, and for verifications of the below claims, see [17, Chap. 2] and [25].

3.2.1 Real Closed Fields

A **real field** or an **ordered field** is a field R together with a total ordering on its elements such that $x \leq y \implies x + z \leq y + z$ and $x \geq 0, y \geq 0 \implies x \cdot y \geq 0$. If R has no nontrivial extensions that are both algebraic and real, R is called a **real closed field**; equivalently, R is real closed precisely when $R[\sqrt{-1}] = R[x]/(x^2 + 1)$ is algebraically closed. Any real field can be algebraically extended to a real closed field compatible with its ordering in an essentially unique way [25, Thm. 1.3.2]. For example, the real closure of \mathbb{Q} is the set of real algebraic numbers \mathbb{R}_{alg} , while \mathbb{R} is its own real closure.

For a real field R , a **formula in the language of ordered fields with coefficients in R** is a first-order logical formula built from statements of the form $P = 0$ or $P \geq 0$, where P is a polynomial in some number of variables with coefficients in R , and where these inequalities are connected with logical operators \wedge, \vee and \neg and variable quantifiers \exists and \forall . A formula with no unbound variables is called a **sentence**. The following “transfer principle” strongly relates the first-order theories of two nested real closed fields:

Theorem 3.2.1 (Tarski-Seidenberg principle, [17, Thm. 2.80]). *If R is a real closed field contained in a larger real closed field R' , and if Ψ is a sentence in the language of ordered fields with coefficients in R , then Ψ is true in R precisely when it is true in R' .*

This principle leverages “quantifier elimination” for real closed fields: any formula in the language of ordered fields with coefficients in R can be replaced by an equivalent formula that avoids the quantifiers \forall and \exists .

For increased clarity when multiple real fields are being used simultaneously, we may affix a subscript to an element or interval to specify which field it belongs to, such as $0_{\mathbb{R}}$ or $[0, 1]_{\mathbb{R}_{\text{alg}}}$.

3.2.2 Adjoining Infinitesimals

If R is a real field and ε a transcendental variable over R , there is a unique ordering on the field $R(\varepsilon)$ of rational functions in which ε is positive but **infinitesimal**, i.e., smaller than any positive element of R . We will always assume $R(\varepsilon)$ is equipped with this ordering.

When R itself is real closed, the real closure of $R(\varepsilon)$ may be succinctly described in terms of Puiseux series as follows: a **Puiseux series** in ε over R is a formal Laurent series in $\varepsilon^{1/q}$ for some positive integer q , i.e., a series of the form $S = \sum_{j \geq k} a_j \varepsilon^{j/q}$, where k is a (not necessarily positive) integer, j ranges over the integers $\geq k$, q is a positive integer, and each a_j is in R . If $R\langle\varepsilon\rangle$ denotes the set of these Puiseux series that are algebraic over $R(\varepsilon)$, then $R\langle\varepsilon\rangle$ is the real closure of $R(\varepsilon)$.

The following Theorem, closely related to the Tarski-Seidenberg principle above, gives a precise way in which $\varepsilon \in R\langle\varepsilon\rangle$ behaves like an “arbitrarily small” element of R :

Theorem 3.2.2 (Transfer Principle for $R\langle\varepsilon\rangle$ [17, Prop. 3.17]). *If R is a real closed field and $\Psi(t)$ is a formula in the language of ordered fields with coefficients in R (where t is the only unbound variable), then $\Psi(\varepsilon)$ is true in $R\langle\varepsilon\rangle$ if and only if there is some positive $t_0 \in R$ such that $\Psi(t)$ holds in R for every $t \in (0, t_0)_R$.*

3.2.3 Sizes of Elements in $R\langle\varepsilon\rangle$

Let R be a real closed field as before. The ring $R\langle\varepsilon\rangle$ comes equipped with a valuation: for a Puiseux series S of the form $S = \sum_{j \geq k} a_j \varepsilon^{j/q}$ as above whose leading coefficient a_k is nonzero, its valuation is defined as $\nu_\varepsilon(S) := k/q$, the exponent of ε in the leading term. When S is identically 0, $\nu_\varepsilon(0) := \infty$.

An element S of $R\langle\varepsilon\rangle$ is **infinitesimal** if it is contained in the interval $(-r, r)$ for every positive $r \in R$, i.e., if $\nu_\varepsilon(S) > 0$, and we write $P \sim Q$ when $Q - P$ is infinitesimal. Similarly, S is **bounded** over R if it is contained in $(-r, r)$ for *some* element $r \in R$, i.e., if $\nu_\varepsilon(S) \geq 0$; otherwise, it is **unbounded**. Let $R\langle\varepsilon\rangle_b \subset R\langle\varepsilon\rangle$ denote the set of bounded elements.

3.2.4 Projecting and Lifting

Let R be a real closed field. For a bounded element $P \in R\langle\varepsilon\rangle_b$, let $\lim_\varepsilon P$ denote the coefficient of ε^0 in the Puiseux series P , i.e., $\lim_\varepsilon P$ is the result of replacing ε by 0 in P . Similarly, for a semialgebraic subset S of $R\langle\varepsilon\rangle^k$, define $\lim_\varepsilon S \subset R^k$ by $\lim_\varepsilon S := \{\lim_\varepsilon P \mid P \in S \cap R\langle\varepsilon\rangle_b^k\}$; this set $\lim_\varepsilon S$ is a semialgebraic and closed subset of R^k [17, Prop. 12.43].

In the other direction, if $S \subset R^k$ is a semialgebraic set, S may be lifted to a semialgebraic set $\text{Ext}(S, R\langle\varepsilon\rangle) \subset R\langle\varepsilon\rangle^k$ in a canonical way: $\text{Ext}(S, R\langle\varepsilon\rangle)$ is described by taking a(ny) quantifier-free formula for S with coefficients in R and reinterpreting the formula over domain $R\langle\varepsilon\rangle^k$. Note that $\text{Ext}(S, R\langle\varepsilon\rangle) \cap R^k = S$, and it may also be verified that $\lim_\varepsilon \text{Ext}(S, R\langle\varepsilon\rangle) = \overline{S}$, using the proof method of [17, Prop. 12.43].

3.3 Expansive Self-Touching Carpenter’s Rule Theorem

3.3.1 Self-Touching Linkages via Infinitesimals

Recall from Section 2.2.1 that an **(abstract) linkage** is an edge-weighted graph $G = (V(G), E(G), \ell_G)$ with positive edge lengths $\ell_G(e) > 0$; we do not consider pins in this chapter. A linkage is a **chain** if its underlying graph is either a path (**open chain**) or a cycle (**closed chain**). A **configuration** of a linkage G is a choice of vertex locations $C : V(G) \rightarrow \mathbb{R}^2$ with correct edge lengths, i.e., $|C(v) - C(w)| = \ell(v, w)$ for each edge $(v, w) \in E(G)$, and this configuration C is **nontouching** if distinct edges intersect only at common endpoints. Note that in Chapter 2 this concept was called **noncrossing**, as is customary in graph-drawing literature; but because this chapter focuses on distinguishing acceptable crossings (i.e., *self-touching*) from forbidden ones, the term **nontouching** better emphasizes when no form of self-intersection is allowed. The space $\text{NConf}(G)$ is the configuration space of all nontouching configurations of G , as in Chapter 2.

If $C : V(G) \rightarrow \mathbb{R}^2$ is a “configuration” of G whose edge lengths are within $\pm r$ of G ’s assigned lengths, C is called an **r -related** configuration of G . The configuration space of

nontouching, r -related configurations of G is denoted $\text{NConf}_r(G)$. This useful space was introduced in [1], and our proposed model is based on an infinitesimal version:

Definition 3.3.1. An **infinitesimal self-touching configuration** of a linkage G is defined to be an infinitesimally-related, nontouching configuration of G over $\mathbb{R}\langle\varepsilon\rangle$; in other words, a configuration $C : V(G) \rightarrow \mathbb{R}\langle\varepsilon\rangle^2$ that is nontouching (in $\mathbb{R}\langle\varepsilon\rangle^2$) and whose edge lengths differ from G 's edge lengths by infinitesimal amounts. We denote the set of all such configurations by $\text{NConf}_\sim(G)$. A **continuous motion of self-touching configurations** of G is defined simply as a continuous, semialgebraic path $[0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \text{NConf}_\sim(G)$.

Note. If $C : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \text{NConf}_\sim(G)$ is a continuous motion as above, we denote the configuration at a given time $T \in [0, 1]_{\mathbb{R}\langle\varepsilon\rangle}$ by C_T instead of $C(T)$, to highlight the fact that a motion is a parameterized family of configurations. Similarly, if T is unspecified, we use the same notation C_T to refer to the entire motion.

We will show that this model relates strongly to the other two models: Sections 3.5.1 and 3.5.2 show how to convert an infinitesimal self-touching configuration into a combinatorial or limiting configuration, and they prove that combinatorial or limiting self-touching configurations are precisely those that come from infinitesimal configurations. These theorems prove, furthermore, that continuous, semialgebraic paths in $\text{NConf}_\sim(G)$ may be converted to continuous, semialgebraic paths in the other two models. Paired with the nearly effortless proof of the expansive self-touching carpenter's rule theorem in the infinitesimal model (Theorem 3.3.3), this proves that expansive motions of combinatorial or limiting self-touching chain linkages always exist (Theorem 3.5.4).

We must note that the correspondence relating the infinitesimal model to the other two models is not perfect: a single combinatorial or limiting configuration has *many* different infinitesimal perturbations, so the correspondence is not one-to-one. Furthermore, we show that motions in the infinitesimal model may be transferred to the other two models, but we do *not* prove the reverse. Ares Ríbo Mor's result that every combinatorially self-touching configuration has arbitrarily close nontouching perturbations ([71, Theorem 3.1]) is already quite involved, and lifting (combinatorial or limiting) motions would amount to a dynamic version of this static theorem. Thankfully, such a result is not needed in this paper.

3.3.2 Expansive Self-Touching Carpenter's Rule Theorem in the Infinitesimal Model

We first state a precise form of the nontouching Expansive Carpenter's Rule Theorem, proven in [79].

Theorem 3.3.2 (Nontouching Expansive Carpenter's Rule Theorem). *For any nontouching configuration C_0 (over \mathbb{R}) of an open or closed chain linkage G , there is a motion $t \in [0, 1]_{\mathbb{R}} \mapsto \varphi_t \in \text{Conf}_{\mathbb{R}}(G)$ from $\varphi_0 = C_0$ to the canonical linear configuration φ_1 of G (if G is open) or a convex configuration φ_1 (if G is closed), such that φ is **expansive**: for any pair of (not necessarily adjacent) vertices $u, v \in V(G)$, the function $t \mapsto |\varphi_t(u) - \varphi_t(v)|$ is (weakly) increasing. Furthermore, this motion φ may be chosen to be semialgebraic, defined by a formula of size at most $F(n)$ over \mathbb{R} . For a chosen $v \in V(G)$, we may assume further that $\varphi_t(v) = (0, 0)$ for all $0 \leq t \leq 1$.*

Proof. All but the formula of size $F(n)$ is proved explicitly in [79]. To verify the expansive motion φ constructed there also has bounded formula size, note that φ is assembled from

at most n^3 semialgebraic paths, each one describable as follows: start with $\text{Conf}(G)$; forbid a subset of edge-lengths from changing; intersect with the set of configurations having $a < |C(u) - C(v)| < b$, where (u, v) is the removed convex-hull edge driving the motion; extract a connected component; compute the closure; and finally, apply a linear transformation to replace the parameter interval $[a, b]$ with the desired interval. Each step is a semialgebraic transformation of bounded complexity in terms of n —for the connected component step, this component can be written as a union of cells in any cylindrical decomposition of $\text{Conf}(G)$ [17, Thm. 5.21], and a cylindrical decomposition can be constructed with complexity and number of components bounded by a function of n [17, Thm 5.34]. \square

The self-touching version of this theorem follows for free, via the Transfer Principle (Theorem 3.2.1):

Theorem 3.3.3. *For any infinitesimally noncrossing configuration C_0 (i.e., a nontouching configuration over $\mathbb{R}\langle\varepsilon\rangle$) of an open chain linkage G , there is a weakly expansive motion (over $\mathbb{R}\langle\varepsilon\rangle$) from C to the canonical linear configuration of G .*

Proof. Apply the Transfer Principle 3.2.1 to Theorem 3.3.2; indeed, the continuity and weak expansiveness of motion φ_t may be described in the first-order language of ordered fields, so the Transfer Principle applies.

Let G' be an abstract linkage over $\mathbb{R}\langle\varepsilon\rangle$ with the same unweighted graph as G but with edge lengths $\ell_{G'}$ matching those of configuration C_0 . The preceding paragraph guarantees the existence of a weakly expansive motion $\Phi_t : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \text{NConf}(G') \subseteq \text{NConf}_{\sim}(G)$ from C_0 to a straightened or convexified configuration that preserves the edge lengths of G' exactly. Because $\Phi_t(v)$ remains fixed at $(0, 0)$ and G' is connected with bounded edge lengths, Φ_t remains bounded over \mathbb{R} for all t . This Φ_t is thus the desired motion. \square

3.4 Gradual Functions

As a first step toward relating our infinitesimal model to prior models, we must show how to convert a continuous, semialgebraic path $\Phi : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ to a corresponding path over the real numbers, by “forgetting” infinitesimals. The simplest attempt to define this real path $\varphi : [0, 1]_{\mathbb{R}} \rightarrow \mathbb{R}^n$ from Φ would be to declare, for each $t \in [0, 1]_{\mathbb{R}}$, that $\varphi(t) := \lim_{\varepsilon} \Phi(t)$.

Unfortunately, this does not always produce a continuous function. For example, if $\Phi : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \mathbb{R}\langle\varepsilon\rangle$ is the continuous, semialgebraic function defined as $\Phi(T) = T/\varepsilon$ when $0 \leq T \leq \varepsilon$ and $\Phi(T) := 1$ otherwise, the corresponding φ is discontinuous: $\varphi(0) = 0$, while $\varphi(t) = 1$ when $t > 0_{\mathbb{R}}$. The problem is that this Φ travels “infinitely fast”: the times 0 and ε are indistinguishable under \lim_{ε} , but the function values at those moments, $\Phi(0) = 0$ and $\Phi(\varepsilon) = 1$, are distinguishable. Functions that avoid such infinite speeds we call *gradual* (Definition 3.4.3), and in this section, we show how any continuous, semialgebraic path over $\mathbb{R}\langle\varepsilon\rangle$ may be “slowed down” or reparameterized to be gradual, allowing the above naïve strategy to work. We also analyze gradual functions more general than paths, which will prove useful when applying these ideas to origami in Chapter 4.

For the remainder of this section, let R be any real-closed field.

Lemma 3.4.1 (Monotonic Subdivision). *Let $f : [0, 1]_R \rightarrow R$ be any continuous semialgebraic function. Then there is a finite subdivision $[0, 1] = \bigcup_{i=0}^{k-1} [t_i, t_{i+1}]$ where $0 = t_0 < t_1 < \dots < t_k = 1$ such that on each interval $[t_i, t_{i+1}]$, f either strictly increases, strictly decreases, or is constant.*

Proof. Define

$$I_+ = \{t \in [0, 1] \mid \forall e \in (0, t) \exists s \in (t - e, t) \text{ s.t. } f(s) \leq f(t)\},$$

$$I_- = \{t \in [0, 1] \mid \forall e \in (0, t) \exists s \in (t - e, t) \text{ s.t. } f(s) \geq f(t)\}.$$

It may be checked that $I_+ \cup I_- = [0, 1]_R$ —this relies on neither continuity nor semialgebraicity. Because f is semialgebraic, the sets I_+ and I_- are also semialgebraic and therefore consist of finitely many points and intervals.

We first show that f is weakly increasing on (each interval contained in) I_+ . So suppose $a < b \in I_+$ and the whole interval $[a, b]$ is contained in I_+ ; we'll show $f(a) \leq f(b)$. Letting $m = \min_{t \in [a, b]} f(t)$ be the least value f attains on $[a, b]$, it suffices to show that $f(a) = m$. The set $M = f^{-1}(m) \cap [a, b]$ is a nonempty, closed, semialgebraic subset of $[a, b]$ and therefore consists of finitely many points and intervals. We claim that M in fact takes the form $M = [a, c]$ for some $a \leq c \leq b$ (where $[a, a] = \{a\}$). Indeed, if this is not the case, then M has some isolated point s different from a and/or a component $[s, t]$ not containing a , and in either situation, $s \in M$ but $(s - e, s)$ is disjoint from M for small enough e . But then f on interval $(s - e, s)$ takes values strictly greater than $m = f(s)$, contradicting the fact that $s \in I_+$. Therefore M indeed has the claimed form, so $a \in M$ and $f(a) = m$, as required.

So f indeed weakly increases on I_+ , and by symmetric arguments, f weakly decreases on I_- . Then f must be constant on (each component of) $F_0 := I_+ \cap I_-$, strictly increasing on $F_+ := I_+ \setminus F_0$, and strictly decreasing on $F_- := I_- \setminus F_0$, and note that these three semialgebraic sets are disjoint but have $F_- \cup F_0 \cup F_+ = [0, 1]$. By continuity, f is strictly decreasing, constant, and strictly increasing on $\overline{F_-}$, $\overline{F_0}$, and $\overline{F_+}$, respectively, and these three sets are interior disjoint. The result follows by letting t_0, \dots, t_k be the sorted list of the interval endpoints defining these three semialgebraic sets. \square

Lemma 3.4.2. 1. Any semialgebraic set $S \subseteq R\langle \varepsilon \rangle_b^n$ is actually contained in $[-m, m]^n$ for some $m \in R$. (In particular, $R\langle \varepsilon \rangle_b^n$ is not a semialgebraic subset of $R\langle \varepsilon \rangle^n$.)

2. If a semialgebraic set $S \subseteq R\langle \varepsilon \rangle$ contains no infinitesimal values, it is disjoint from $[-r, r]$ for some positive $r \in R$.

Proof. We first prove part 1 in the case $n = 1$. Then the semialgebraic set $S \subset \mathbb{R}\langle \varepsilon \rangle$ consists of finitely many intervals (and isolated points, which we consider degenerate intervals here). Because S contains no unbounded values, none of the endpoints of these intervals can be unbounded, so S is indeed contained in $[-m, m]$, where $m \in R$ is a bound on the magnitude of these (finitely many) interval endpoints.

For general n , it is sufficient to show that the projection of S onto each coordinate axis is contained in $[-m_i, m_i]$ for some $m_i \in R$. But this projection is semialgebraic and bounded over R , so this follows by the $n = 1$ case above.

To prove part 2, apply the first part to the semialgebraic set $\{\frac{1}{x} \mid x \in S\}$. \square

Definition 3.4.3 (Gradual Function). For a semialgebraic set $S \subset \mathbb{R}\langle \varepsilon \rangle^n$ and a semialgebraic function $\Phi : S \rightarrow \mathbb{R}\langle \varepsilon \rangle^m$, say that Φ is **gradual** if $\Phi(a) \sim \Phi(b)$ whenever $a \sim b$.

Note. Gradual functions need not be continuous, such as the piecewise-constant function $\Phi : \mathbb{R}\langle \varepsilon \rangle \rightarrow \mathbb{R}\langle \varepsilon \rangle$ defined by $\Phi(T) := 0$ when $T \leq 0$ and $\Phi(T) := \varepsilon$ when $T > 0$.

Theorem 3.4.4 (Gradual Projections are Continuous). Consider a semialgebraic set $S \subset R\langle \varepsilon \rangle^n$ and a semialgebraic function $\Phi : S \rightarrow R\langle \varepsilon \rangle^m$ such that $\Phi(S \cap R\langle \varepsilon \rangle_b^n) \subseteq R\langle \varepsilon \rangle_b^m$. If

Φ is gradual on $S \cap R\langle\varepsilon\rangle_b^n$ (but not necessarily continuous), the function $\varphi : \lim_\varepsilon S \rightarrow R^m$ defined by sending $p \in \lim_\varepsilon S$ to $\varphi(p) := \lim_\varepsilon \Phi(p)$ is semialgebraic and continuous.

Proof. Consider the graph $\Gamma = \{(T, \Phi(T)) \mid T \in S\} \subset R\langle\varepsilon\rangle^{n+m}$ of Φ , and let $\gamma = \lim_\varepsilon \Gamma$, which is semialgebraic and closed. This set γ contains the graph of φ because it contains $\lim_\varepsilon(p, \Phi(p))$ for each $p \in \lim_\varepsilon S$. We'll show that γ equals the graph of φ . Indeed, for any $p \in \lim_\varepsilon S$, suppose there were some point $(p, y) \in \gamma$ different from $(p, \varphi(p))$. Then there is some $(P, \Phi(P)) \in \Gamma$ with $\lim_\varepsilon P = p$ and $\lim_\varepsilon \Phi(P) \neq \varphi(p)$, but this exactly says that $P - p$ is infinitesimal while $\Phi(P) - \Phi(p)$ is not, contradicting the assumption that Φ is gradual on $S \cap R\langle\varepsilon\rangle_b^n$; note that P is bounded because $P \sim p$. So $(p, \varphi(p))$ is the only point in γ with first coordinate p . This shows that γ is indeed the graph of φ , and in particular, φ is a semialgebraic function.

To show that φ is continuous, suppose the contrary at some point $p \in \lim_\varepsilon S$. By the definition of continuity, there is some positive $d \in R$ such, for any positive $r \in R$, there is some point $q \in \lim_\varepsilon S$ with $|q - p| < r$ but $|\varphi(q) - \varphi(p)| > d$. The semialgebraic set $T := \{|Q - p| \mid Q \in S, |\Phi(Q) - \Phi(p)| > d\} \subset R\langle\varepsilon\rangle$ therefore intersects the interval $(0, r]$ for every positive $r \in R$, so by Lemma 3.4.2 part 2, T contains some infinitesimal value. But then there must be some $Q \in S$ such that $Q \sim p$ (so Q is bounded) but $|\Phi(Q) - \Phi(p)| \approx 0$, contradicting the assumption that Φ is gradual on the bounded portion of S . So φ is continuous, as claimed. \square

Theorem 3.4.5 (Gradual Reparametrization). *Let $I = [I_{\min}, I_{\max}] \subset R\langle\varepsilon\rangle$ be a closed interval, and let $\Phi : I \rightarrow R\langle\varepsilon\rangle_b^n$ be a continuous semialgebraic path. Then there exists a continuous, semialgebraic, increasing reparameterization $R : [0, 1]_{R\langle\varepsilon\rangle} \rightarrow I$ with $R(0) = I_{\min}$ and $R(1) = I_{\max}$ such that $\Phi \circ R$ is gradual.*

Proof. By Lemma 3.4.1 (and changing signs as necessary), it suffices to prove the result when each coordinate function Φ_i is either strictly increasing on I or is constant on I . By translation we may assume $\Phi(I_{\min}) = \vec{0}$. The case where Φ_1, \dots, Φ_n are all constant on I is trivial, so we may assume at least one of them strictly increases; in particular, the function $F(T) := |\Phi(T)|^2$ is continuous and strictly increasing from $0 = F(I_{\min})$ to $M := F(I_{\max}) \in R\langle\varepsilon\rangle_b$.

We choose to reparameterize Φ such that $|\Phi(T)|^2$ increases at a constant rate: specifically, we define our chosen reparameterization as $R(U) := F^{-1}(U \cdot M)$ for $U \in [0, 1]_{R\langle\varepsilon\rangle}$. Because F strictly increases, R is continuous.

We'll verify that $|\Phi(R(V)) - \Phi(R(U))|^2 \leq M \cdot (V - U)$ for any $0 \leq U \leq V \leq 1$; because $M \cdot (V - U)$ is infinitesimal whenever $V - U$ is infinitesimal (as $M \in R\langle\varepsilon\rangle_b$), this suffices to show that $\Phi \circ R$ is gradual. To this end, write $A := \Phi(R(U))$ and $B := \Phi(R(V))$, so that $|A|^2 = UM$ and $|B|^2 = VM$, and furthermore, the vectors A , B , and $B - A$ have non-negative entries. Then $A \cdot B \geq A \cdot A = |A|^2 = UM$, so

$$|B - A|^2 = |A|^2 + |B|^2 - 2A \cdot B \leq UM + VM - 2UM = (V - U) \cdot M,$$

as claimed. \square

3.5 Translating to the Combinatorial and Limiting Models

In this section we demonstrate that configurations and motions in the infinitesimal linkage model are also valid according to the combinatorial and limiting models. These models are

known to be equivalent [1], so directly relating the infinitesimal model to *both* models is unnecessary from a correctness perspective. We do so anyway to concretely compare the strengths and weaknesses of these models. In the process, we locate and fix a minor error in the original combinatorial model.

3.5.1 From Infinitesimal to Combinatorial

We first review the combinatorial model of self-touching linkages from [33] more thoroughly. As described in Section 3.1.3, a configuration has the form (c, M) , where $c : V(G) \rightarrow \mathbb{R}^2$ specifies the geometry and the **magnified views** M_p combinatorially describe the stacking order in the form of a planar graph within a circle for each point $p \in \mathbb{R}^2$ in the image of c . It remains to specify when a motion, $t \in [0, 1]_{\mathbb{R}} \mapsto (c_t, M_t)$, is considered valid by this model. We may assume without loss of generality that all of G 's edge lengths are 1 or greater. There are three forbidden interactions that a valid motion must avoid:

Vertex-Vertex Constraint In a combinatorial configuration (c, M) , pick a vertex u and an edge (v, w) . Node u is said to lie **to the left of edge (v, w)** if either

$$\det(c(u), c(v), c(w)) > 0_{\mathbb{R}} \quad (3.2)$$

(so u is *geometrically* to the left), or if this determinant is 0 but $c(u)$ lies in the relative interior of $c(vw)$ and is *combinatorially* assigned to the left according to the magnified view $M_{c(u)}$. Lying **to the right of edge (v, w)** is defined similarly. For convenience, denote condition (3.2) by $L_{VE}(c, u, vw)$.

The Vertex-Edge constraint requires that in a valid motion, if u lies to the left of (v, w) at time r and to the right of this edge at time $t > r$, then at some moment in interval $[r, t]$, u must have touched the line containing segment vw away from the relative interior of this segment. In other words, u cannot switch sides without going *around* edge (v, w) .

Vertex-Chain Constraint For a combinatorial configuration (c, M) , a chain of distinct edges v_1v_2 and v_2v_3 , and a node u where $c(u)$ is at most $1/2$ away from $c(v_2)$, we say that u lies **to the left of chain $v_1v_2v_3$** in c (either geometrically or combinatorially) if, either,

- $c(u)$ does not touch $c(v_1v_2)$ or $c(v_2v_3)$ and it holds that

$$\det(\text{unit}(c(v_3) - c(v_2)), \text{unit}(c(u) - c(v_2)), \text{unit}(c(v_1) - c(u))) > 0, \quad (3.3)$$

- or $c(u)$ lies in the relative interior of edge $c(v_1v_2)$ or $c(v_2v_3)$ and is combinatorially assigned to the left of this edge in $M_{c(u)}$, or
- $c(u) = c(v_2)$ and is combinatorially assigned to the left of this chain in $M_{c(u)}$.

To explain Equation (3.3), note that u lies to the left of chain $v_1v_2v_3$ precisely when the angles of the three vectors $C(v_2v_3)$, $C(v_2u)$, and $C(v_2v_1)$ are sorted in counterclockwise cyclic order (see Figure 3-3), and this is precisely what the given determinant tests. We name this inequality $L_{VC}(c, u, v_1v_2v_3)$. Node u lying **to the right** of the chain is defined analogously.

The Vertex-Chain condition stipulates that during motion (c_t, M_t) , for as long as node u remains at most $1/2$ away from v_2 , u stays on the same side of chain $v_1v_2v_3$. Note that, on its

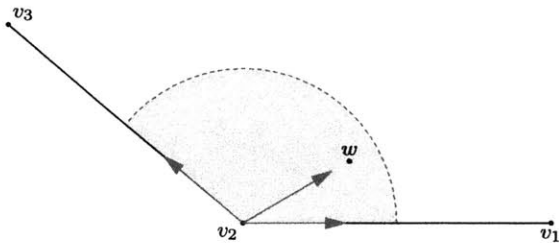


Figure 3-3: For a node u near chain $v_1v_2v_3$, u lies to the left of the chain when the unit vectors in the directions of v_2v_1 , v_2w , and v_2v_3 respectively are arranged in counterclockwise cyclic order.

own, this is more permissive than the Vertex-Chain constraint described in [33]. However, the additional cases permitted here—namely, u passing through both edges of a 360° chain or passing through the single edge in a degenerate chain, as depicted in Figure 3-4—are already forbidden by the Vertex-Edge constraint above, so this modification does not affect the overall set of motions accepted by the model.



(a) Node u passing through a 360° chain $v_1v_2v_3$. (b) Node u passing through a degenerate $v_1v_2v_3$.

Figure 3-4: Two motions allowed by our Vertex-Chain condition but not that in [33]. These motions are forbidden by the Vertex-Edge constraint, however, so this weaker Vertex-Chain condition does not affect the model overall.

Edge-Edge Constraint The Edge-Edge constraint forbids a small fringe-case overlooked by the previous two constraints: an edge $c(u_1u_2)$ passing through another edge $c(v_1v_2)$ when both vertices align, i.e., $c(u_1) = c(v_1)$ and $c(u_2) = c(v_2)$, for otherwise the Vertex-Edge constraint would catch it. More precisely, when a combinatorial configuration (c, M) satisfies

$$|c(u_1) - c(v_1)| \leq \frac{1}{8} \quad \text{and} \quad |c(u_2) - c(v_2)| \leq \frac{1}{8}, \quad (3.4)$$

say that u_1u_2 is positioned **to the left of edge v_1v_2** when (see Figure 3-5):

- segments $c(u_1u_2)$ and $c(v_1v_2)$ are disjoint and the statement

$$(L_{VE}(c, u_1, v_1v_2) \wedge L_{VE}(c, u_2, v_1v_2)) \vee (L_{VE}(c, v_1, u_2u_1) \wedge L_{VE}(c, v_2, u_2u_2)) \quad (3.5)$$

holds, or

- $c(u_1)$ or $c(u_2)$ lies on segment $c(v_1v_2)$, and edge u_1u_2 is to the left of v_1v_2 in the magnified view at that point; or
- $c(v_1)$ or $c(v_2)$ lies on segment $c(u_1u_2)$, and edge u_1u_2 is to the left of v_1v_2 in the magnified view at that point.

The condition (3.5), which we abbreviate $L_{EE}(c, u_1u_2, v_1v_2)$, says that either both of u_1u_2 's endpoints lie to the left of line v_1v_2 , or both of v_1v_2 's endpoints lie to the right of line u_1u_2 ,

as illustrated in Figure 3-5. Edge (u_1, u_2) lying **to the right of edge** (v_1, v_2) is defined analogously.

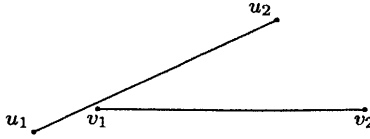


Figure 3-5: Both halves of condition $L_{EE}(c, u_1u_2, v_1v_2)$ need not hold simultaneously.

The Edge-Edge constraint stipulates that, for as long as condition (3.4) remains true during a valid motion, u_1u_2 must remain on the same side of v_1v_2 . Note that this differs from the Edge-Edge Constraint of [33] in a subtle but crucial way: our constraint remains active only when *both* $|c(u_1) - c(v_1)|$ and $|c(u_2) - c(v_2)|$ remain short, whereas [33] only required $|c(u_1) - c(v_1)|$ to stay short. We believe the latter places too strong a restriction on movement: for example, the Edge-Edge constraint of [33] inadvertently forbids any chain from growing from 179° to 181° . Requiring *both* pairs of endpoints to remain nearby fixes this error.

We are now prepared to convert an infinitesimal configuration $C \in \text{NConf}_\sim(G)$ to its corresponding combinatorial configuration $(c, M) = \text{Comb}(C)$. The geometry is determined by C 's geometry after ignoring infinitesimals (i.e., $c = \lim_\varepsilon C$), and the combinatorial magnified views M_p are determined by those same infinitesimals. More concretely, if $c(u)$ lies on the relative interior of bar $c(v, w)$ —in other words, $C(u)$ lies infinitesimally close to segment $C(vw)$ but not to its endpoints—then in the magnified view $M_{c(u)}$ at $c(u)$, we declare that u is combinatorially positioned to the left of edge (v, w) if $L_{VE}(C, u, vw)$ holds over $\mathbb{R}\langle\varepsilon\rangle$, and to the right of (v, w) if $L_{VE}(C, u, vw)$ holds instead. The determinant defining L_{VE} is infinitesimal, but it cannot equal $0_{\mathbb{R}\langle\varepsilon\rangle}$ because C is nontouching.

Similarly, for a node u and a chain of distinct edges v_1v_2 and v_2v_3 , if $c(u) = c(v_2)$, then combinatorially assign u to the left of chain (v_1, v_2, v_3) if $L_{VC}(C, u, v_1v_2v_3)$ holds over $\mathbb{R}\langle\varepsilon\rangle$, and to the right if $L_{VC}(C, u, v_3v_2v_1)$ holds instead. As above, the determinant defining L_{VC} cannot be 0 because C is nontouching.

Finally, for two distinct edges (u_1, u_2) and (v_1, v_2) whose images in c overlap for positive distance (with $c(u_1u_2)$ parallel to $c(v_1v_2)$ and not antiparallel), we assign edge (u_1, u_2) combinatorially to the left of (v_1, v_2) in $(c, M) = \text{Comb}(C)$ if $L_{EE}(C, u_1u_2, v_1v_2)$ holds, and to the right if $L_{EE}(C, v_1v_2, u_1u_2)$ holds instead. As above, precisely one be true because $C(u_1u_2)$ and $C(v_1v_2)$ are disjoint.

These designations suffice to fully specify the magnified views of $(c, M) = \text{Comb}(C)$, as they provide a combinatorial disambiguation for any pair of self-touching features according to c 's geometry. We must verify, however, that these magnified views are well-defined (i.e., the designations above are never self-contradictory) and planar. This may be demonstrated by appealing directly to C 's geometry: For any node u , let $d \in \mathbb{R}$ be the least distance from $C(u)$ to any node or edge in c that is not incident with point $c(u)$ itself, and consider the disk of radius $d/2$ around $C(u)$ in the plane $\mathbb{R}\langle\varepsilon\rangle^2$. This disk intersects precisely the edges that come infinitesimally close to $C(u)$, and it may be (painstakingly) verified that the order in which the disk's boundary intersects these edges corresponds precisely to the order of the nodes along the boundary of planar graph $M_{c(u)}$. Planarity then follows from the fact that C is non-touching: if four nodes on the boundary of the magnified view of $c(u)$ were out of order, their corresponding edges in C would intersect inside the disk by the Intermediate

Value Theorem for continuous semialgebraic functions over arbitrary real-closed fields [17, Prop. 3.4].

We'll show that *motions* can be likewise translated from infinitesimal to combinatorial.

Theorem 3.5.1. *Let C_T , mapping $T \in [0, 1]_{\mathbb{R}\langle \varepsilon \rangle}$ into $\text{NConf}_{\sim}(G)$, be a continuous, semialgebraic, and gradual motion of G in the infinitesimal self-touching linkage model, and define $(c_t, M_t) := \text{Comb}(C_t)$ for each $t \in [0, 1]_{\mathbb{R}}$. Then $t \in [0, 1]_{\mathbb{R}} \mapsto (c_t, M_t)$ defines a valid motion in the combinatorial self-touching linkage model. In particular, if C_T is expansive, then (c_t, M_t) is also expansive.*

Note. Any continuous, semialgebraic path over $\mathbb{R}\langle \varepsilon \rangle$ may be reparameterized by Theorem 3.4.5 to be gradual, so assuming C_T is gradual is not a real restriction.

Proof. As before, we may assume that all of G 's edges have length at least 1. Because C_T is gradual, Lemma 3.4.4 guarantees that motion c_t is continuous. Edge-lengths (over $\mathbb{R}\langle \varepsilon \rangle$) stray from G 's edge lengths by no more than an infinitesimal amount in C_T , so edge-lengths in c_t remain constant. Motion C_T avoids touching entirely, so c_t avoids edges properly crossing in their relative interiors. It remains only to show that (c_t, M_t) avoids the three forbidden movements. We tackle these individually.

Vertex-Edge Constraint. We may check that u lying to the left of (v, w) in configuration $(c_r, M_r) = \text{Comb}(C_r)$ implies $\det(C_r(u), C_r(v), C_r(w)) > 0_{\mathbb{R}\langle \varepsilon \rangle}$, by the way $\text{Comb}(C_r)$ was defined. Similarly, this determinant must be strictly negative at time t , so by the Intermediate Value Theorem [17, Prop. 3.4], there is some $S \in [r, t]_{\mathbb{R}\langle \varepsilon \rangle}$ at which $\det(C_S(u), C_S(v), C_S(w)) = 0_{\mathbb{R}\langle \varepsilon \rangle}$, i.e., $C_S(u)$ lies on line $C_S(vw)$. Since C_S is nontouching, $C_S(u)$ cannot lie on closed segment $C_S(vw)$, so it must lie somewhere on the extension of this line. It follows that at time $s = \lim_{\varepsilon} S$, point $c_s(u)$ lies either on the extension of edge $c_s(vw)$ or at one of its endpoints—here we use the fact that C is gradual, so $C_S(u)$ and $C_s(u)$ are infinitesimally close.

Vertex-Chain Condition. To verify that the motion (c_t, M_t) satisfies this Vertex-Chain condition, observe simply that, when $C(u)$ is strictly closer than 1 unit away from $C(v_2)$, $c(u)$ lies to the left of $c(v_1v_2v_3)$ precisely when

$$\det(\text{unit}(C(v_3) - C(v_2)), \text{unit}(C(u) - C(v_2)), \text{unit}(C(v_1) - C(v_2)))$$

is positive in $\mathbb{R}\langle \varepsilon \rangle$, and $c(u)$ lies to the right when this determinant is negative; it cannot be 0, because C is non-touching over $\mathbb{R}\langle \varepsilon \rangle$. But this determinant changes continuously and cannot equal 0 so long as $C(u)$ stays close to $C(v_2)$, so the side of $v_1v_2v_3$ containing u in motion c_t indeed does not change.

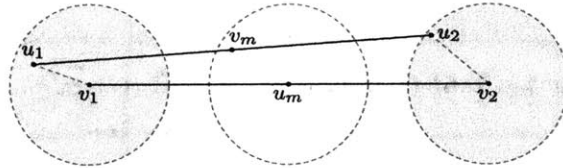


Figure 3-6: When $|u_1v_1|$ and u_2v_2 remain small, midpoint v_m is constrained to remain near the midpoint of edge u_1u_2 .

Edge-Edge Constraint It remains to show that the motion (c_t, M_t) obeys this Edge-Edge constraint. When Equation (3.4) holds for configuration c_t , it may be seen that $u_1 u_2$ is to the left of $v_1 v_2$ in c_t if and only if $L_{EE}(C_t, u_1 u_2, v_1 v_2)$ holds over $\mathbb{R}\langle\varepsilon\rangle$. In turn, this holds if and only if $v_m := (C_t(u_1) + C_t(u_2))/2$, the midpoint of edge $C_t(u_1 u_2)$, lies to the left of $v_1 v_2$, $\det(m, C_t(v_1), C_t(v_2)) > 0$ —this observation uses condition (3.4) to conclude that v_m stays bounded away from either endpoint of $C_t(v_1 v_2)$ (cf. Figure 3-6). Midpoint v_m can never lie on segment $C_t(v_1 v_2)$ because C is nontouching, so $\det(m, C_t(v_1), C_t(v_2))$ must stay positive so long as Equation (3.4) holds, as required. \square

3.5.2 From Infinitesimal To Limiting

For contrast, we now connect our infinitesimal model directly to the limiting model [1]. Observe that the arguments in this section are significantly more streamlined, as they need not consider separate discrete conditions and events.

Recall that a self-touching configuration in the limiting model takes the form (c, A) , where $c \in \text{Conf}(G)$ and $A : E(G) \times E(G) \rightarrow \mathbb{R}$ is a real number “annotation” for each pair of edges that disambiguates layer ordering for overlapping pairs. These annotations are based on the Ord function, introduced in [1], which we now describe. For directed line segments e_1 and e_2 in the plane, intersect e_2 with the left half-plane of e_1 , project this portion onto line e_1 , and let $d_+(e_1, e_2)$ be the length of this projection intersected with segment e_1 . Define $d_-(e_1, e_2)$ analogously with the right closed half-plane. Then the Ord function is given by $\text{Ord}(e_1, e_2) := d_+(e_1, e_2) - d_-(e_1, e_2)$. This function has been carefully crafted to be continuous for all pairs of segments (e_1, e_2) *except* when e_1 and e_2 intersect for positive distance: in this case, if the length of $e_1 \cap e_2$ is d , nearby *interior disjoint* segments e'_1 and e'_2 will have $\text{Ord}(e'_1, e'_2)$ near d (when e'_1 is to the right of e'_2) or $-d$ (when e'_1 is to the left).^{*} By contrast, note that $\text{Ord}(e_1, e_2) = 0$ whenever e_1 and e_2 are collinear. See [1] for details.

Recall from Section 3.3.1 that an r -related configuration $c \in \text{NConf}_r(G)$ is a nontouching configurations whose edge lengths differ from G 's edge lengths by at most r . From any 1-related[†] nontouching configuration $c \in \text{NConf}_1(G)$ we may compute an **annotation** $A_c : E(G) \times E(G) \rightarrow \mathbb{R}$ given by its pairwise Ord values, $A_c(e_1, e_2) := \text{Ord}(c(e_1), c(e_2))$, where we assume G 's edges come with canonical orientations. Then the pair (c, A_c) is called an **annotated** nontouching 1-related configuration, and the set of all such configurations is denoted $\text{Annot}_G(\text{NConf}_1(G))$. Finally, the configuration space $\text{LimitConf}(G)$ of self-touching configurations of G in the limiting model is defined to be the limits of configurations in $\text{Annot}_G(\text{NConf}_1(G))$, i.e.,

$$\text{LimitConf}(G) := \left(\text{Conf}(G) \times \mathbb{R}^{|E(G)| \times |E(G)|} \right) \cap \overline{\text{Annot}_G(\text{NConf}_1(G))}.$$

Continuous self-touching motions are simply continuous paths in $\text{LimitConf}(G)$.

We may now directly relate this to the infinitesimal model, as follows:

Theorem 3.5.2. *For such a configuration $C \in \text{NConf}_\sim(G)$ in the infinitesimal model, define its annotation $A_C : E(G) \times E(G) \rightarrow \mathbb{R}\langle\varepsilon\rangle$ by $A_C(e_1, e_2) := \text{Ord}(C(e_1), C(e_2)) \in \mathbb{R}\langle\varepsilon\rangle$*

^{*}This is meant in a limiting sense: if two sequences of segments e_1^n and e_2^n (for $n \in \mathbb{N}$) converge to e_1 and e_2 respectively and e_1^n and e_2^n are interior disjoint for all n , then the sequence $\text{Ord}(e_1^n, e_2^n)$ has at most two accumulation points, at d and $-d$.

[†]The 1 is not crucial—the same results are obtained if we replace NConf_1 by NConf_r for any $r > 0$ (including ∞), as shown in [1, Lemma 3].

as above, and denote the set of all such (C, A_C) by $\text{Annot}_G(\text{NConf}_\sim(G))$. Then

$$\lim_{\varepsilon} \text{Annot}_G(\text{NConf}_\sim(G)) = \text{LimitConf}(G).$$

Proof. This follows by the transfer principle, Theorem 3.2.2. Indeed, if (c, A) is a configuration in $\text{LimitConf}(G)$, then for any real $\delta > 0$ there is a nontouching annotated configuration (c_δ, A_{c_δ}) at distance at most δ from (c, A) . By Theorem 3.2.2, there therefore must exist a nontouching annotated configuration $(c_\varepsilon, A_{c_\varepsilon})$, defined over $\mathbb{R}\langle\varepsilon\rangle^2$, at distance at most ε from (c, A) . Then $\lim_{\varepsilon}(c_\varepsilon, A_{c_\varepsilon}) = (c, A)$, as claimed. The argument for the reverse inclusion is identical, using the opposite direction of Theorem 3.2.2. \square

Theorem 3.5.3. *If $\Phi : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \text{NConf}_\sim(G)$ is a continuous, gradual, semialgebraic path (i.e., a continuous motion of self-touching configurations of G in the infinitesimal model), then the path $t \in [0, 1]_{\mathbb{R}} \mapsto (\Phi(t), A_{\Phi(t)})$ is a valid motion of self-touching configurations of G in the limiting model.*

Proof. The $\text{Ord}(e_1, e_2)$ function is semialgebraic, continuous, and gradual when restricted to non-intersecting, positive-length segments $e_1, e_2 \in \mathbb{R}\langle\varepsilon\rangle^2$, so this result follows from Theorem 3.4.4. \square

3.5.3 Expansive Self-Touching Carpenter’s Rule Theorem

We have already verified the Expansive Self-Touching Carpenter’s Rule Theorem in the infinitesimal model (Theorem 3.3.3); we may now transfer this to the combinatorial and limiting models as well. Since the latter are equivalent, we may use only the limiting model.

Theorem 3.5.4 (Expansive Self-Touching Carpenter’s Rule Theorem). *For any self-touching configuration (c, A) (in the limiting model) of an open or closed chain linkage G , there is a continuous, expansive motion from (c, A) to the fully straightened configuration (if G is open) or a convex configuration (if G is closed).*

Proof. We simply combine our earlier results. By Theorem 3.5.2 there is some nontouching configuration $C \in \text{NConf}_\sim(G)$ corresponding to $(c, A) \in \text{LimitConf}(G)$. The Expansive Carpenter’s Rule Theorem in the infinitesimal model (Theorem 3.3.3) guarantees the existence of some continuous, semialgebraic, bounded motion Φ_T from C to a straightened or convex position, and by reparameterizing using Theorem 3.4.5, we may assume Φ_T is a gradual function of T . Finally, sampling this motion at real values $t \in [0, 1]_{\mathbb{R}}$ produces a continuous path in $\text{LimitConf}(G)$ starting at $\lim_{\varepsilon}(C, A_C) = (c, A)$ by Theorem 3.5.3, as required. \square

3.6 Linkages with Slender Adornments

In [31], Connelly et al. discuss the natural problem of linkages with regions or *adornments* affixed to the edges of the linkage. They prove the elegant fact that, if a nontouching linkage has “slender” adornments (defined precisely below) that initially do not intersect, then an expansive motion of the underlying linkage prevents the adornments from intersecting. In this section, we extend this result to self-touching linkages.

A weaker form of this result was shown in [1], requiring the adornments to be polygonal and strictly slender. We remove both of these restrictions.

To prove that expansive motions do not cause slender adornment overlaps in the self-touching context, it is tempting to lift the concept of adornments into $\mathbb{R}\langle\varepsilon\rangle$ and apply the transfer principle again. Indeed this can likely be made to work for semialgebraic adornments (though not all slender adornments are semialgebraic), or, with extra effort, by approximating general slender adornments with semialgebraic ones (similar to an approach suggested in [1]). Instead, for maximum generality and simplicity, we choose to stay in \mathbb{R} with the limiting linkage model and to echo the proofs in [31] directly. Essentially the same arguments prove to work; we only need to amend the definition to allow self-touching, and to remove the assumption of strict nontouching from the proof. The formal definitions and arguments follow.

3.6.1 Linkage Adornments

For a linkage G , an **adornment** on an edge $(u, v) \in E(G)$ (with length $\ell = \ell_G(u, v)$) is a simply-connected, compact region D together with two boundary points u', v' (with $|u' - v'| = \ell$) where segment $u'v'$ is contained in D and is called the **base**. Region D must be bounded by two **boundary arcs**—one simple path that traces counter-clockwise around the region's boundary from u' to v' , and another simple, clockwise path from u' to v' . (These paths may intersect each other along base $u'v'$, so D itself is not necessarily a Jordan region.) This adornment D moves in tandem with nodes u and v of G : in any configuration (c, A) of G , we additionally draw region $c(D)$, which is D translated and rotated so that u' coincides with $c(u)$ and v' coincides with $c(v)$.

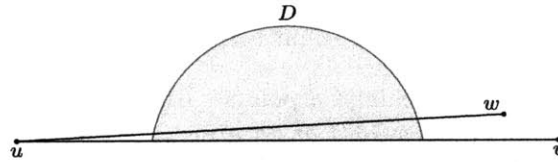


Figure 3-7: An example motivating the modification of our definition of adornment overlap. In the figure, edge uw is touching along the left side of edge uv and therefore does not intersect the interior of adornment D (only its boundary). But D and uw should still be considered overlapping.

In [31], it is declared that two adornments *overlap* if any part of one intersects the *interior* of the other; touching or sliding along common boundaries is allowed. A small adjustment to this definition is needed to adapt to the self-touching setting: indeed, in the example depicted in Figure 3-7, edge uw does not intersect the interior of adornment D , but this configuration should still be considered an overlap.

A simple modification suffices, as follows: Given adornment D and a point p on the relative interior of its base $u'v'$, say that p is **exposed on the left** if the counter-clockwise boundary path from u' to v' passes through p ; otherwise, p is **non-exposed on the left**. Exposure on the right is defined analogously.

Now, for two adornments D_1 and D_2 on edges (u_1, v_1) and (u_2, v_2) respectively, say that D_1 and D_2 **overlap** in a given configuration if:

- (a) D_1 intersects the interior of D_2 , or vice-versa; or
- (b) the base u_1v_1 touches a point p on base u_2v_2 on a side where p is not exposed, or vice-versa.

3.6.2 Slender Adornments

Let D be an adornment as above. As defined in [31], D is called a **slender** adornment if, when a point q traverses either boundary arc from u' to v' , the distances $|q - u'|$ and $|q - v'|$ monotonically increase and decrease, respectively. These boundary arcs need not be differentiable everywhere, but near a point that *is* differentiable, being slender is equivalent to the inward normal intersecting base segment $u'v'$.

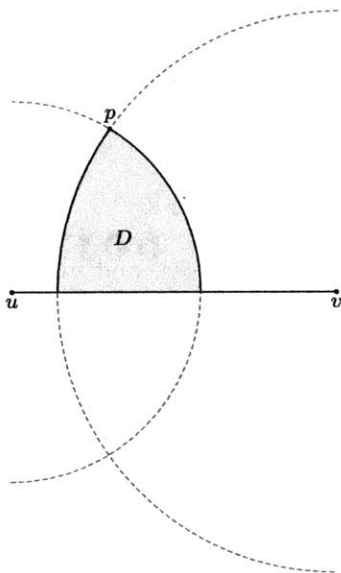


Figure 3-8: Half-lenses are the fundamental examples of slender adornments; $\text{half-lens}(p, uv)$ is shown. Slender adornments are precisely those that are expressible as a union of half-lenses.

Slender adornments may also be understood in terms of fundamental examples called **half-lenses**. Suppose uv is a given base with length ℓ . For a point p not on segment uv with $|pu| \leq \ell$ and $|pv| \leq \ell$, the **half-lens** at p , denoted $\text{half-lens}(p, uv)$, is the intersection of the closed disks through p centered at u and v , restricted to the closed uv half-plane containing p , as in Figure 3-8. Segment uv also belongs to this adornment (as it must). Half-lenses are slender, and in fact, any adornment D is slender if and only if it is a (possibly infinite) union of half-lenses on its base. Said differently, D is slender precisely if $\text{half-lens}(p, uv) \subseteq D$ for each point $p \in D$ away from the base [31, Prop. 3 and 4].

3.6.3 Chains with Slender Adornments Can Always Open

Theorem 3.6.1. *If a configured (combinatorial) self-touching linkage has non-overlapping slender adornments, then any valid expansive motion of the underlying (unadorned) linkage is also free of adornment overlap. In particular, any open chain with non-overlapping slender adornments can always be straightened and is therefore universally configurable without adornment overlap.*

Proof. Suppose we are given an expansive motion φ_t (defined for $t \in [0, 1]$) of a self-touching linkage, as well as adornments that do not overlap at time $t = 0$. For the sake of contradiction, assume that there *is* overlap at some later time.

We first rule out intersections of type (b), i.e., at some time $u > 0$, a point p on base u_1v_1 meets a point q on base u_2v_2 on a side where q is not exposed. Because φ is expansive, the distance between base points $\varphi_t(p)$ and $\varphi_t(q)$ increases monotonically, even when these points are not endpoints of their edges [31]; since this distance is 0 at time $t = u$, p and q must have coincided since $t = 0$. Point p cannot move to the opposite side of base u_2v_2 during this time, so p must have rested on an unexposed side of q since $t = 0$. This contradicts our assumption that φ_0 was overlap-free, so indeed, no type (b) intersections are possible.

Type (a) intersections, where a base or an adornment intersects the interior of another adornment, can be ruled out by the same arguments used in the original non-touching proof [31, Thm. 3, Cases 2 and 3], using the four-circle case of Kirszbraun’s theorem [61][31, Thm. 1]. \square

3.7 Extensions and Future Work

The framework developed here is quite general, and many strengthenings and extensions seem feasible.

Greater Expressibility for Linkages. This model generalizes straightforwardly to linkages in higher dimensions: a self-touching linkage configuration in \mathbb{R}^n is defined simply as a nontouching configuration in $\mathbb{R}(\varepsilon)^n$. This seems to be the first proposed model that expresses linkages in 3-dimensions or higher that may touch but not cross: the limiting model [1] uses the specially-designed Ord function which does not work for co-dimension 2 or greater.

In 2 dimensions and higher dimensions alike, it should be possible to extend the infinitesimal model to allow more flexibility (so to speak) in describing self-touching linkage configurations. For example, allowing “zero-length edges” that are invisible to the geometry has proven useful, such as when constructing locked orthogonal trees as perturbations of rigid self-touching trees [14]. The difficult “ δ -Perturbation Theorem” [71] also becomes much simpler with this allowance [1]. It is known, however, that for the purposes of describing self-touching linkage configurations, these zero-length edges are a convenience only, in that they do not change the linkage’s configuration space [1]. These can easily be incorporated into the infinitesimal model by allowing edges of infinitesimal length, and the stronger analog of Theorems 3.5.2 and 3.5.3 should follow by similar methods.

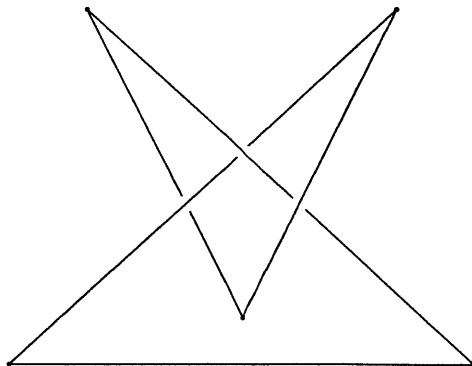


Figure 3-9: A self-touching linkage configuration in \mathbb{R}^3 that has no nontouching perturbations (in either \mathbb{R}^3 or $\mathbb{R}(\varepsilon)^3$) without artificially inserting more vertices.

Similarly, we could allow bars to be subdivided *anywhere*, so long as the newly created

vertex is bent only infinitesimally far and therefore geometrically appears to remain straight. It is likely that this too does not change the resulting configuration space for linkages in the plane, but it *does* in higher dimensions. The self-touching, 5-bar trefoil configuration depicted in Figure 3-9, which lives in ambient 3-dimensional space, has a nontouching perturbation in $\mathbb{R}\langle\varepsilon\rangle^3$ if such subdivision with infinitesimal bending is allowed (by introducing slight kinks near the three overlaps), but no such perturbation exists if all 5 bars must remain straight. Indeed, if a nontouching 5-bar trefoil knot existed in $\mathbb{R}\langle\varepsilon\rangle^3$ then it would exist in \mathbb{R}^3 as well by the Tarski-Seidenberg Principle, but the trefoil knot is known to have stick number 6 [50].

Difficulties with Rigidity. This paper has focused on defining *motions* of self-touching linkages, but there is also interest in detecting *rigidity* of certain linkages [33, 14]. Our infinitesimal model has much intrinsic flexibility—a single combinatorial configuration has many infinitesimal perturbations—and therefore may not be well equipped for working with rigid self-touching linkages. We therefore do not see the infinitesimal model as a global improvement over the combinatorial and limiting models, but rather as a useful tool when working with self-touching linkages *in motion*.

Beyond Linkages: Origami. The idea of using infinitesimals to describe self-touching geometry does not only apply to linkages. We show in Chapter 4 how this idea provides a versatile new framework for origami, by using it to formally analyze the *orderly squash* algorithm for flattening convex polyhedral surfaces and folding origami tessellations.

Chapter 4

Continuously Flattening Polyhedra Using Straight Skeletons

Chapter Summary. We prove that a surprisingly simple algorithm folds the surface of every convex polyhedron, in any dimension, into a flat folding by a continuous motion, while preserving intrinsic distances and avoiding crossings. The flattening respects the straight-skeleton gluing, meaning that points of the polyhedron touched by a common ball inside the polyhedron come into contact in the flat folding, which answers an open question in the book *Geometric Folding Algorithms*. The primary creases in our folding process can be found in nearly quadratic time, though necessarily, creases must roll continuously, and the full crease pattern can be exponential in size in the worst case. We generalize to continuously flattening the facets of a d -dimensional convex polyhedron to a $(d - 1)$ -dimensional hyperplane. As an additional application, we show how a limiting form of our algorithm gives a general design technique for flat origami tessellations, for any spiderweb (planar graph with all-positive equilibrium stress).

We formally analyze this folding algorithm as an *origami* motion, carefully proving that the surface does not “pierce” itself as it flattens. Self-touching origami configurations and motions are notoriously complicated to define and even moreso to reason about, but we propose a novel formal model of origami that aims to reverse this trend. Instead of the traditional approach of combinatorially labelling touching layers as “above” or “below” each other with a slew of local consistency conditions, we appeal to more direct geometric intuition by “infinitesimally” separating the layers to a nearby, nontouching configuration. Our model uses the field $\mathbb{R}(\varepsilon)$ of algebraic Puiseux series, which extends \mathbb{R} to include formal infinitesimals, to aid in expressing these infinitesimal perturbations while also exposing powerful tools like the Tarski-Seidenberg principle.

4.1 Introduction

4.1.1 Flattening Polyhedra with Origami

We consider the problem of flattening the surface of a given polyhedron with an origami motion, meaning the surface should deform continuously, without stretching and without penetrating itself, to a completely flattened state. This is impossible with rigid origami: if finitely many creases are chosen in advance, and the panels between creases must remain flat and rigid, then flexible polyhedra do not change volume as they flex by the Bellows

Theorem [35], and in particular their volume cannot drop to 0 continuously. Even worse, convex polyhedra remain entirely rigid no matter how many extra creases are added [30]. To overcome this, we allow “rolling” creases that continuously move across the surface over time.

Previously, Bern and Hayes [22] used a circle-packing algorithm to show that any genus-0 polyhedron has a flattened state, but no continuous motion to this flattened state is known. Itoh, Nara, and Vilcu [49] proved that convex polyhedra may be continuously flattened; their method “pinches” one vertex at a time, using Alexandrov’s Gluing Theorem [12] to reconfigure the rest of the polyhedron as this happens. The end result is easy to describe in terms of intrinsic shortest paths on the surface, but reliance on Alexandrov’s Theorem renders the intermediate stages of the motion difficult to compute [24, 53].

For convex polyhedra, Demaine and O’Rourke have asked whether there exist flattenings that adhere to the so-called *straight-skeleton gluing* [43, §18.4]: the straight-skeleton gluing identifies two points of the surface whenever they project onto the same point of the polyhedron’s straight skeleton, or equivalently, whenever there is a sphere inside the polyhedron that is tangent at both points. Because the straight skeleton gluing derives from the intrinsic geometry of the polyhedron, such flattenings would be especially simple and natural. As an example, the standard crease pattern for a paper shopping bag follows the bag’s straight skeleton gluing. Neither flattening described above adheres to the straight skeleton gluing.

Our Geometric Results We show that every convex polyhedron may be flattened according to its straight skeleton gluing with a continuous origami motion. The motion we introduce, called *orderly squashing*, is surprisingly simple to describe. First, give each face of the polyhedron a rank from 1 to n arbitrarily. Now, consider the planes H_1, \dots, H_n containing these faces, and imagine moving these planes along their inward normals simultaneously at the same rate. Each point on the surface moves with its plane until it meets and “sticks to” a more important (i.e., lower-ranked) plane. It moves together with this plane until sticking to an even more important plane, and so on. Eventually, the whole surface sticks to plane H_1 and is therefore flat. We verify that this motion indeed satisfies the properties claimed above. Its simple description also makes it easily computable: an implementation in Python was used to generate all of this paper’s figures.

Orderly squashing is quite general: it continuously flattens any convex polytope in any dimension $d \geq 2$ down to a $d - 1$ -dimensional hyperplane, and it also generalizes to some non-convex structures, which we call *positive hyperplane arrangements*.

4.1.2 Theoretical Foundations of Origami: Challenges of Self-Touching

On a more theoretical level, self-touching origami is notoriously difficult to define. Much work has gone into *geometric* conditions, expressing when a configuration does not stretch or shear the paper [36, 47, 37], especially in the context of rigid origami [48, 80, 83, 81] with flat panels and fixed edge positions. But modeling *layer ordering* is significantly more challenging: it is much the same task as modeling self-touching linkages as in Chapter 3, but with the added complications of additional dimensions (when $d > 2$), no promise of piecewise linearity, and creases that may continuously change positions (i.e., “roll”).

Certain special cases are easier to describe, such as flat folded configurations that lie in the plane, and multiple approaches exist for this. As with self-touching linkages, some models of flat-folded origami describe layer ordering *combinatorially* [52], with local consistency checks, and others describe flat origami in terms of *limits* of nearby, nontouching

configurations [22, 41].

The de-facto standard formal model of general (not necessarily flat) origami [40] (elaborated in [43]) falls squarely in the combinatorial camp and is an extension of Justin’s model for flat origami [52]. It bears strong resemblance to the combinatorial self-touching linkage model from [33], discussed at length in Section 3.5.1, but as might be expected, it is significantly more involved. Briefly, for origami of 1-dimensional “paper” in \mathbb{R}^2 , this model makes use of an “order function” $\lambda(p, q) \in \{-1, 1\}$ that specifies, for each pair of collocated points p and q , whether p lies “above” q ($\lambda(p, q) = 1$) or below q ($\lambda(p, q) = -1$), together with local combinatorial conditions (the Antisymmetry, Transitivity, Consistency, and Noncrossing conditions) that express when the order is well-defined and does not cross itself. There are additional time-continuity constraints governing origami motions. Finally, for higher-dimensional origami, this definition is applied *recursively* by dimension in a small sphere surrounding each point.

Unfortunately, this model too proves quite complicated to use in full rigor, and many papers resort to intuitive arguments instead of engaging with the complexities of the formal model [49, 18, 38], just as was done before the general model’s availability [39, 19]. For example, we are not aware of any arguments using [40, 43] that explicitly address the model’s fundamental recursive nature. This is just as well for simple applications, but for complicated algorithms like Lang’s *Treemaker* [65], theoretical results such as Demaine and Tachi’s *Origamizer* [44], and/or higher-dimensional origami scenarios (e.g., [55]), where intuitions can easily break down or mislead, a higher level of certainty in the form of concise yet water-tight arguments is desirable.

Our Model Results By analogy to the infinitesimal self-touching linkage model developed in Chapter 3, we propose an alternative self-touching origami model based on infinitesimals that aims to bridge the gap between intuitive and formal origami. Our model is succinct to define and geometrically intuitive: an origami configuration that may self-touch but not pierce through itself is specified simply as a *nontouching* configuration of the paper over the field $\mathbb{R}\langle\varepsilon\rangle$. The underlying, self-intersecting geometry can be recovered by sending infinitesimals to 0, and the configuration over $\mathbb{R}\langle\varepsilon\rangle$ may mildly deform the paper (for example, to detour around another layer) so long as the distortions disappear once infinitesimals are ignored.

As evidence of the practical usability of this model, we fully, formally analyze orderly squashing in this infinitesimal origami model, as well as an infinitesimally flattened analog of orderly squashing applied to origami tessellations. These proofs remain conceptually simple, though admittedly a bit notation-heavy.

4.1.3 Acknowledgements

Much of this chapter is a re-exposition of much of [8], which is joint work with Erik D. Demaine, Martin L. Demaine, Jin-ichi Itoh, Anna Lubiw, Chie Nara, and Joseph O’Rourke. Diminished orderly squashing (Section 4.6) and the infinitesimal origami model (Section 4.5.1) represent new work with Erik D. Demaine.

4.2 Positive Hyperplane Arrangements

Our algorithm folds a more general class of objects than (surfaces of) convex polygons or polytopes, which we call **positive hyperplane arrangements** and define as follows. Let

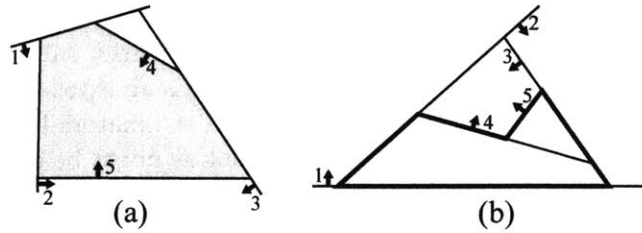


Figure 4-1: Examples of 2D positive hyperplane arrangements: (a) positive cell P_5 is shaded; (b) a nonconvex polygonal subset is shown in bold.

H_1, H_2, \dots, H_n be a sequence of hyperplanes in \mathbb{R}^d . For each $1 \leq i \leq n$ choose a unit vector u_i normal to hyperplane H_i (there are two choices for each), and let the **positive half-space** H_i^+ be the H_i -half-space that u_i points toward. The **positive cell** P_i is defined as the intersection of positive half-spaces $P_i := H_1^+ \cap \dots \cap H_i^+$ (define $P_0 := \mathbb{R}^d$ for convenience), and the **face** F_i is the portion of plane H_i contained in the previous positive half-spaces, i.e., $F_i := H_i \cap P_{i-1}$. Then the union of the faces $F = F_1 \cup \dots \cup F_n$ is a **positive hyperplane arrangement**. Some 2-dimensional examples of positive hyperplane arrangements are shown in Figure 4-1, and a 3-dimensional example is visible in Figure 4-4.

For a convex d -dimensional polytope P , if we choose the hyperplanes H_1, \dots, H_n to be the facets of P (in arbitrary order) with their normals pointing inward, then the union of P 's facets is contained in the resulting positive hyperplane arrangement F , so folding F is enough to fold P . Some nonconvex polytopes can also be expressed as subsets of positive hyperplane arrangements (Figure 4-1(b)), so our orderly squashing algorithm can fold these as well.

We will require our positive hyperplane arrangements to satisfy $u_1 \neq u_2$, so planes H_1 and H_2 cannot be both parallel and similarly oriented (but $u_1 = -u_2$ is acceptable). This is not a meaningful restriction, because when $u_1 = u_2$, the first face $F_1 = H_1$ is disjoint from $F_2 \cup \dots \cup F_n \subseteq P_2 = H_2^+$.

4.3 Orderly Squashing

Our algorithm to flatten any (bounded piece of a) positive hyperplane arrangement $F = F_1 \cup \dots \cup F_n$, which we call **orderly squashing**, is defined as follows: every hyperplane H_i moves in its positive normal direction u_i at unit speed, and when a point of face F_i becomes incident to a lower numbered hyperplane $H_j, j < i$, the point “sticks” to that hyperplane and moves with it instead, until the point becomes incident to an even lower numbered hyperplane, and so on. (When a portion of face F_i sticks to a lower-ordered plane H_j , we say that it **folds onto** H_j .) If $u_1 = u_2$ then none of $F_2 \cup \dots \cup F_n$ will ever fold onto H_1 ; this is the reason for our assumption that $u_1 \neq u_2$. Examples of this motion applied to convex polyhedra are shown in Figures 4-2 and 4-3. Orderly squashing applied to a larger portion of a positive hyperplane arrangement is illustrated in Figure 4-4.

It will be helpful to specify this motion more thoroughly. For $v \in F_i$ and for $1 \leq j < i$, define the collision time $c_{i,j}(v)$ as the time at which v , when moving at unit speed in direction u_i , would hit hyperplane H_j as this plane moves at unit speed in direction u_j . In other words, $c_{i,j}(v)$ is the time at which v would “stick” to plane H_j during orderly squashing, assuming no other planes interfered. By definition of positive hyperplane arrangement, $c_i(v)$ is always

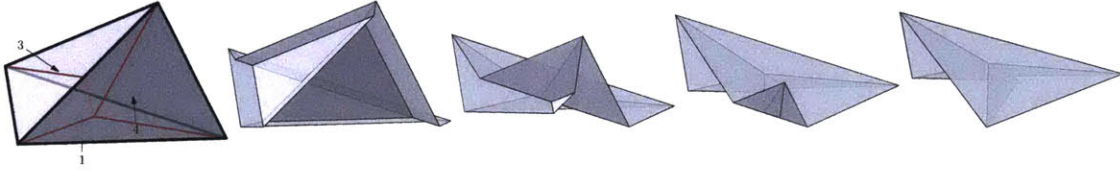


Figure 4-2: Orderly squashing of a tetrahedron. The left pane shows the ordering of faces and the crease pattern, with primary creases in thick dark red and other creases in thin light red. The other panes show an animation of the flattening. Face 2, originally at the back, ends up on top.

nonnegative. When $u_i \neq u_j$, this function $c_{i,j}(v)$ is a linear function of v . If instead $u_i = u_j$ then planes H_i and H_j will never meet, so $c_{i,j}(v) = \infty$. Then during orderly squashing, by definition, v will move forward with plane H_i for distance and time $c_i(v) := \min_{1 \leq j < i} c_{i,j}(v)$ before sticking to a lower-indexed plane. As a notable special case, it may happen that $c_i(v) = 0$, meaning v *already* lies on a lower-ordered plane (or possibly multiple such planes) in addition to H_i . Then the above rule seems ambiguous: should v follow plane H_i or a different plane? Thankfully, there is no ambiguity, as each plane that v lies on will give it the same instructions. Indeed, the H_i rule says that v should move with plane H_i for distance $c_i(v) = 0$, i.e., not at all, before switching allegiance to the lower-indexed plane(s). This ensures that our definition of orderly squashing is well-defined. Let $\text{Sq}_t(v)$ denote the location $v \in F$ has moved to after orderly squashing for time $t \geq 0$.

Take any $v \in F_i$ with $i \geq 2$, and let $t = c_i(v)$ be the time v collides with some lower-order plane H_j . Let v' be the reflected image of v through the bisecting plane between H_i^+ and H_j^+ , and note that by symmetry, $\text{Sq}_t(v) = v + u_i \cdot t = v' + u_j \cdot t$. We would like to show that v and v' meet at time t (the same time v meets plane H_j), but it is conceivable that v' might get swept away by a lower-order plane before time t , i.e., that $c_j(v') < t$. We will show that this cannot happen, so it always holds that $c_j(v') \geq t$ and therefore $\text{Sq}_t(v) = \text{Sq}_t(v')$. To prove this, note that $c_i(v) = t$ implies that the point $\text{Sq}_t(v)$ must lie in front of the moved positions of each H_k , $1 \leq k < i$ at time t , because otherwise $c_{i,k}(v)$ would be smaller than t . Now if $c_{j,k}(v') < t$ for some $1 \leq k < j$, the point $v' + u_j \cdot t = h_t(v)$ would lie *behind* the moved position of plane H_k at time t , contradicting our earlier deduction. So $c_{j,k}(v') \geq t$ for each $1 \leq k < j$, meaning $c_j(v') \geq t$ and $\text{Sq}_t(v) = \text{Sq}_t(v')$ as claimed. Note that v and v' move in unison from time t onward.

Fix some time $T \geq 0$. Armed with the above foundation, we may describe the crease pattern resulting from the orderly squash Sq_T . First, we decompose each face of F into regions according to which lower-indexed plane each point *first* folds onto. Specifically, for any pair $1 \leq j < i$, let $F_i[j] \subseteq F_i$ be the set of points $v \in F_i$ that fold onto plane H_j before encountering any other plane, i.e., $c_i(v) = c_{i,j}(v)$ (v hits H_j first, but ties are allowed) and $c_i(v) \leq T$ (this collision happens during time $[0, T]$). Similarly, let $F_i[0]$ be the set of points that *do not* encounter another plane before time T , i.e., $c_i(v) \geq T$. (When the chosen time T is not clear from context, we will write $F_i^T[j]$.) These regions $F_i[j]$ for $0 \leq j < i$ exhaust F_i , and because they are defined by closed linear inequalities, they are closed, convex, and polygonal (but possibly unbounded, degenerate, or even empty). Furthermore, they may be checked to be interior-disjoint (their pairwise intersections are $n - 2$ -dimensional or lower), but they do share boundary: for example, all points along the seam $F_1 \cap F_2$ belong to $F_2[1]$ as well as $F_1[0]$. Because these regions $F_i[j]$ describe the *first* folding each point of F experiences, we refer to their boundaries as the **primary crease pattern**.

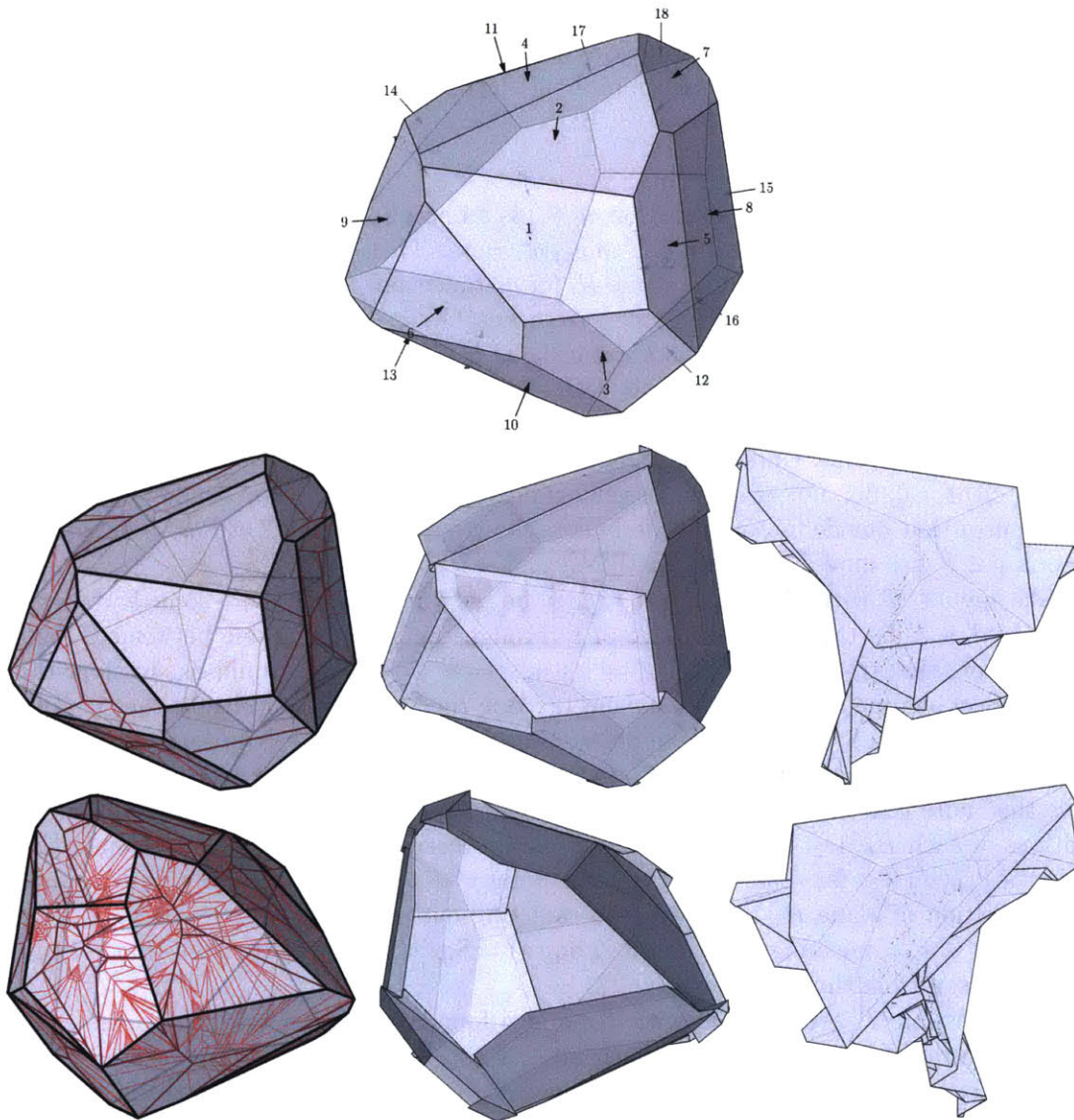


Figure 4-3: Orderly squashing of a convex polyhedron: (top) the ordering of front faces; (middle row) front view of the creases and two stages of the flattening; (bottom row) back view of the creases and two stages of the flattening. Primary creases are shown in thick dark red and other creases in thin light red.

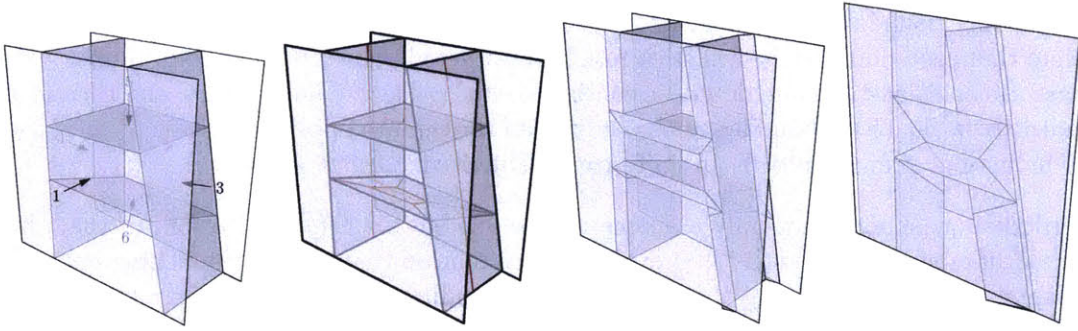


Figure 4-4: Orderly squashing of a 3D positive hyperplane arrangement: (from left to right) the ordering of hyperplanes; the crease pattern, with primary creases in thick dark red and others in thin light red; two frames of the flattening.

The **total** crease pattern may be computed inductively from the primary crease pattern, as follows: we showed above that region $F_i[j]$ folds onto its mirror image in F_j during orderly squashing, so the total crease pattern inside $F_i[j]$ is simply reflections of the creases from the mirrored subset of F_j . A more careful description is possible along the lines of the above treatment of primary creases: instead of partitioning F based on the *first* collision, we may partition based on the *complete sequence of collisions* encountered. Indeed, we inductively define the regions $F_i[j_1, j_2, \dots, j_k, 0] \subset F_i$, where $i > j_1 > j_2 > \dots > j_k > 0$, as follows: a point $v \in F_i$ lies in $F_i[j_1, \dots, j_k, 0]$ if its reflection onto plane H_{j_1} lies in region $F_{j_1}[j_2, \dots, j_k, 0]$. This region represents the points $v \in F_i$ that collide with all of the planes $H_{j_1}, H_{j_2}, \dots, H_{j_k}$ and no others, with some allowances for ties.* As before, these are closed, convex, polygonal regions, and it follows by induction that they are also interior disjoint. Also as before, we'll write $F_i^T[j_1, \dots, j_k, 0]$ when T is not clear from context.

We may now easily verify some basic geometric properties of orderly squashing. The condition that origami paper may not “stretch” may be expressed by requiring an origami configuration to be described by a **rigid map**, which is a piecewise C^1 function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ whose gradient, where defined, is an $m \times n$ orthogonal matrix [36].

Theorem 4.3.1. *Let F be a positive hyperplane arrangement $F = F_1 \cup \dots \cup F_n \subset \mathbb{R}^d$ as above, with $u_1 \neq u_2$. Then orderly squashing $\text{Sq}_t(v)$ deforms F continuously (as a function of both t and v) and does not stretch surface F , i.e., for each $t \geq 0$, Sq_t is a **rigid map** on each face of F . Furthermore, if $S \subset F$ is any bounded subset of F , orderly squashing folds S onto H_1 in finite time.*

Proof. Membership in each region $F_i^t[j_1, \dots, j_k, 0]$ is a closed, linear condition on both $v \in F$ and $t \geq 0$, and orderly squashing restricted to each such region is a continuous linear function, so $\text{Sq}_t(v)$ is indeed a continuous function of v and t . As shown above, for each fixed $t \geq 0$, the function Sq_t sends $F_i^t[j_1, \dots, j_k, 0]$ to its image after reflecting from H_i onto H_{j_1} , then onto H_{j_2} , and so on onto H_{j_k} . In particular, Sq_t is a linear isometry on each region $F_i^t[j_1, \dots, j_k, 0]$, so Sq_t is a piecewise isometry on F and is therefore rigid.

Finally, we'll show by induction on n (the number of planes in F) that bounded subsets $S \subset F$ fold onto H_1 in finite time. In the base case $n = 1$, there is nothing to prove. Let

*In detail, v must collide with all of the listed planes, but it is not required to follow them for positive time. On the other hand, v may collide with planes that are *not* listed as well, but *must* not follow them for positive time.

$S \subset F$ be any bounded subset, and let $S_n = S \cap F_n$ be the portion of S on the n^{th} face of F . Note that each point on S_n will stick to a lower-ordered plane by some finite time $u \geq 0$, because the collision-time function c_n is continuous and piecewise linear on S_n and therefore bounded. Now $\text{Sq}_u(S)$ is bounded and contained in the updated positions of $F_1 \cup \dots \cup F_{n-1}$, so by induction, it folds onto H_1 in finite time, as desired. \square

Orderly squashing is not only a geometric motion but an *origami* motion, in that the surface F does not “pierce itself” as it evolves. Even moreso than with self-touching linkages in Chapter 3, formally defining these non-piercing conditions for self-touching origami is a nontrivial task, as is verifying that orderly squashing indeed satisfies them. We therefore postpone this task to Section 4.5, after discussing the algorithmic properties of orderly squashing’s geometry.

4.4 Computing Orderly Squashing

In this section we outline a simple algorithm to compute orderly squashing, as well as bounds on the complexity of the results.

4.4.1 The Algorithm

We wish to compute the results of orderly squashing for a time of $T \geq 0$. The main algorithmic task is computing the primary crease pattern, i.e., the regions $F_i[j]$. From these, the total crease pattern, i.e., the regions $F_i[j_1, \dots, j_k, 0]$, may be computed by repeated reflections and intersections, as described in Section 4.3.

Theorem 4.4.1. *For a d -dimensional positive hyperplane arrangement $F = F_1 \cup \dots \cup F_n$, the primary crease pattern that results from orderly squashing F for a given time T (the limiting case $T = \infty$ is also valid) has complexity $O(n^{\lfloor (d+2)/2 \rfloor})$, and it may be computed in time $O(n^{\lfloor (d+2)/2 \rfloor} + n^2 \log n)$. (The $n^2 \log n$ term is only relevant for $d = 2, 3$.)*

Proof. We run the following process on each $1 \leq i \leq n$ individually, so consider i fixed. Without loss of generality we may assume that F_i lies in the plane $\mathbb{R}^{d-1} \times \{0\}$ with normal vector $u_i = (0, \dots, 0, 1)$.

The regions $F_i[j]$, $0 \leq j < i$ are determined by which of the collision functions $c_{i,j}$ is smallest, so for each $1 \leq j < i$ let $C_{i,j}$ be the graph of linear function $c_{i,j} : F_i \rightarrow \mathbb{R}$ in the space $F_i \times \mathbb{R} \subset \mathbb{R}^d$. Likewise, let $C_{i,0}$ be the graph of the constant function $c_{i,0} \equiv T$. Then the graph C_i of function c_i is simply the lower hull of the planes $C_{i,0}, \dots, C_{i,i-1}$, and each region $F_i[j]$ is the subset of F_i on which $c_i = c_{i,j}$. In other words, if A_j is the subset of plane $C_{i,j}$ lying on lower hull C_i , then the projection of A_j onto F_i is precisely the region $F_i[j]$. So computing $F_i[0], \dots, F_i[i-1]$ reduces to computing the intersection of $i+1$ half-spaces in \mathbb{R}^d —the lower half-spaces of planes $C_{i,0}, \dots, C_{i,i-1}$ above, as well as the half-space above F_i itself. This half-space intersection problem is equivalent—in fact, projectively dual—to that of computing the convex hull of $i+1$ points in \mathbb{R}^d , so the result has complexity $O(n^{\lfloor d/2 \rfloor})$ and can be accomplished in time $O(n^{\lfloor d/2 \rfloor} + n \log n)$ using Chazelle’s convex hull algorithm [29].

Since we repeat the above procedure for each $i = 1, \dots, n$, the complexity and runtime bounds at most multiply by n , matching those bounds claimed in Theorem 4.4.1. \square

We note that, in the $T = \infty$ case, lower hull C_i lies on the **medial axis** of positive cell P_i , also known as its **straight skeleton** [11, 10, 16] since P_i is convex. This follows

from the observation that the points between facet F_i and lower hull C_i are precisely the points in positive cell P_i that are closer to plane H_i than to any other plane H_1, \dots, H_{i-1} . The algorithm presented above is closely related to a technique that uses “roofs” to compute straight skeletons of convex polytopes [10]. But computing the full straight skeleton of P_i is wasteful in this case, because only the region adjacent to F_i is needed.

Unlike the primary crease pattern, there are not polynomial bounds on the size of the overall crease pattern.

Theorem 4.4.2. *For a positive hyperplane arrangement $F = F_1 \cup \dots \cup F_n$ in \mathbb{R}^d , the number of $d - 1$ -dimensional regions in the total crease pattern is $O(2^n)$.*

Proof. In Theorem 4.3.1, we identified these regions as $F_i[j_1, \dots, j_k, 0]$, each a convex (and therefore connected) polytopal region labelled by a decreasing subsequence (i, j_1, \dots, j_k) of $(n, n - 1, \dots, 1)$. There are at most 2^n such regions. \square

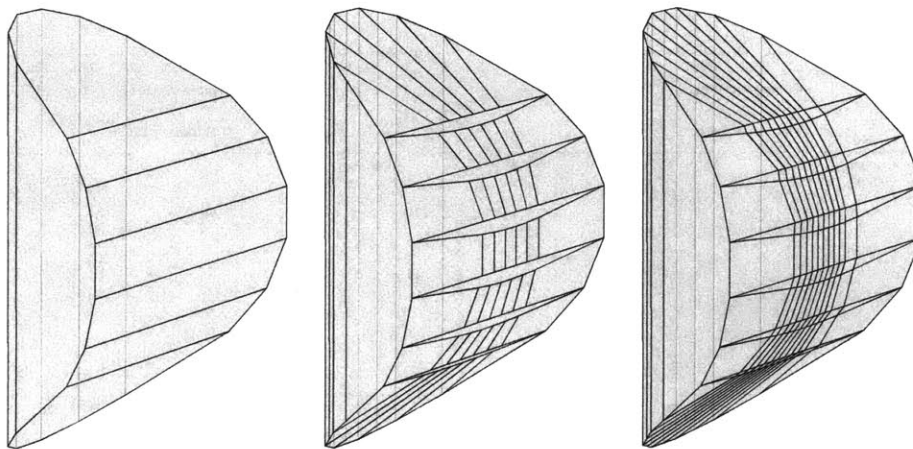


Figure 4-5: A convex polyhedron where any face ordering produces $\Theta(n^2)$ primary creases: (left) the polyhedron; (middle) the primary creases for one face ordering; (right) all creases for that face ordering.

Note. The bounds on the complexity of the primary crease pattern and total crease pattern in Theorems 4.4.1 and 4.4.2 are tight in the worst case. A simple example of a polyhedron in dimension $d = 3$ whose primary crease structure has size $\Theta(n^2)$, no matter what order is given to the faces, is shown in Figure 4-5. It is based on an example from [46, 16]. Matching lower bound examples for the other cases may be found in [8]; the author of this thesis was not closely involved with these examples from [8], so we have chosen to omit them here.

4.4.2 An Implementation

We implemented the above algorithm for $d = 3$ dimensions, which allowed us to automatically generate all of the 3-dimensional images in this Chapter. Our implementation calls Qhull [15] as a subroutine for the half-plane intersection step, but it prioritizes convenience over optimization, invoking Qhull more times than necessary and using quadratic algorithms where faster but more complicated methods exist. Even so, it remains practical on decently-sized examples: on a laptop featuring a 2.90GHz Intel Core i7-3520M processor and 4GB RAM, each three dimensional image in this paper was computed in no more than 15 seconds,

and we computed the 642,295 faces in the overall crease pattern of a randomly generated 3D hyperplane arrangement of 64 planes in just over 70 minutes.

For more careful benchmarking, we used a desktop featuring a 3.5GHz Intel Core i7-3970X processor and 64GB RAM to test two datasets. The first dataset generates n planes with normal vectors chosen uniformly on the sphere, and offsets chosen to guarantee the correct ordering. The second dataset is our example with $\Theta(n^2)$ primary creases from Figure 4-5. The results are shown in Figure 4-6.

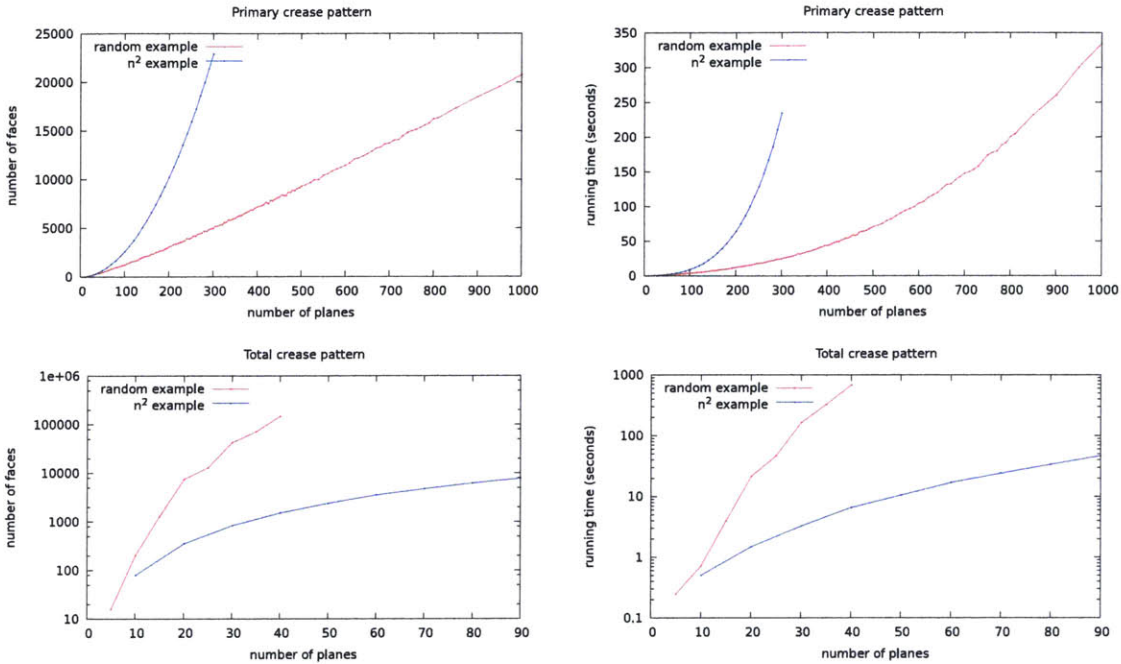


Figure 4-6: Combinatorial complexity of crease patterns (left column) and performance of our implementation (right column) for the primary crease pattern (top row) and full crease pattern (bottom row).

4.5 Challenges of Layer Ordering

We have carefully described the geometry of $Sq_t(v)$ in Section 4.3, but to fully specify orderly squashing as an *origami* motion, we must also specify the *layer ordering* information to clarify the ordering of pieces of paper stacked (geometrically) on top of each other, and to ensure no part of the paper pierces through another part.

4.5.1 Origami via Infinitesimals

By analogy to linkage work from Chapter 3, we propose an alternative origami model based on nontouching configurations over $\mathbb{R}\langle\varepsilon\rangle$.

Definition 4.5.1. Let our sheet of paper $P \subset \mathbb{R}^n$ be an arbitrary semialgebraic set—it need not be connected or have consistent local dimension. Define $P' := \text{Ext}(P, \mathbb{R}\langle\varepsilon\rangle)$ as the lift of P into $\mathbb{R}\langle\varepsilon\rangle^n$, and let $P'_b = P' \cap \mathbb{R}\langle\varepsilon\rangle_b^n$ be the bounded portion of P' . We define

a (possibly self-touching) **origami configuration** of P as a semialgebraic map $C : P' \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ satisfying these properties:

1. Paper stays bounded: C maps P'_b to $\mathbb{R}\langle\varepsilon\rangle_b^n$.
2. Geometry is consistent: C is gradual and continuous on P'_b .
3. No piercing: C is injective on P'_b .
4. No stretching: $\lim_\varepsilon C : P \rightarrow \mathbb{R}^n$ is rigid.

An **origami motion** C_T is likewise defined as a semi-algebraic function $[0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \times P' \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ satisfying the following conditions:

5. Valid at each moment: for each $T \in [0, 1]_{\mathbb{R}\langle\varepsilon\rangle}$, the function $C_T : P' \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ is an origami configuration of S , and
6. Consistent across time: the whole motion C_T is continuous and gradual as a function of $P'_b \times [0, 1]_{\mathbb{R}\langle\varepsilon\rangle}$.

Note. For a semi-algebraic function $C : P' \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ satisfying conditions 1–3 above, the function $\lim_\varepsilon C : P \rightarrow \mathbb{R}^n$ is well-defined and continuous by Theorem 3.4.4. For a family C_T satisfying 5 and 6, the function $\lim_\varepsilon C_T : P \times [0, 1]_{\mathbb{R}} \rightarrow \mathbb{R}^n$ is likewise well-defined and continuous for the same reason.

Note. To prove that a function $f : S_1 \times S_2 \rightarrow \mathbb{R}\langle\varepsilon\rangle^d$ is gradual, it is enough to show that $f(s_1, s_2)$ is a gradual function of $s_1 \in S_1$ when s_2 is held constant, and $f(s_1, s_2)$ is likewise a gradual function of $s_2 \in S_2$ when s_1 is held constant. Indeed, if these are satisfied, then $(s_1, s_2) \sim (t_1, t_2)$ implies

$$f(s_1, s_2) \sim f(t_1, s_2) \sim f(t_1, t_2),$$

meaning f is gradual, as claimed.

We also propose a definition for when two points of paper are “adjacent” in a folded origami state *with no intervening layers*. For example, if a sheet of paper is accordion-folded into thirds, collocated points in the top and middle layers are adjacent, but those in the top and bottom layers are not, because the middle layer is in the way.

Definition 4.5.2. For an origami configuration C of a semi-algebraic set $P \in \mathbb{R}^n$, and for two points $p, q \in P$ that are collocated, i.e., $C(p) \sim C(q) \in \mathbb{R}\langle\varepsilon\rangle^n$, we say that p and q are **adjacent** in C if there is a continuous, semi-algebraic path $Q : [0, 1]_{\mathbb{R}\langle\varepsilon\rangle} \rightarrow \mathbb{R}\langle\varepsilon\rangle^n$ such that:

- Q does not intersect the image of C except at its endpoints $Q(0) = p$ and $Q(1) = q$, and
- the whole path Q is infinitesimally close to $C(p)$ and $C(q)$.

To prove the versatility of these definitions, we will use them to provide formal specifications for two nontrivial origami motions: standard orderly squashing (introduced in Section 4.3 and given origami structure in Section 4.6), and a limiting version of orderly squashing for folding tessellations (Section 4.7).

We make no attempt to connect this definition of origami to that of [40, 43] or any other, but it seems hopeful that they may compare favorably. Our definition is strictly less general, as it can only discuss configurations and motions that are semialgebraic. Perhaps it is possible to generalize this technique without sacrificing usability?

4.6 Diminished Orderly Squashing

We now define **diminished orderly squashing**, a motion that behaves much like standard orderly squashing but completely avoids intersections at the expense of slightly distorting the surface F . In diminished orderly squashing, the surface F moves forward at a speed $s < 1$ that is slightly slower than the unit speed of planes H_i , and standard orderly squashing may be recovered by increasing s all the way to 1. Details follow.

Let $F = F_1 \cup \dots \cup F_n$ be a positive hyperplane arrangement defined by half-spaces H_1^+, \dots, H_n^+ as above. Diminished orderly squashing of F will be defined not only for points on F , but also for the points in \mathbb{R}^d that lie “behind” F : for each i let the **negative half-space** H_i^- be the closed H_i -half-space opposite to H_i^+ , and define the **i^{th} negative cell** by $N_i := H_1^+ \cap \dots \cap H_{i-1}^+ \cap H_i^-$, i.e., the set of points (weakly) behind H_i but (weakly) in front of the earlier planes H_1, \dots, H_{i-1} . These negative cells are closed, convex, polygonal, and interior disjoint, but they do *not* necessarily cover all of \mathbb{R}^d , since points in the interior of positive cell $P_n = H_1^+ \cap \dots \cap H_n^+$ do not lie behind any of the H_i . For convenience, define $N := N_1 \cup \dots \cup N_n = H_1^- \cup \dots \cup H_n^-$.

Diminished orderly squashing, $\text{Sq}_t^s(v)$, is defined for a point $v \in N$, a time $t \geq 0$, and a speed $0 < s \leq 1$, as follows: each plane H_i moves along positive normal u_i at unit speed as before, and each point $v \in N_i$ moves in direction u_i at speed s until hitting a lower-indexed plane $H_j, j < i$, at which time it starts moving in direction u_j at speed s until hitting an even lower-indexed plane, and so on.

Lemma 4.6.1. *For any speed $0 < s \leq 1$, diminished orderly squashing $\text{Sq}_t^s(p)$ is a continuous function of $t \geq 0_{\mathbb{R}}$ and $p \in N$.*

Proof. As with standard orderly squashing, the rule governing points’ trajectories during diminished orderly squashing is consistent for points in multiple negative cells. So by partitioning $N \times [0, \infty)$ into closed, convex regions identified by the sequence of planes hit during t seconds of motion (analogous to the regions $F_i^t[j_1, \dots, j_k, 0]$), we find that diminished orderly squashing is linear and therefore continuous on each such piece. \square

In contrast to standard orderly squashing, diminished orderly squashing at speed $s < 1$ completely avoids collisions:

Lemma 4.6.2. *For any speed $0 < s < 1$ and any time $t \geq 0$, diminished orderly squashing $\text{Sq}_t^s : N \rightarrow \mathbb{R}^d$ is an injective function of N .*

Proof. Suppose Sq_t^s were not injective, so different points $v \neq w$ in N have $\text{Sq}_t^s(v) = \text{Sq}_t^s(w)$ for some time $t \geq 0$. Let $t_0 > 0$ be the *earliest* time at which the images of v and w agree. In which directions were v and w moving leading up to time t_0 ? They cannot both be trailing the same plane H_i , because they would then both be translating along the same velocity vector su_i , meaning they must have met before time t_0 . So say v was trailing H_i and w was trailing H_j leading up to time t_0 , where $j < i$ (without loss of generality). Because w has followed H_j for positive time leading up to time t_0 and because $s < 1$, the point $\text{Sq}_{t_0}^s(w)$ must lie in the interior of H_j^- at this time. But then $\text{Sq}_{t_0}^s(v)$ also lies in the interior of H_j^- , so v must have encountered plane H_j strictly before time t_0 and therefore should be trailing plane H_j (or lower), not the assumed H_i . This contradiction proves that $\text{Sq}_t^s(v) = \text{Sq}_t^s(w)$ cannot happen, as claimed. \square

Diminished orderly squashing does slightly stretch the positive hyperplane arrangement $F \subset N$, but we’ll show it does not do so very much: in the limit as s approaches 1, we recover

standard orderly squashing. In light of the infinitesimal view of origami in Definition 4.5.1, we choose to work over $\mathbb{R}\langle\varepsilon\rangle$, showing that when $s \sim 1$, diminished and undiminished orderly squash are infinitesimally close on domain F .

Theorem 4.6.3. *For a positive hyperplane arrangement $F \in \mathbb{R}^d$, consider F instead as a positive hyperplane arrangement in $\mathbb{R}\langle\varepsilon\rangle^d$ (i.e., replace F with $\text{Ext}(F, \mathbb{R}\langle\varepsilon\rangle)$). Then diminished orderly squash $\text{Sq}_t^{1-\varepsilon}(v)$ at speed $s = 1 - \varepsilon$ is continuous and gradual as a function of $t \in [0, \infty)_{\mathbb{R}\langle\varepsilon\rangle} \cap \mathbb{R}\langle\varepsilon\rangle_b$ and $v \in N \subset \mathbb{R}\langle\varepsilon\rangle^d$. Furthermore, diminished and undiminished orderly squash are infinitesimally close: $\text{Sq}_t^s(v) \sim \text{Sq}_t(v)$ for any bounded $t \geq 0_{\mathbb{R}\langle\varepsilon\rangle}$ and any $v \in N$. In particular, $\text{Sq}_t^{1-\varepsilon}$ endows orderly squashing Sq_t with the structure of an origami motion (Definition 4.5.1).*

Note. We emphasize that F is initially defined over the real numbers; this theorem is not true for an arbitrary positive hyperplane arrangement over $\mathbb{R}\langle\varepsilon\rangle$. The crucial feature is that F 's planes cannot be “too close to parallel”: when two planes have different positive normals $u_i \neq u_j$, meaning $u_i \cdot u_j < 1$, we require more than infinitesimal separation, $1 - u_i \cdot u_j \approx 0$. In Theorem 4.6.3, $u_i \cdot u_j$ is a real number less than 1, which suffices.

Proof. Our proof that diminished orderly squashing is continuous (Lemma 4.6.1) applies equally well to any positive hyperplane arrangement over $\mathbb{R}\langle\varepsilon\rangle$, so $\text{Sq}_t^{1-\varepsilon}$ is indeed continuous as a function of $t \in [0, \infty)_{\mathbb{R}\langle\varepsilon\rangle}$ (t need not be bounded here) and $N \subset \mathbb{R}\langle\varepsilon\rangle^d$.

We now argue that $\text{Sq}_t^{1-\varepsilon}(v)$ is infinitesimally close to $\text{Sq}_t(v)$ for any point $v \in F$ and any nonnegative, bounded $t \in \mathbb{R}\langle\varepsilon\rangle_b$. In fact we'll prove a slight strengthening: if two points $v \in F$ and $w \in N$ are infinitesimally close, then so are $\text{Sq}_t(v)$ and $\text{Sq}_t^{1-\varepsilon}(w)$.

When point v is following plane H_i during orderly squashing, say that v 's **heading** is the vector u_i . Likewise, when w trails plane H_i in diminished orderly squashing, say that w 's **heading** is the vector u_i , even though its velocity is actually $(1 - \varepsilon)u_i$. Note that v and w may have equal headings even when following different planes, so long as those planes have identical normal vectors.

If v and w begin infinitesimally close and then both move with the *same* heading u_i for y seconds (where $y \in \mathbb{R}\langle\varepsilon\rangle_b$), their new positions $v + y \cdot u_i$ and $w + y \cdot (1 - \varepsilon)u_i$ are still infinitesimally close, because y and u_i are both bounded. If instead v and w do not (necessarily) have the same headings but advance only for an infinitesimal amount of time z , they remain infinitesimally close, because each has moved at most a total distance of z . So it is enough to show that v 's and w 's headings disagree for only an infinitesimal amount of time in the interval $[0, t]$.

So suppose v and w are currently trailing planes H_i and H_j respectively, where $u_i \neq u_j$. We'll show that one of v or w (or both) must encounter a lower-ordered plane after an infinitesimal amount of time. Indeed, suppose first that $i < j$, so v lies on plane H_i , meaning w 's distance d to plane H_i is infinitesimal. If w hits plane H_i before changing heading, it will meet H_i at time $\frac{d}{1 - (1 - \varepsilon)u_i \cdot u_j}$, which is infinitesimal because $1 - (1 - \varepsilon)u_i \cdot u_j \gtrsim 0$, as explained in the Note preceding this proof. It may happen that v or w hits a different plane *before* w hits H_i , but in any case, one such event must happen within infinitesimal time, as claimed. The case $j < i$ is similar: the time for v to hit plane H_j in this case is governed by the simpler denominator $1 - u_i \cdot u_j$, which is likewise greater than infinitesimal.

By analogous reasoning, $\text{Sq}_t^{1-\varepsilon}(v)$ is a gradual function on N (for any $t \in \mathbb{R}\langle\varepsilon\rangle_b$): if two points $v, w \in N$ begin infinitesimally near each other, then during diminished orderly squashing, similar arguments show that their headings disagree for only an infinitesimal

total amount of time. And $\text{Sq}_t^{1-\varepsilon}(v)$ is a gradual function of t for fixed $v \in N$, because v 's speed $1 - \varepsilon$ is bounded. So by the Note following Definition 4.5.1, $\text{Sq}_t^{1-\varepsilon}(v)$ is gradual. \square

We also show that orderly squashing on convex polytopes obeys the medial axis gluing:

Theorem 4.6.4. *Let $P \subseteq \mathbb{R}^d$ be a convex polytope. The **medial axis gluing** of P identifies two boundary points p and q whenever there is a sphere contained entirely within P that touches P 's boundary at p and q (and possibly elsewhere). Then any such pair of points p and q are adjacent with no intervening layers (as in Definition 4.5.2) in any flattening of P obtained by orderly squashing P 's facets (with any ordering).*

Proof. Let F be a positive hyperplane arrangement built from P 's facets facing inward. Suppose $p \in F_i$ and $q \in F_j$ with $i > j$, and let c and t be the center and radius of the sphere initially tangent at p and q . By arguments similar to those in Section 4.3, none of the planes H_1, \dots, H_n meets point c during orderly squashing for time $u < t$, so orderly squashing brings p and q together at time t . We wish to show that there are no intervening layers at this time, as in Definition 4.5.2.

Consider the state of affairs after applying *diminished* orderly squashing for time t at speed $s = 1 - \varepsilon$. By similar arguments, p and q have followed planes i and j uninterrupted during this time, so their new locations $p' := p + (1 - \varepsilon)tu_i$ and $q' := q + (1 - \varepsilon)tu_j$ are infinitesimally close to c . We claim that the segment connecting p' to c is disjoint from $\text{Sq}_t^{1-\varepsilon}(F)$ except at $p' = \text{Sq}_t^{1-\varepsilon}(p)$, and similarly for q' . Indeed, each point $w \in F$ travels at speed $1 - \varepsilon$ during diminished orderly squashing and begins at distance at least t away from c , so after t seconds of diminished orderly squashing, w must be at least $(1 - \varepsilon)t$ away from c . Segment $p'c$ lies strictly within that distance except at p' , and by injectivity, there is no point $w \in F$ other than p with $\text{Sq}_t^{1-\varepsilon}(w) = p'$. This proves the claim.

Let Q be the path $p'cq'$ in $\mathbb{R}\langle\varepsilon\rangle^d$, which is disjoint from $\text{Sq}_t^{1-\varepsilon}(F)$ except at its endpoints, as above. Now continue the diminished orderly squashing from time t to any bounded time $t' \geq t$, and note that Q goes along for the ride, because $p'c$ lies behind plane H_i and cq' lies behind plane H_j at time t . Let Q' be the image of Q at time t' . By injectivity, Q' is a simple path and is disjoint from $\text{Sq}_{t'}^{1-\varepsilon}(F)$ except at its endpoints, $\text{Sq}_{t'}^{1-\varepsilon}(p)$ and $\text{Sq}_{t'}^{1-\varepsilon}(q)$. And because diminished orderly squashing is gradual, path Q' is contained in an infinitesimal neighborhood of (both of) its endpoints. \square

4.7 Folding Tessellations

By applying orderly squashing to a shallow, almost planar polyhedron, it seems reasonable that orderly squashing should allow us to fold planar tessellations as well, as in Figure 4-7. This is indeed true for certain patterns, but modifications are necessary. Plain orderly squashing as described in Section 4.3 clearly does not work directly on a flat planar figure, because if each face moves along its inward normal, the entire pattern simply translates and no creasing occurs. Instead consider the following wishful thinking. We try to move each face of a planar graph G in the *same* plane as the figure itself: say each face f_i of G translates continuously with velocity vector w_i for one second, adding creases between polygons *somehow* to maintain the paper's integrity and creating crimps along the edges of the original crease pattern as shown in Figure 4-7.

Before even considering the *somehow*, what conditions must velocity vectors w_i satisfy for such an origami tessellation fold to even be feasible? If faces f_i and f_j share an edge e , to prevent the paper from shearing or tearing along e , we need f_i and f_j to move directly toward

each other (in a relative sense), that is, the relative velocity $w_j - w_i$ must be orthogonal to edge e in the direction pointing from f_j to f_i . (Their shared crimp will then have width $|w_i - w_j|/2$ after one second.) This orthogonality condition means precisely that the vectors w_i define an orthogonal embedding of the dual graph of G , and in particular that G is a spiderweb. In the language of stresses, this constraint exactly says that $e \mapsto |w_i - w_j|/|e|$ is a positive equilibrium stress [75]. Note that the same spiderweb condition arises in Lang and Bateman’s related origami tessellation technique using twist-folds [66].

In the rest of this section, we show that any spiderweb may be folded into an origami tessellation using an orderly squashing technique. It is notable that infinitesimals play a crucial role, not only when providing infinitesimal offsets for origami structure, but even in describing the underlying geometry of these tessellation folds.

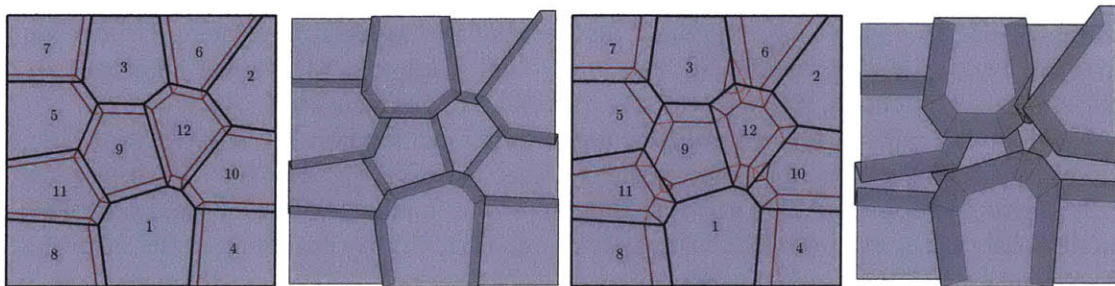


Figure 4-7: Origami tessellations made from a spiderweb graph through a limiting orderly squash. Small offsets (left) result in elegant tessellations, while large offsets (right) have more interaction.

Our method proceeds roughly as follows: beginning with a spiderweb G , an orthogonal dual embedding $\{w_i\}$ as above is equivalent to giving a strictly convex polyhedral lifting of G : each vertex $v_i = (x_i, y_i)$ of G is assigned a z -coordinate z_i such that each face of G lifts to a planar polygon in \mathbb{R}^3 and such that all of G lifts to a convex surface. If we instead use the z -coordinates $r \cdot z_i$ for some small $r > 0$, the lifting is still convex but very shallow, and orderly squashing on this shallow surface resembles a folding of the original planar figure. In the limit as r tends toward 0, we recover the desired folding of G . This limit, however, is subtle: on a shallower lift (r closer to 0), orderly squashing must proceed for a longer time (approaching infinity) because the dihedral angle between neighboring planes is very near 180° . To overcome this challenge, we appeal again to $\mathbb{R}\langle\varepsilon\rangle$: we will scale z -coordinates by an *infinitesimal*, $r = \varepsilon$, and let orderly squashing proceed for infinite time c/ε , for a chosen real number c .

With small face offsets (smaller c), the tessellations produced with this method are rather tame, involving only a crimp along each edge and small compensations near each vertex. But larger offsets are also possible without modification, and the results are “fatter” tessellations with more interaction between the layers.

4.7.1 Formal Setup for Spiderweb Folding

Begin with an embedded spiderweb graph $G = (V, E)$ in \mathbb{R}^2 , where as usual we will blur the distinction between a vertex and its position in the plane. We wish to fold a sheet of paper A corresponding to the union of G ’s bounded faces, $\{f_1, \dots, f_n\}$, which are numbered in the priority order that will be used for orderly squashing. Let G' be an embedded orthogonal dual graph of G , also in \mathbb{R}^2 , with vertices $w_i = (a_i, b_i)$ corresponding to faces f_i . (The existence

of G' is equivalent to G being a spiderweb.) The Maxwell-Cremona correspondence [84] says that G' corresponds to a specific convex polyhedral **lifting**, i.e., a convex, piecewise linear function $L : A \rightarrow \mathbb{R}^3$ where each $(x, y) \in A$ is lifted to some point $L((x, y)) = (x, y, z)$ projecting onto (x, y) . Specifically, this lifting L has the property that each face f_i gets lifted to a plane H_i whose (unnormalized) normal is $(-a_i, -b_i, 1)$. Note that $L(f_i)$ lies on plane H_i and *in front of* each other plane $H_j, j \neq i$, because L is convex; in other words, the positive hyperplane arrangement defined by these planes contains $L(A)$. For simplicity, we may assume $w_1 = (0, 0)$ (so $L(f_1)$ has normal $(0, 0, 1)$), and $L(f_1)$ contains the origin $o = (0, 0, 0)$.

We will run orderly squashing, but not directly on the planes H_i . Work instead over $\mathbb{R}\langle\varepsilon\rangle$, and let T^ε be the transformation $T^\varepsilon(x, y, z) = (x, y, \varepsilon z)$ that scales z coordinates by the infinitesimal ε . Now, $L^\varepsilon := T^\varepsilon \circ L$ is an “infinitesimally flattened” version of L , with faces lying in planes $T^\varepsilon(H_i)$. (Note: the lift L^ε is the result of applying the Maxwell-Cremona correspondence directly to the scaled orthogonal dual graph $\varepsilon \cdot G'$.) We will apply orderly squashing to *these* planes, $T^\varepsilon(H_i)$, whose (unnormalized) normals are $(-\varepsilon a_i, -\varepsilon b_i, 1)$.

In full detail, our **spiderweb folding algorithm** is as follows: lift A to $A^\varepsilon := L^\varepsilon(A)$, orderly squash this surface for time c/ε for a chosen positive *real* number c , translate the result so that the origin $o \in A^\varepsilon$ returns to the origin, and finally take \lim_ε of the result. This describes only the *geometry* of the folding; layer ordering will be considered in the following subsection.

Theorem 4.7.1. *Let lift $A^\varepsilon = L^\varepsilon(A)$ be as described above. Then for any nonnegative $t \in \mathbb{R}\langle\varepsilon\rangle$, even unbounded t , the function $\Delta\text{Sq}_t(p) := \text{Sq}_t(p) - \text{Sq}_t(o)$, defined for $p \in A^\varepsilon$, is continuous, gradual, piecewise isometric, and has image in $\mathbb{R}\langle\varepsilon\rangle_b^3$.*

Proof. Continuity and piecewise isometry follow by the same arguments in Theorem 4.6.3 and hold for *any* positive hyperplane arrangement over $\mathbb{R}\langle\varepsilon\rangle$. To prove ΔSq_t is gradual, suppose $v = (v_1, v_2, v_3)$ and $w = (w_1, w_2, w_3)$ lie on $L^\varepsilon(A)$ and are infinitesimally close. The line segment S connecting $(v_1, v_2) \in A$ and $(w_1, w_2) \in A$ has infinitesimal length, as does the piecewise linear path $L^\varepsilon(S) \subset A^\varepsilon$ connecting v with w . Then the piecewise isometric and piecewise linear function ΔSq_t transforms path $L^\varepsilon(S)$ into another piecewise linear path of the same infinitesimal length, so we indeed have $\Delta\text{Sq}_t(v) \sim \Delta\text{Sq}_t(w)$. The proof that $\Delta\text{Sq}_t(A) \subseteq \mathbb{R}\langle\varepsilon\rangle_b^3$ is similar: any $v, w \in A^\varepsilon$ are connected by a polygonal path in A^ε of bounded total length, and ΔSq_t preserves this path’s length. \square

Note. When considered as a function on the entire positive hyperplane arrangement, instead of just the subset A^ε , ΔSq_t may fail to be gradual when t : it is no longer true that points on this positive hyperplane arrangement that are infinitesimally close are connected by an infinitesimal-length path that is *contained in the arrangement*.

Theorem 4.7.2. *For a positive real number c , the function sending $p \in A \subseteq \mathbb{R}^2$ to $\text{Tess}_c(p) := \lim_\varepsilon \Delta\text{Sq}_{c/\varepsilon}(L^\varepsilon(p))$ is continuous and piecewise isometric, and each point $\text{Tess}_c(p)$ lies in the plane $z = 0$. Furthermore, for any point $p \in A$ such that $L^\varepsilon(p)$ lies in the region $F_i^{c/\varepsilon}[0]$ —i.e., lies on face F_i of the hyperplane arrangement and does not stick to a lower ordered plane during time $[0, c/\varepsilon]$ —the function Tess_c sends p simply to $p - c \cdot (a_i, b_i)$. (We have identified the plane \mathbb{R}^2 containing A with the plane $z = 0$ in \mathbb{R}^3 .)*

Note. This Theorem formalizes the intuition that Tess_c works by “moving each face f_i with velocity $-w_i = -(a_i, b_i)$ for c seconds and resolving overlaps near the edges and vertices,” as described at the beginning of Section 4.7.

Proof. For any point $p \in A$ (which we interpret as lying in the plane $z = 0$ of \mathbb{R}^3 or $\mathbb{R}(\varepsilon)^3$), the third coordinate of $L^\varepsilon(p)$ is infinitesimal, so $p \sim L^\varepsilon(p)$. The continuity of Tess_c now follows from the fact that $\Delta\text{Sq}_{c/\varepsilon}$ is gradual and has image in $\mathbb{R}(\varepsilon)_b$. Piecewise isometry of Tess_c likewise follows from piecewise isometry of $\Delta\text{Sq}_{c/\varepsilon}$ on A^ε , because distances in A and in A^ε are infinitesimally related.

If $q = L^\varepsilon(p)$ lies in $F_i^{c/\varepsilon}[0]$, so it does not encounter another plane during orderly squashing before time c/ε , then

$$\text{Sq}_{c/\varepsilon}(q) = q + \frac{c}{\varepsilon} \cdot \frac{(-\varepsilon a_i, -\varepsilon b_i, 1)}{\sqrt{1 + \varepsilon^2(a_i^2 + b_i^2)}}.$$

Because $o \in F_1$ and $u_1 = (0, 0, 1)$, we have $\text{Sq}_{c/\varepsilon}(o) = o + \frac{c}{\varepsilon}(0, 0, 1)$, and the difference $\Delta\text{Sq}_{c/\varepsilon}(q)$ is $q + (-ca_i, -cb_i, z)$, where

$$z = \frac{c}{\varepsilon} \left(\frac{1}{\sqrt{1 + \varepsilon^2(a_i^2 + b_i^2)}} - 1 \right).$$

And because $\sqrt{1 + \varepsilon^2(a_i^2 + b_i^2)} = 1 + (a_i^2 + b_i^2)\varepsilon^2/2 + O(\varepsilon^4)$, it may be verified that $z \sim 0$, so $\lim_\varepsilon \Delta\text{Sq}_{c/\varepsilon}(q)$ indeed lies on plane $z = 0$.

Finally, if $q \in A^\varepsilon$ lies on F_i but *not* in $F_i^{c/\varepsilon}[0]$, we must verify that $\lim_\varepsilon \Delta\text{Sq}_{c/\varepsilon}(q)$ also lands on plane $z = 0$. But q lies in some region $F_i^{c/\varepsilon}[j_1, \dots, j_k, 0]$, so $\text{Sq}_{c/\varepsilon}(q) = \text{Sq}_{c/\varepsilon}(q')$ for some point $q' \in F_{j_k}^{c/\varepsilon}[0]$, and we know the result is true for q' . This completes the proof. \square

4.7.2 Layer Ordering for Tessellation Folding

Now that we have described the geometry of our spiderweb folding, Tess_c , we proceed to give Tess_c an origami structure by carefully separating its layers with diminished orderly squashing, as in Section 4.6. Our proof is similar to that of Theorem 4.6.3, but an additional subtlety arises: Theorem 4.6.3 exploits the fact that non-parallel normals are more than infinitesimally far from parallel, but in Tess_c , *all* of the planes are infinitesimally close to parallel! This means that even infinitesimally close points $v, w \in N$ might follow different planes for an unbounded time, moving more than infinitesimally far from each other in the process. We thus have to work harder to make sure our diminished orderly squash does not stray too far from Tess_c , and the key insight is to simply force them to drift apart more slowly: instead of diminished orderly squashing with speed $s = 1 - \varepsilon$, we use speed $s = 1 - \varepsilon^{(n^2)}$.

Lemma 4.7.3. *The planes defining A^ε , as above, are not too parallel: two distinct unit normals u_i and u_j among these planes satisfy $1 - u_i \cdot u_j = \Theta(\varepsilon^2)$.*

Proof. Let $w_i = (a_i, b_i)$ and $w_j = (a_j, b_j)$, and recall that these vectors are *real*. Set $v_i = (-\varepsilon a_i, -\varepsilon b_i, 1)$ so that $u_i = v_i/|v_i|$, and define v_j similarly. We're given that $u_i \neq u_j$, which implies $w_i \neq w_j$. We have

$$1 - u_i \cdot u_j = 1 - \frac{v_i \cdot v_j}{|v_i| \cdot |v_j|} = \frac{\sqrt{(1 + \varepsilon^2(a_i^2 + b_i^2))(1 + \varepsilon^2(a_j^2 + b_j^2))} - (1 + \varepsilon^2(a_i a_j + b_i b_j))}{\sqrt{(1 + \varepsilon^2(a_i^2 + b_i^2))(1 + \varepsilon^2(a_j^2 + b_j^2))}}.$$

Using the facts that $\sqrt{1+e} = 1 + \frac{e}{2} + O(e^2)$ and $\frac{1}{1+e} = 1 + O(e)$ for any infinitesimal e , the above quantity simplifies to

$$(|w_i - w_j|^2 \cdot \varepsilon^2/2 + O(\varepsilon^4)) (1 + O(\varepsilon^2)).$$

The term $|w_i - w_j|^2$ is a nonzero real number, so this quantity is $\Theta(\varepsilon^2)$, as claimed. \square

Theorem 4.7.4. *Let A^ε and c be as above. Then for $s = 1 - \varepsilon^{(n^2)}$, the translation of diminished orderly squashing $\Delta\text{Sq}_{c/\varepsilon}^s(q) := \text{Sq}_{c/\varepsilon}^s(q) - \text{Sq}_{c/\varepsilon}^s(o)$ is a continuous, gradual, and injective function of $q \in A^\varepsilon$, and furthermore, $\Delta\text{Sq}_{c/\varepsilon}^s(q) \sim \Delta\text{Sq}_{c/\varepsilon}(q)$. Because $\Delta\text{Sq}_{c/\varepsilon}(L^\varepsilon(p)) \sim \text{Tess}_c(p)$ for all $p \in A$, this shows that $\Delta\text{Sq}_{c/\varepsilon}^s \circ L^\varepsilon$ defines an origami structure for Tess_c .*

Proof. Diminished orderly squashing Sq_t^s is continuous and injective for any positive hyperplane arrangement defined over $\mathbb{R}(\varepsilon)$, any $t \geq 0$ (even when unbounded) and any $s \in [0, 1)$, by the same proof used in Theorem 4.6.3 (or, alternatively, by one of the transfer principles, Theorem 3.2.2).

Note that $\Delta\text{Sq}_{c/\varepsilon}^s(o)$ (diminished) and $\Delta\text{Sq}_{c/\varepsilon}(o)$ (not diminished) differ from each other by $\frac{c}{\varepsilon} \cdot \varepsilon^{(n^2)} \cdot (0, 0, 1) \sim (0, 0, 0)$, so showing that $\text{Sq}_{c/\varepsilon}^s(q) \sim \text{Sq}_{c/\varepsilon}(q)$ is enough to prove the desired $\Delta\text{Sq}_{c/\varepsilon}^s(q) \sim \Delta\text{Sq}_{c/\varepsilon}(q)$. For this, we proceed much like the proof of Theorem 4.6.3 by comparing q 's trajectories under $\text{Sq}_{c/\varepsilon}$ and $\text{Sq}_{c/\varepsilon}^s$, and by separately considering the effects of (i) unequal heading and (ii) equal heading.

First, suppose at some moment during $[0, c/\varepsilon]$, point $q \in A^\varepsilon$ has moved to v under orderly squash and to w under diminished orderly squash, and let $|v - w| = d$. Suppose v and w travel for some interval of time with fixed, unequal headings u_i and u_j respectively. As in the proof of Theorem 4.6.3, the time until either v or w sticks to a lower-ordered plane is at most $\frac{d}{1 - (1-s)u_i \cdot u_j} = O(d/\varepsilon^2)$, so the distance between v and w grows to at most $O(d/\varepsilon^2)$ during this interval. There are at most $\binom{n}{2} = n(n-1)/2$ such time intervals during q 's travels because indices i and j decrease monotonically over time and therefore (i, j) and (j, i) cannot both occur.

On the other hand, if these points v and w travel with the *same* heading for some interval of $r \leq c/\varepsilon$ seconds, their distance d from each other grows to at most $d + s \cdot r = d + O(\varepsilon^{n^2-1})$ during this interval.

Because v and w start at distance 0 apart at time $t = 0$ (since they both start at q), putting the above two cases together proves that the distance between v and w grows to at most

$$O\left(\frac{\varepsilon^{n^2-1}}{(\varepsilon^2)^{\binom{n}{2}}}\right) = O(\varepsilon^{n-1}) \sim 0$$

during their full trajectories over time $[0, c/\varepsilon]$. In other words, $\text{Sq}_{c/\varepsilon}^s(q) \sim \text{Sq}_{c/\varepsilon}(q)$, as required.

This also proves that $\Delta\text{Sq}_{c/\varepsilon}^s$ is gradual on A^ε , because if distinct points $v, w \in A^\varepsilon$ have $v \sim w$, then $\Delta\text{Sq}_{c/\varepsilon}^s(v) \sim \Delta\text{Sq}_{c/\varepsilon}(v) \sim \Delta\text{Sq}_{c/\varepsilon}(w) \sim \Delta\text{Sq}_{c/\varepsilon}^s(w)$, using the fact that (translated) orderly squashing $\Delta\text{Sq}_{c/\varepsilon}$ is gradual from Theorem 4.7.2. This concludes the proof. \square

Chapter 5

Hardness of Polyhedron Edge Unfolding

Chapter Summary. We prove that it is strongly NP-complete to decide whether a given orthogonal polyhedron has a (nonoverlapping) edge unfolding. The result holds even when the polyhedron is topologically convex, i.e., is homeomorphic to a sphere, has faces that are homeomorphic to disks, and where every two faces share at most one edge.

5.1 Introduction

An *edge unfolding* of a polyhedron consists of cutting the surface along a subset of its edges in such a way that the surface can be unfolded into one planar piece without overlap.* Edge unfoldings have a long history, dating back to Albrecht Dürer in 1525; see [43]. The most famous open question is whether every convex polyhedron has an edge unfolding [76], but nonconvex polyhedra are even more interesting for practical manufacturing applications. The theoretical study of such unfoldings began at CCCG 1998 [23] and CCCG 1999 [20, 21]. Biedl et al. [23] found some orthogonal polyhedra with no edge unfoldings, but the examples had faces with holes or two faces that shared two edges. Bern et al. [21] found a triangulated polyhedron with no edge unfolding that is homeomorphic to a sphere, implying that the polyhedron is *topologically convex*—has the graph (1-skeleton) of a convex polyhedron. In the journal version of their CCCG 1999 paper [21], they asked for the computational complexity of deciding whether a given triangulated polyhedron has an edge unfolding.

In this paper, we settle the computational complexity of the closely related problem of deciding whether a topologically convex orthogonal polyhedron has an edge unfolding. Specifically, we prove this *Orthogonal Edge Unfolding* problem is strongly NP-complete.

This Chapter was published in the 23rd Canadian Conference on Computational Geometry (CCCG 2011) with Erik D. Demaine [3].

5.2 Unique Coordinate Square Packing

The Square Packing problem asks, given n squares s_1, \dots, s_n of side-lengths a_1, \dots, a_n and a target distance d , whether there is some (non-overlapping) orthogonal packing of the squares

*We allow boundary edges to touch in unfoldings, requiring only that the interior of the cut surface does not overlap itself.

s_i into a square of side-length d . This is known to be strongly NP-complete [67]. We first show that we may impose a few simplifying assumptions on the packings produced by this problem:

Definition 5.2.1 (Unique Coordinate Square Packing). An instance of the Unique Coordinate Square Packing (UCSP) promise problem has the form $(d, (a_1, \dots, a_n))$, where all values are positive integers and $a_i \leq d - 2$ for each $1 \leq i \leq n$. In a YES instance, there exists an orthogonal packing of n squares s_1, \dots, s_n of side-lengths a_1, \dots, a_n into the square $D = [0, d] \times [0, d] \subset \mathbb{R}^2$ satisfying the following additional properties:

- all vertices of all squares in the packing have integer coordinates,
- no two vertices of two different squares have the same x - or y -coordinate, and
- no square in the packing touches the boundary of D .

In a NO instance, there does not exist any orthogonal packing of the s_i into D .

Theorem 5.2.2. *The Unique Coordinate Square Packing problem is strongly NP-hard.*

Proof. Say that a square s is **positioned at** (x, y) if its lower-left corner has coordinates (x, y) . For a Square Packing instance $(d, (a_1, \dots, a_n))$ with s_1, \dots, s_n , and D as above, there exists a packing of the squares s_i in D if and only if there exists a packing with integer coordinates: indeed, repositioning each square position (x_i, y_i) in a packing to $(\lfloor x_i \rfloor, \lfloor y_i \rfloor)$ yields an integer packing.

We reduce from the Square Packing problem. Given an instance $(d, (a_1, \dots, a_n))$ of Square Packing, return the UCSP instance $(d', (a'_1, \dots, a'_n))$ where $d' = d(2d + 1)(n + 1)$ and $a'_i = 2d(n + 1)a_i$.

Suppose $(d', (a'_1, \dots, a'_n))$ is a YES instance of UCSP. Scaling down by a factor of $2d(n + 1)$, it follows that squares of side lengths a_1, \dots, a_n can be packed inside the square $[0, d + \frac{1}{2}]^2$. As above this packing can be taken to have integer coordinates, meaning the squares actually fit in $[0, d]^2 = D$. So $(d, (a_1, \dots, a_n))$ is a YES instance.

Conversely, suppose $(d, (a_1, \dots, a_n))$ is a YES Square Packing instance, so there is some integer packing where each s_i is positioned at (x_i, y_i) . Let b'_i be a square with side length a'_i . It may be checked that positioning b'_i at $((2d + 1)(n + 1)x_i + i, (2d + 1)(n + 1)y_i + i)$ in $D' = [0, d']^2$ produces a valid packing. Furthermore, the coordinates of each vertex of b'_i are congruent to $i \pmod{n + 1}$, which proves the last two required properties. \square

5.3 Overview

This section provides an overview of the detailed constructions to follow.

We first consider the problem of unfolding orthogonal polyhedra *with boundary* in Section 5.4, proving hardness by reduction from a UCSP instance $(d, (a_1, \dots, a_n))$. We construct a polyhedron B with boundary (Figure 5-4) involving n squares b_i with side-lengths a_i (call these “blocks”) surrounded by filler material. The polyhedron is designed to force the blocks to unfold inside a “cage” of shape $d \times d$, such that an unfolding exists if and only if there exists a square packing. (In the construction below, the blocks and cage are scaled up by a large factor q .) As the unfolding must remain connected, we use thin “wires” made from the filler material to “wind” around the blocks and connect them to the boundary of the cage. The un-needed filler material winds itself out of the way. The universally windable wires

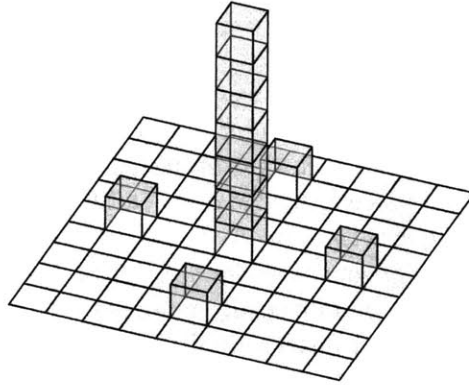


Figure 5-1: A depiction of an “atom,” the polyhedral surface with 9×9 square boundary that enables universal wire unfolding as in Theorem 5.4.1. An atom is composed of 125 unit-square faces.

are described and proved in Section 5.4.1, and the details of the unfolding are presented in Sections 5.4.2, 5.4.3, and 5.4.4.

In Section 5.5 we reduce to orthogonal polyhedra without boundary by extending B into a polyhedron C with the property that C has an unfolding if and only if B does. This is accomplished with two U-shaped polygons (Figure 5-8) that must be separated from B to avoid overlap, which forces the extra material of C not to interfere with the unfolding of B .

5.4 Polyhedron With Boundary

In this section we show that the edge unfolding problem for an orthogonal polyhedron *with boundary* is NP-hard.

5.4.1 Atoms and Universal Wire Unfolding

As described in the overview, we require “winding wires” that can unfold into an arbitrarily chosen orthogonal path. We construct those here.

Define an **atom** as the polyhedral surface with boundary in Figure 5-1, whose boundary is a 9×9 square. The width of an atom is called the **atomic width**, $w_A = 9$. Atoms are named thus since they are the basic “winding wire” unit, and also due to their tiny size relative to many constructions to follow.

For a finite or infinite grid G of $u \times u$ squares in the xy -plane and integers i, j , write $G[i, j]_u$ for the $(i, j)^{\text{th}}$ cell in G , i.e., the $u \times u$ cell positioned at (ui, uj) . Similarly, if e is a directed line segment of length ℓ and u evenly divides ℓ , express e as the union of ℓ/u directed segments of length u and let $e[i]_u$ be the i^{th} such segment.

Define a **wire W of length k in G** as a simple path of connected squares in G : specifically, a collection of distinct squares $c_i = c_i(W)$ ($0 \leq i \leq k - 1$) in G and distinct, oriented edges $e_i = e_i(W)$ ($0 \leq i \leq k$) such that e_i is the common edge of cells c_{i-1} and c_i for each $1 \leq i \leq k - 1$, edge e_0 (the **starting edge**) is an edge of the **starting cell** c_0 , and e_k (the **ending edge**) is an edge of the **ending cell** c_{k-1} . Edge e_i is oriented to trace the boundary of c_{i-1} clockwise, or equivalently, to trace the boundary of c_i counterclockwise. (Use the former condition for e_k and the latter for e_0 .) It is convenient to discuss the **medial path** of a wire that connects the centers of $e_0, c_0, e_1, \dots, c_{k-1}, e_k$ sequentially. The

wire turns **right**, **straight**, or **left** at square c_i if the medial path turns right, straight, or left there.

If W is a wire of $w_A \times w_A$ squares in the xy -plane, we can form the associated **wire of atoms** $A(W)$, a polyhedral surface with boundary, by replacing each square c_i with an atom a_i pointing in the positive z -direction such that atoms a_{i-1} and a_i are connected along the edge corresponding to e_i . Each unit-length edge $e_i[s]_1$ (for $0 \leq s \leq w_A - 1 = 8$) corresponds to an edge of one or two unit-square faces on $A(W)$.

Define a **flatom*** as a $w_F \times w_F$ square where $w_F = 27$ is the **flatomic width**. We will now show that wires of atoms can be universally unfolded in the following sense: roughly, any wire of k atoms can be unfolded inside any desired wire of k flatoms, while ensuring that the middle of each atom edge unfolds to the center of the corresponding flatom edge (or one unit away from center). An example is shown in Figure 5-2: there, a path of atoms that turns Right, Right, Left, then Straight is unfolded into the path of flatoms with the entirely different turn sequence Right, Left, Straight, Left.

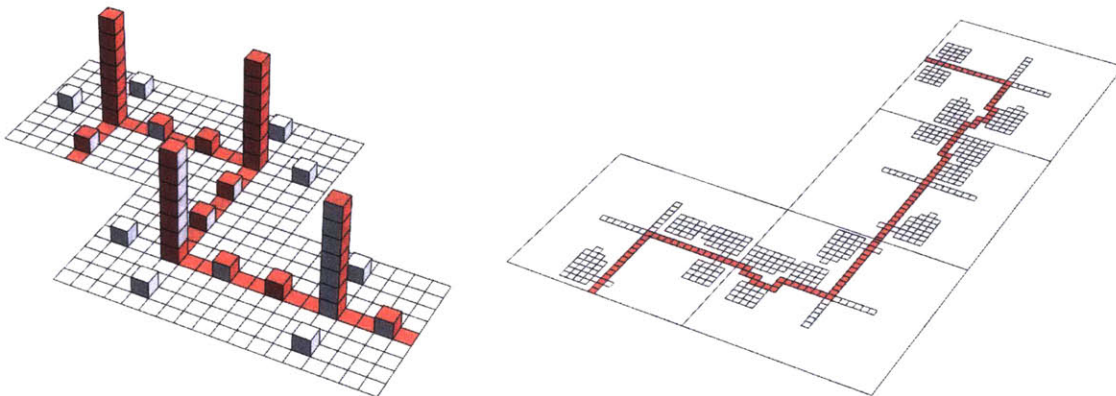


Figure 5-2: An illustration of Theorem 5.4.1: a path of atoms that turns right, right, left, straight in sequence is unfolded into the path of flatoms with the entirely different turn sequence right, left, straight, left.

Theorem 5.4.1. *Let W and W' be any two wires of length k with side-lengths w_A and w_F respectively. Then for each $t \in \{12, 13, 14\}$ there is an edge unfolding of the wire of atoms $A(W)$ that lies inside W' such that $e_0(W)[4]_1$ (the middle unit edge of $e_0(W)$) unfolds to $e_0(W')[t]_1$ (i.e., the middle edge of $e_0(W')$ or one unit away) and $e_k(W)[4]_1$ unfolds to $e_k(W')[u]_1$ for some $u \in \{12, 13, 14\}$. Furthermore, this unfolding can be accomplished so that t and u have the same (resp., different) parity when W and W' together have an even (resp., odd) total number of left and right turns.*

Proof. By induction on k , it suffices to prove only the case $k = 1$. There are thus 27 cases: W turns right, straight or left; W' turns right, straight, or left; and $e_0(W)[4]_1$ unfolds to $e_1(W')[12]_1$, $e_1(W')[13]_1$, or $e_1(W')[14]_1$. We label these unfoldings of an atom by a quadruple $[X, Y, t, u]$, where $X, Y \in \{L, S, R\}$ [†] indicate the directions of the turns of wires W and W' respectively, and t and u are as above. We must show that each of the 27

*short for “flat atom”

[†]These are abbreviations for Left, Straight, and Right turns.

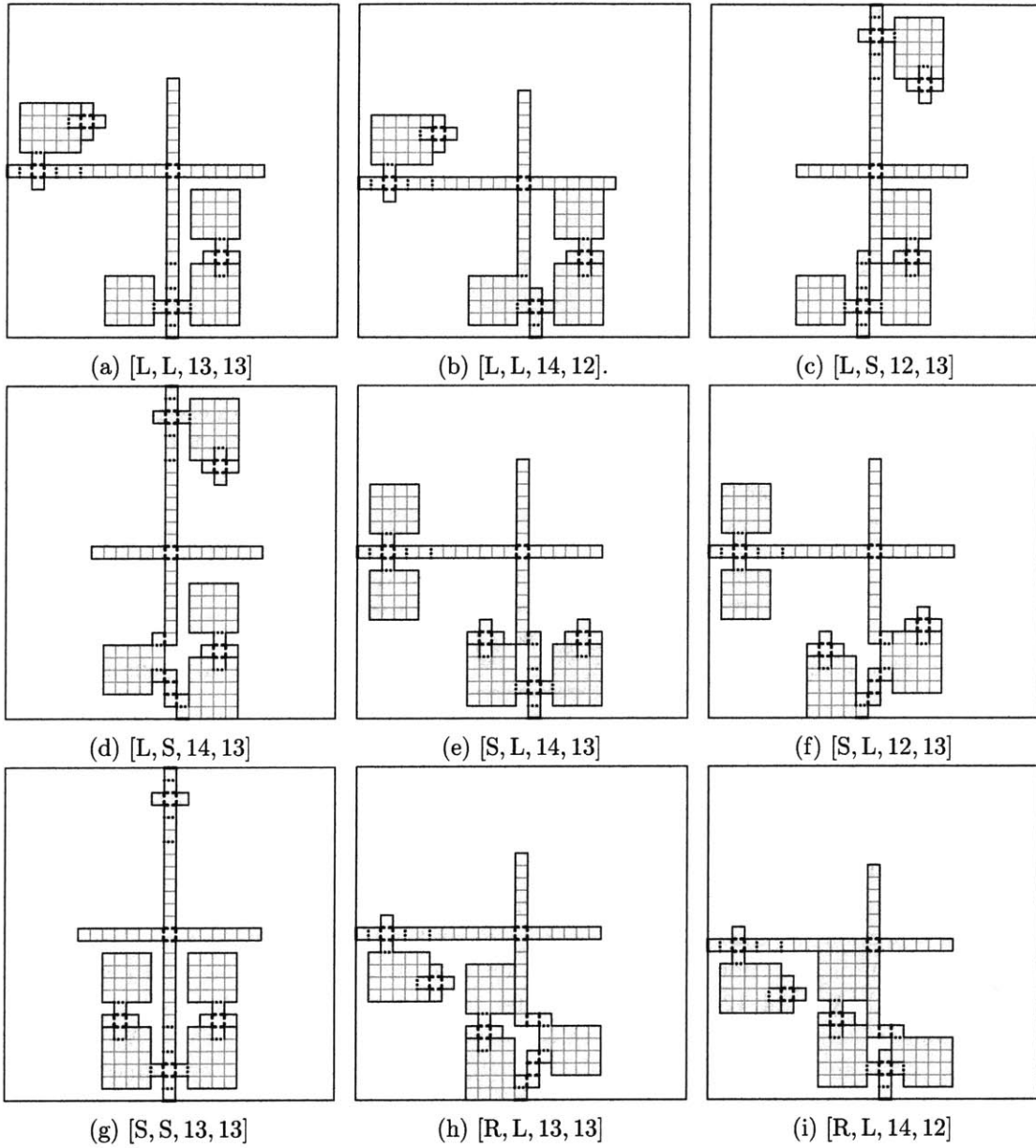
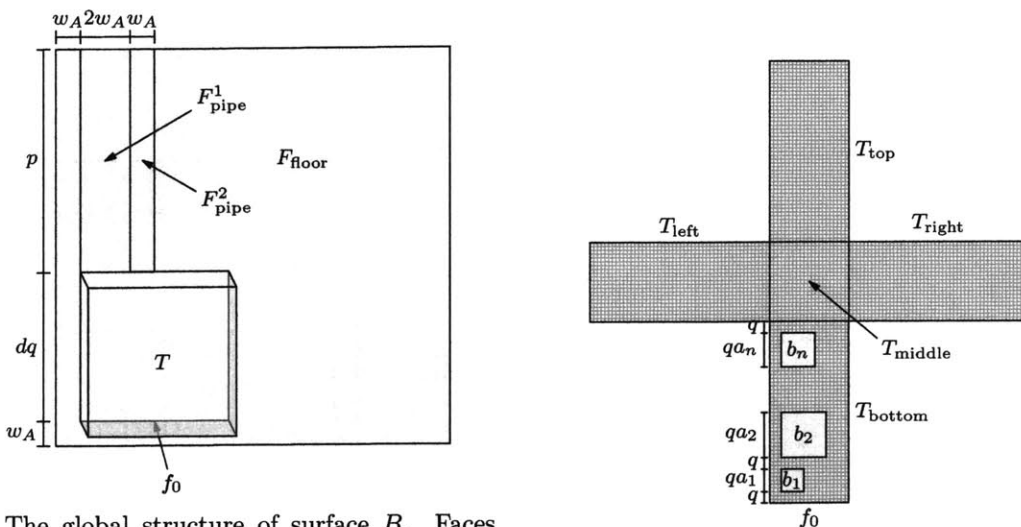


Figure 5-3: All nine required unfoldings of an atom inside a flatom. Each unfolding is labeled with $[X, Y, t, u]$ as described in the proof of Theorem 5.4.1. Solid black lines indicate cuts, dotted lines are valley folds, and dashed lines are mountain folds. Gray lines are unchanged edges.



(a) The global structure of surface B . Faces F_{floor} , F_{pipe}^1 , and F_{pipe}^2 are each a single polygon, but tower T is mostly covered with molecules as detailed in section 5.4.2. For ease of viewing, the image here is not drawn to scale: for example, the width of the two pipes is significantly smaller than the width of the tower T .

(b) A detail of the tower, T . Each surface of T is entirely tiled with molecules except for T_{bottom} , which has bricks b_1, \dots, b_n —each a single square face—arranged as shown.

Figure 5-4: Detailed depiction of polyhedral surface B .

tuples (X, Y, t) appears in some unfolding $[X, Y, t, u]$ with the required parity constraints on t and u .

If $X \in \{L, S, R\}$, let $-X$ be the opposite turn direction, i.e., $-L = R$, $-S = S$, and $-R = L$. Notice that reflecting an $[X, Y, t, u]$ configuration produces a $[-X, -Y, 26-t, 26-u]$ unfolding, whereas reversing the direction (i.e., swapping $e_0(W)$ with $e_1(W)$ and $e_0(W')$ with $e_1(W')$) produces a $[-X, -Y, 26-u, 26-t]$ unfolding. Applying both transformations therefore gives rise to an $[X, Y, u, t]$ unfolding.

We now claim that the 9 unfoldings shown in Figure 5-3 cover all 27 required cases. Indeed, by reflection, we may assume that $Y \neq R$, and if $Y = S$ then $X \neq R$. Next, if X and Y both turn or both don't turn, then it suffices to show $[X, Y, 13, 13]$ and just one of $[X, Y, 12, 14]$ and $[X, Y, 14, 12]$, because the latter two may be obtained from each other by reversing and reflecting. Similarly, if there is one turn among X and Y , then $[X, Y, 12, 13]$ and $[X, Y, 14, 13]$ suffice. Finally, the $(X, Y) = (S, S)$ cases are covered by translations of Figure 5-3g. It is apparent that in each unfolding, t and u have the same parity if W and W' together have no turns (both go straight) or two turns (each goes left or right), and otherwise t and u have opposite parity. \square

5.4.2 The Construction

Here we specify the polyhedron with boundary used in the reduction. The remainder of Section 5.4 is devoted to proving its correctness.

It will be useful to package atoms into a **molecule**: a 2×2 grid of atoms whose boundary is a $w_M \times w_M$ square, where $w_M = 2w_A = 18$. Much of the reduction below uses a molecule as a basic unit of construction.

Begin with a Unique Coordinate Square Packing instance $(d, (a_1, \dots, a_n))$. Define $q = 2^5 \cdot 3^4 \cdot nd$ (a large scale factor), and let $q_M = q/w_M$ be the number of molecules that fit across a distance q . Also set $t = (n + 1 + a_1 + \dots + a_n)q_M$, and $p = 500(4dq_M t + t^2) + 3w_A$; these choices will be explained shortly. Define the polyhedron with boundary $B(d, (a_1, \dots, a_n))$ as the surface shown in Figure 5-4, to be described in more detail presently. The diagram is oriented so that the positive x and y directions are right and up respectively, and z is out of the page toward the reader.

The face F_{floor} in Figure 5-4a is a single polygon formed by creating a $w_F \times p$ hole, H_{drain} , and a $dq \times dq$ hole, H_{cage} , in a large square of size $\ell = p + dq + w_A$. The two faces F_{pipe}^1 and F_{pipe}^2 , of widths $2w_A$ and w_A respectively, exactly fill H_{drain} . Five (not flat!) polyhedral surfaces $T_{\text{bottom}}, T_{\text{left}}, T_{\text{mid}}, T_{\text{right}}, T_{\text{top}}$, shown in detail in Figure 5-4b, form the sides of the **tower**, T , which connects along the boundary of H_{cage} . The polyhedral surface T_{bottom} , whose boundary is a $dq \times tw_M$ rectangle, is a $dq_M \times t$ grid of molecules facing away from the tower except for the n square faces b_1, \dots, b_n —called **bricks**—of side-lengths qa_1, \dots, qa_n , where brick b_i is positioned at $(q, (i + a_1 + \dots + a_{i-1})q)$ relative to the bottom-left corner of T_{bottom} . (For T_{bottom} , “right” and “up” refer to the positive x and z directions, respectively, as in Figure 5-4b.) The parameter t was chosen so that these bricks exactly fit with q separation from each other and from the bottom and top edges. Recall that $a_i \leq d - 2$ for each i , so there is at least q separation between each brick and the right edge of T_{bottom} . The other four sides of T , which have dimensions $dq \times tw_M$ or $dq \times dq$, are completely tiled with outward-facing molecules.

A single molecule has surface area 500, so the total surface area of T is strictly less than what the surface area would be if each brick were also tiled with molecules, namely $500(4dq_M t + t^2) < p - 3w_A$. Furthermore, the height of a molecule (out of the plane of its boundary) is 7, so the projection of T onto the plane containing F_{floor} extends beyond H_{cage} by only seven units. In particular, this projection lies strictly in the interior of the bounding box of F_{floor} , and is at least $p - 7 > w_A$ units away from the top edge of F_{floor} .

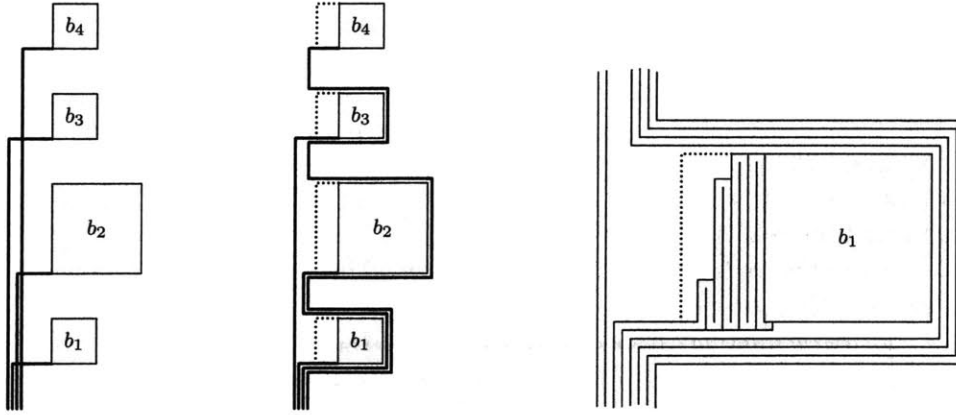
We will show in the next two subsections that $B(d, (a_1, \dots, a_n))$ has an edge unfolding if and only if $(d, (a_1, \dots, a_n))$ is a YES instance of UCSP. One direction is straightforward:

Lemma 5.4.2. *If $(d, (a_1, \dots, a_n))$ is a UCSP instance and $B(d, (a_1, \dots, a_n))$ has an edge unfolding, then $(d, (a_1, \dots, a_n))$ is a YES instance.*

Proof. Fix some unfolding of $B = B(d, (a_1, \dots, a_n))$. Let F_{pipe}^{1*} be the bottom height-1 subrectangle of F_{pipe}^1 , and similarly for F_{pipe}^{2*} , and consider the polyhedral surface B^* obtained by replacing F_{pipe}^1 and F_{pipe}^2 with F_{pipe}^{1*} and F_{pipe}^{2*} . The unfolding of B induces an unfolding of B^* . In this unfolding of B^* , all of $T \cup \{F_{\text{pipe}}^{1*}, F_{\text{pipe}}^{2*}\}$ unfolds into $H_{\text{cage}} \cup H_{\text{drain}}$: indeed, p was chosen to ensure that the surface area of $T \cup \{F_{\text{pipe}}^{1*}, F_{\text{pipe}}^{2*}\}$ is strictly less than p , so there is not enough material to reach the top of H_{drain} . It follows that each brick b_i unfolds into $H_{\text{cage}} \cup H_{\text{drain}}$, and since H_{drain} is too narrow for the bricks, each b_i unfolds into H_{cage} . So there exists a packing of the b_i (with side-lengths $a_i \cdot q$) into H_{cage} (with side-length $d \cdot q$), which proves the Lemma. \square

5.4.3 Wiring the Tower

Think of T_{bottom} as a grid of molecules, with origin $(0, 0)$ at its lower left corner. In this section we demonstrate how to connect each brick to the bottom-left corner of the tower by a chain of molecules. For convenience, we ensure all such chains have the same length, L .



(a) In the first step, (b) Wires are mod- (c) Finally, each wire W_i is modified with zig-
overlapping wires are ified with detours zags in the empty $a_{\sigma(i)}q_M/2 \times q_M/2$ grid next
drawn from u_i to $v_{\sigma(i)}$ around bricks to avoid to brick $b_{\sigma(i)}$ in order to bring its length up to
with just one right turn. intersections. exactly $L = 4ndq_M$.

Figure 5-5: The three steps in the construction of molecule wires W_i of Lemma 5.4.3. Each wire W_i connects molecule u_i to $v_{\sigma(i)}$, where u_1, \dots, u_n are lined up along the bottom edge from left to right, and each molecule v_i is just under the lower left corner of brick b_i . The figures correspond to $\sigma(1) = 3$, $\sigma(2) = 1$, $\sigma(3) = 2$, and $\sigma(4) = 4$.

Brick b_i is positioned at $(q, y_i; q)$ where $y_i = i + a_1 + \dots + a_{i-1}$. Let f_i ($1 \leq i \leq n$) be the lower edge of b_i , oriented left-to-right, and let f_0 be the lower edge of T_{bottom} , also oriented left-to-right. For $1 \leq i \leq n$ define $u_i = T_{\text{bottom}}[6i, 0]_{w_M}$; these are lined along the left of f_0 in T_{bottom} , spaced 6 molecules apart. Also let $v_i = T_{\text{bottom}}[q, y_i - 1]_{w_M}$ be the molecule just under the lower left corner of b_i in T_{bottom} .

Lemma 5.4.3. *For any permutation σ of $\{1, \dots, n\}$, there exist n non-overlapping wires W_1, \dots, W_n of molecules in T_{bottom} such that each wire W_i has length exactly $L = 4ndq_M$ and connects $c_0(W_i) = u_i$ to $c_L(W_i) = v_{\sigma(i)}$, with starting and ending edges along f_0 and $f_{\sigma(i)}$ respectively. Furthermore, no wire touches the two leftmost columns of molecules on T_{bottom} , and finally, the complement of the bricks b_i and wires W_i in T_{bottom} forms a single edge-connected polyomino of molecules.*

Proof. Provisionally define each W_i as the wire that goes straight up from u_i and turns right to $v_{\sigma(i)}$, as in Figure 5-5a. As defined, these wires may intersect: the horizontal segment of W_i hits the vertical segment of W_j when $i < j$ but $\sigma(i) > \sigma(j)$. To fix these, for each i , take all wires W_j that hit the horizontal part of W_i and insert a detour around brick b_i as illustrated in Figure 5-5b, keeping a 1-molecule gap between two detouring wires, and between these wires and b_i . Because $q_M > 4n$, there is ample room for the detours. Before the detours, each wire had length less than

$$t + q_M = (n + 2 + a_1 + \dots + a_n) q_M \leq 2ndq_M,$$

and each of fewer than n detours adds at most $2dq_M$ molecules, so the total length of each W_i is less than $4ndq_M$. Furthermore, by the parity of the positions of u_i and $v_{\sigma(i)}$, W_i has even length.

We now bring the length of each W_i up to exactly $L = 4ndq_M$. The $a_i q_M \times \frac{q_M}{2}$ grid of molecules to the left of b_i is empty, its bottom edge is adjacent to wire W_i , and the top

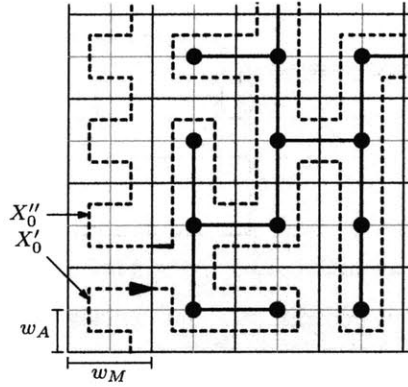


Figure 5-6: A closeup of the bottom-left corner of T_{bottom} illustrating how to write G as a wire of atoms as in the proof of Lemma 5.4.4. This wire, X_0 , is formed by traversing the four atoms of wire X'_0 , then walking around the spanning tree S , and finally following atom-wire X''_0 .

and left edges are not adjacent to any wires. This grid has even width $q_M/2$, and by zig-zagging up and down in this region as shown in Figure 5-5c, we can add any even number of molecules up to $a_i q_M^2/2 > L$. This indeed allows each wire to reach its destination with total length exactly L . Finally, the left two columns of molecules were not touched by the wires, and the 1-molecule gaps inserted above ensure that the complement of the bricks and wires remains connected. \square

Now think of T (partially unfolded as in Figure 5-4b) as a grid of atoms, not molecules. Edges f_0, f_1, \dots, f_n are as defined above, and let g be the bottom edge of $F_{\text{pipe}}^1 \cup F_{\text{pipe}}^2$ of length $3w_A$, oriented left to right.

Lemma 5.4.4. *It is possible to write T as an interior-disjoint union of the following pieces:*

- bricks b_1, \dots, b_n ,
- wires X_1, \dots, X_n of atoms where each X_i has length exactly $4L$ and connects the bottom-right corner of molecule u_i (with starting edge along f_0) to the top-right corner of molecule $v_{\sigma(i)}$ (with ending edge along $f_{\sigma(i)}$), and
- a wire X_0 connecting $(T_{\text{bottom}}[1, 0]_{w_A}, f_0[1]_{w_A})$ to edge $g[1]_{w_A}$ along with its adjacent atom on T_{top} .

Proof. Let wires of molecules W_i be as in Lemma 5.4.3. Wire X_i is obtained from W_i by starting at the bottom-right atom in molecule u_i and ensuring that $c_{4k}(X_i), \dots, c_{4k+3}(X_i)$ are the four atoms in molecule $c_k(W_i)$ for each $0 \leq k \leq L-1$. This can be done uniquely, and by parity, this wire X_i will terminate at the top-right atom of molecule $v_{\sigma(i)}$.

It remains to construct X_0 . Let W_0 be the wire of molecules in T that starts at $T_{\text{bottom}}[0, 0]_{w_M}$ and traces the left edge of T_{bottom} , the bottom, left, and top edges of T_{left} , and the left edge of T_{top} up to its top-left corner. Let X'_0 be the length-four wire of atoms that traces $c_0(W_0)$ as in Figure 5-6, and define X''_0 as the wire of atoms that follows the rest of W_0 as in the Figure. Let G be the region of T outside of the bricks b_1, \dots, b_n and wires $X_1, \dots, X_n, X'_0, X''_0$; by Lemma 5.4.3, G forms a connected polyomino of molecules.

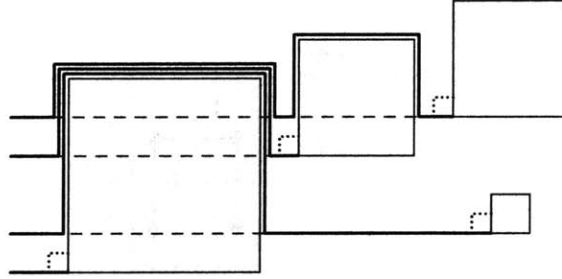


Figure 5-7: The bricks in a unique-coordinate square packing are attached to the bottom-left corner of the cage via non-overlapping wires of flatoms, as follows: First draw wires directly up and right. Then resolve wire-brick overlaps by detours as shown here. Finally, make each wire zig-zag in the small dotted region to bring its length up to $4L$.

Lemma 5.4.3 guarantees that molecules $G[1, 0]_{w_M}$ and $G[1, 1]_{w_M}$ are in G , so pick any spanning tree S of the molecules in G in which these two molecules are connected. The desired wire of atoms X_0 is obtained by traversing X'_0 , walking all the way around S to the starting edge of X''_0 , and then following X''_0 . \square

5.4.4 Unfolding Surface B

We are now able to prove the converse of Lemma 5.4.3:

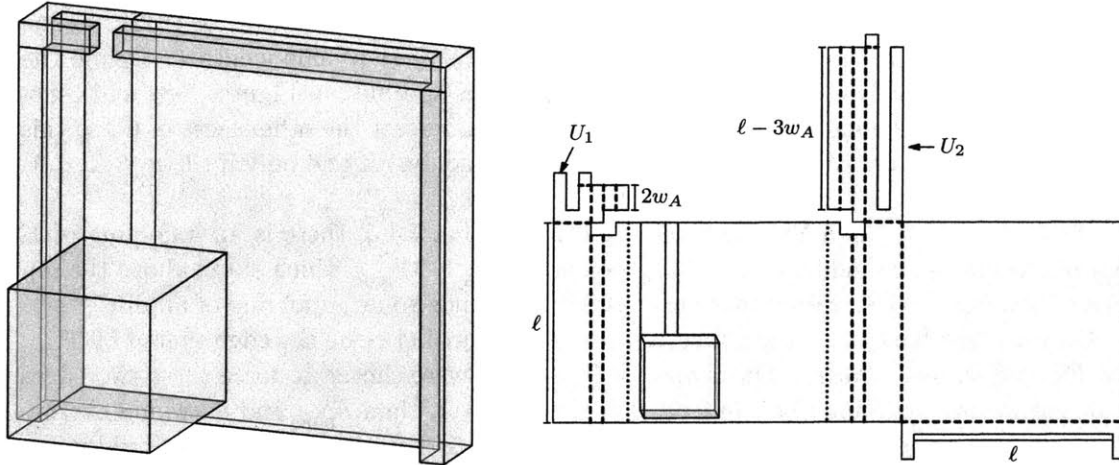
Lemma 5.4.5. *If $(d, (a_1, \dots, a_n))$ is a YES instance of UCSP, then $B(d, (a_1, \dots, a_n))$ has an edge unfolding.*

Proof. Think of H_{cage} as a $dq_F \times dq_F$ grid of flatoms, where $q_F = q/w_F$, with origin in the lower-left corner. Let f_0 and f_i ($1 \leq i \leq n$) be the bottom edges of H_{cage} and brick b_i respectively, as above. Pick a packing of squares with side-lengths a_1, \dots, a_n into $[0, d]^2$ with all the guarantees of the YES-promise of UCSP, and say the i^{th} square is positioned at (x_i, y_i) . Scale this up to a packing of bricks b_i into H_{cage} , with b_i positioned at $(x_i q_F w_F, y_i q_F w_F)$. Since the bricks b_i do not meet each other or the edges of H_{cage} , there is at least a q_F -flatom separation between them. For $1 \leq i \leq n$, define the flatoms $h_i = H_{\text{cage}}[4i, 0]_{w_F}$ (along f_0) and $k_i = H_{\text{cage}}[x_i q_F, y_i q_F - 1]_{w_F}$ (just under edge f_i).

Since the coordinates y_1, \dots, y_n are all different, let σ be the permutation so that $y_{\sigma(1)} > y_{\sigma(2)} > \dots > y_{\sigma(n)}$. We will create wires Z_1, \dots, Z_n of flatoms in $H_{\text{cage}} \setminus \bigcup_{i=1}^n b_i$ where wire Z_i connects flatom h_i with its bottom edge to flatom $k_{\sigma(i)}$ with its top edge, using a method very similar to the proof of Lemma 5.4.3. Initially, let Z_i be the wire of flatoms that goes straight up from h_i then turns right to meet $k_{\sigma(i)}$. These do not intersect each other by choice of σ , but they may intersect some of the bricks. So for each brick b_i , take all wires Z_j crossing b_i and insert detours around the top of b_i as in Figure 5-7; since $q_F/2 > n$, there is more than enough room. Now all the wires and bricks are interior-disjoint.

Initially, each wire had length less than $2dq_F$, and each detour adds at most $2dq_F$ length, so each Z_i has total length less than $2dq_F + 2dnq_F < 4dnq_F < 4L$. The $q_F/2 \times q_F/2$ grid of flatoms adjacent to Z_i and the left edge of $b_{\sigma(i)}$ is empty, and since this has even width $q_F/2$ and has $q_F^2/4 > 4L$ flatoms, wire W_i may zig-zag up and down in this grid (as in the proof of Lemma 5.4.3) to increase its length to exactly $4L$.

Now we can describe the unfolding of $B = B(d, (a_1, \dots, a_n))$. Using permutation σ defined here, apply Lemma 5.4.4 to B to obtain $n + 1$ wires of atoms X_0, X_1, \dots, X_n . Each



(a) The faces U_1 and U_2 must be cut away from B in any unfolding of C in order to avoid overlapping F_{floor} , F_{pipe}^1 , or F_{pipe}^2 .

(b) A partial unfolding of C showing that any unfolding of B extends to an unfolding of C .

Figure 5-8: The polyhedron $C = C(d, (a_1, \dots, a_n))$ without boundary.

brick b_i will unfold to its position $(x_i q, y_i q)$ in the UCSP unfolding above. For each $1 \leq i \leq n$, wires X_i and Z_i were designed so that their initial edges are centered on the same unit-length segment along f_0 :

$$e_0(X_i)[4]_1 = f_0[108i + 13]_1 = e_0(Z_i)[13]_1,$$

and similarly their final edges are centered in the same place on $f_{\sigma(i)}$:

$$e_{4L}(X_i)[4]_1 = f_{\sigma(i)}[13]_1 = e_{4L}(Z_i)[13]_1.$$

Furthermore, wires X_i and Z_i have the same length, $4L$, and each has an even number of left and right turns because their initial and final edges are parallel. It is thus possible, by Lemma 5.4.1, to unfold wire X_i into the region of H_{cage} described by Z_i while keeping X_i connected to both F_{floor} and $b_{\sigma(i)}$ along edges $f_0[108i + 13]_1$ and $f_{\sigma(i)}[13]_1$ respectively.

It remains to describe the unfolding of X_0 , F_{pipe}^1 and F_{pipe}^2 . In H_{cage} , the wires Z_1, \dots, Z_n do not intersect leftmost column of modules, so define Z_0 as the wire of flats in $H_{\text{cage}} \cup H_{\text{drain}}$ that starts at $H_{\text{cage}}[0, 0]_{w_F}$ with its bottom edge and proceeds straight up into H_{drain} with a total length equal to the length of X_0 . By Lemma 5.4.1, we may unfold X_0 into Z_0 while keeping the center of its initial edge connected to F_{floor} at $f_0[13]_1$ and the center of its final edge connected to $F_{\text{pipe}}^1 \cup F_{\text{pipe}}^2$ along $g[13]_1$. In the unfolding, therefore, $F_{\text{pipe}}^1 \cup F_{\text{pipe}}^2$ simply slides up relative to F_{floor} and partially juts out of the top of H_{drain} . \square

5.5 Eliminating the Boundary

With the construction from the previous section, we are ready for the main result:

Theorem 5.5.1. *The Orthogonal Edge Unfolding problem is strongly NP-complete.*

Proof. This problem is in NP because any unfolding has integer coordinates and can thus be checked to be non-overlapping in polynomial time.

For hardness, we reduce from UCSP. For an instance $(d, (a_1, \dots, a_n))$ of UCSP, define $B = B(d, (a_1, \dots, a_n))$ as above, whose boundary is a square of side-length ℓ . Define the closed, orthogonal polyhedron $C = C(d, (a_1, \dots, a_n))$ as specified in Figures 5-8a and 5-8b. The tower T (and the molecules on the tower) do not intersect the other faces of C , so this is a simple polyhedron. We will show C has an edge unfolding if and only if $(d, (a_1, \dots, a_n))$ is a YES instance.

If $(d, (a_1, \dots, a_n))$ is a YES instance, then by Lemma 5.4.5, there is an unfolding of B that fits inside the bounding box of F_{floor} except for $F_{\text{pipe}}^1 \cup F_{\text{pipe}}^2$ which sticks above the top edge. Then Figure 5-8b shows that this unfolding extends to an unfolding of all of C .

On the other hand, suppose C has an edge unfolding. Let t_1 be the edge shared by F_{pipe}^1 and U_1 , and similarly for t_2 . The shapes of U_1 and U_2 were chosen to force these two edges to be cut in the unfolding of C : indeed, if t_1 were not cut, then F_{pipe}^1 and U_1 would overlap in the plane; the argument for t_2 is the same. It follows that the unfolding of C induces a connected unfolding of B , so by Lemma 5.4.2, $(d, (a_1, \dots, a_n))$ is a YES instance. \square

Bibliography

- [1] Timothy G. Abbott, Erik D. Demaine, and Blaise Gassend. A generalized carpenter’s rule theorem for self-touching linkages. *arXiv:0901.1322*, 2009.
- [2] Timothy Good Abbott. Generalizations of Kempe’s universality theorem. Master’s thesis, Massachusetts Institute of Technology, June 2008. Joint work with Reid W. Barton and Erik D. Demaine.
- [3] Zachary Abel and Erik D. Demaine. Edge-unfolding orthogonal polyhedra is strongly NP-complete. In *Proceedings of the 23rd Canadian Conference on Computational Geometry (CCCG)*, Toronto, Ontario, Canada, August 2011.
- [4] Zachary Abel, Erik D. Demaine, and Martin L. Demaine. A topologically convex vertex-unfoldable polyhedron. In *Proceedings of the 23rd Canadian Conference on Computational Geometry (CCCG)*, Toronto, Ontario, Canada, August 2011.
- [5] Zachary Abel, Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Anna Lubiw, André Schulz, Diane Souvaine, Giovanni Viglietta, and Andrew Winslow. Algorithms for designing pop-up cards. In *Proceedings of the 30th International Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 269–280, Kiel, Germany, February–March 2013.
- [6] Zachary Abel, Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Jayson Lynch, and Tao B Schardl. Who needs crossings? Hardness of plane graph rigidity. In *Proceedings of the 32nd International Symposium on Computational Geometry (SoCG)*, Boston, Massachusetts, June 2016.
- [7] Zachary Abel, Erik D. Demaine, Martin L. Demaine, Sarah Eisenstat, Jayson Lynch, and Tao B. Schardl. Who needs crossings? Hardness of plane graph rigidity. *Journal of Computational Geometry*, Special Issue of Selected Papers from SoCG 2016, Forthcoming.
- [8] Zachary Abel, Erik D. Demaine, Martin L. Demaine, Jin-ichi Itoh, Anna Lubiw, Chie Nara, and Joseph O’Rourke. Continuously flattening polyhedra using straight skeletons. In *Proceedings of the 30th annual symposium on Computational geometry (SoCG)*, page 396. ACM, 2014.
- [9] Herb Adams. *Chassis engineering*. Penguin, 1993. As cited by https://en.wikipedia.org/wiki/Watt%27s_linkage, accessed on September 8, 2016.
- [10] Oswin Aichholzer and Franz Aurenhammer. Straight skeletons for general polygonal figures in the plane. In *Proceedings of the 2nd Annual International Conference of*

- Computing and Combinatorics (COCOON)*, volume 1090 of *Lecture Notes in Computer Science*, pages 117–126, Hong Kong, June 1996. Springer.
- [11] Oswin Aichholzer, Franz Aurenhammer, David Alberts, and Bernd Gärtner. A novel type of skeleton for polygons. *Journal of Universal Computer Science*, 1(12):752–761, 1995.
 - [12] A. D. Alexandrov. *Convex Polyhedra*. Springer, 2005. First published in Russian, 1950.
 - [13] Kenneth Appel and Wolfgang Haken. Every planar map is four colorable, Part I: Discharging. *Illinois Journal of Mathematics*, 21:429–490, 1977.
 - [14] Brad Ballinger, David Charlton, Erik D. Demaine, Martin L. Demaine, John Iacono, Ching-Hao Liu, and Sheung-Hung Poon. Minimal locked trees. In *Algorithms and Data Structures*, pages 61–73. Springer, 2009.
 - [15] C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The Quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software*, 22(4):469–483, 1996.
 - [16] Gill Barequet, David Eppstein, Michael T. Goodrich, and Amir Vaxman. Straight skeletons of three-dimensional polyhedra. In *European Symposium on Algorithms (ESA 2008)*, volume 5193 of *Lecture Notes in Computer Science*, pages 148–160. Springer, 2008.
 - [17] Saugata Basu, Richard Pollack, and Marie-Francoise Roy. *Algorithms in real algebraic geometry*, volume 20033. Springer, 2005.
 - [18] Nadia Benbernou, Erik D. Demaine, Martin L. Demaine, and Aviv Ovadya. A universal crease pattern for folding orthogonal shapes. *arXiv:0909.5388*, 2009.
 - [19] Marshall Bern, Erik Demaine, David Eppstein, and Barry Hayes. A disk-packing algorithm for an origami magic trick. In *Proceedings of the 3rd International Meeting of Origami Science, Math, and Education (OSME)*, pages 17–28, 2002.
 - [20] Marshall Bern, Erik D. Demaine, David Eppstein, and Eric Kuo. Ununfoldable polyhedra. In *Proceedings of the 11th Canadian Conference on Computational Geometry (CCCG)*, Vancouver, Canada, August 1999.
 - [21] Marshall Bern, Erik D. Demaine, David Eppstein, Eric Kuo, Andrea Mantler, and Jack Snoeyink. Ununfoldable polyhedra with convex faces. *Computational Geometry: Theory and Applications*, 24(2):51–62, February 2003.
 - [22] Marshall Bern and Barry Hayes. Origami embedding of piecewise-linear two-manifolds. *Algorithmica*, 59(1):3–15, 2011.
 - [23] Therese Biedl, Erik D. Demaine, Martin L. Demaine, Anna Lubiw, Mark Overmars, Joseph O’Rourke, Steve Robbins, and Sue Whitesides. Unfolding some classes of orthogonal polyhedra. In *Proceedings of the 10th Canadian Conference on Computational Geometry (CCCG)*, Montréal, Canada, August 1998.
 - [24] Alexander I. Bobenko and Ivan Izvestiev. Alexandrov’s theorem, weighted Delaunay triangulations, and mixed volumes. *Annales de l’institut Fourier*, 58(2):447–506, 2008.

- [25] Jacek Bochnak, Michel Coste, and Marie-Françoise Roy. *Real algebraic geometry*, volume 36. Springer Science & Business Media, 2013.
- [26] Sergio Cabello, Erik D. Demaine, and Günter Rote. Planar embeddings of graphs with specified edge lengths. *Journal of Graph Algorithms and Applications*, 11(1):259–276, 2007.
- [27] John Canny. Some algebraic and geometric computations in PSPACE. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing (STOC)*, pages 460–469, Chicago, Illinois, May 1988.
- [28] Jason H. Cantarella, Erik D. Demaine, Hayley N. Iben, and James F. O’Brien. An energy-driven approach to linkage unfolding. In *Proceedings of the 20th Annual Symposium on Computational Geometry (SoCG)*, pages 134–143, Brooklyn, New York, June 2004. ACM.
- [29] Bernard Chazelle. An optimal convex hull algorithm in any fixed dimension. *Discrete & Computational Geometry*, 10(1):377–409, 1993.
- [30] Robert Connelly. The rigidity of certain cabled frameworks and the second-order rigidity of arbitrarily triangulated convex surfaces. *Advances in Mathematics*, 37(3):272–299, 1980.
- [31] Robert Connelly, Erik D. Demaine, Martin L. Demaine, Sándor P Fekete, Stefan Langerman, Joseph S. B. Mitchell, Ares Ribó, and Günter Rote. Locked and unlocked chains of planar shapes. *Discrete & Computational Geometry*, 44(2):439–462, 2010.
- [32] Robert Connelly, Erik D. Demaine, and Günter Rote. Straightening polygonal arcs and convexifying polygonal cycles. In *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science (FOCS)*, pages 432–442, Redondo Beach, California, November 2000. IEEE.
- [33] Robert Connelly, Erik D. Demaine, and Günter Rote. Infinitesimally locked self-touching linkages with applications to locked trees. *Contemporary Mathematics*, 304:287–312, 2002.
- [34] Robert Connelly, Erik D. Demaine, and Günter Rote. Straightening polygonal arcs and convexifying polygonal cycles. *Discrete and Computational Geometry*, 30:205–239, 2003.
- [35] Robert Connelly, Idzhad Sabitov, and Anke Walz. The bellows conjecture. *Contributions to Algebra and Geometry*, 38(1):1–10, 1997.
- [36] Bernard Dacorogna, Paolo Marcellini, and Emanuele Paolini. Lipschitz-continuous local isometric immersions: rigid maps and origami. *Journal de mathématiques pures et appliquées*, 90(1):66–81, 2008.
- [37] Erik D. Demaine, Martin L. Demaine, David A Huffman, Duks Koschitz, and Tomohiro Tachi. Characterization of curved creases and rulings: Design and analysis of lens tessellations. In Koryo Miura, Toshikazu Kawasaki, Tomohiro Tachi, Ryuhei Uehara, Robert J. Lang, and Patsy Wang-Iverson, editors, *Origami⁶*, pages 209–230. American Mathematical Society, December 2015.

- [38] Erik D. Demaine, Martin L. Demaine, and Jason Ku. Folding any orthogonal maze. In *Origami5: Fifth International Meeting of Origami Science, Mathematics, and Education*, pages 449–454, 2011.
- [39] Erik D. Demaine, Martin L. Demaine, and Anna Lubiw. Folding and cutting paper. In *Japanese Conference on Discrete and Computational Geometry (JCDCG)*, pages 104–118. Springer, 1998.
- [40] Erik D. Demaine, Satyan L. Devadoss, Joseph S. B. Mitchell, and Joseph O’Rourke. Continuous foldability of polygonal paper. In *CCCG*, pages 64–67, 2004.
- [41] Erik D. Demaine, David Eppstein, Adam Hesterberg, Hiro Ito, Anna Lubiw, Ryuhei Uehara, and Yushi Uno. Folding a paper strip to minimize thickness. *Journal of Discrete Algorithms*, 2015.
- [42] Erik D. Demaine, Stefan Langerman, and Joseph O’Rourke. Geometric restrictions on producible polygonal protein chains. In *International Symposium on Algorithms and Computation*, pages 395–404. Springer, 2003.
- [43] Erik D. Demaine and Joseph O’Rourke. *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*. Cambridge University Press, July 2007.
- [44] Erik D. Demaine and Tomohiro Tachi. *Origamizer: A practical algorithm for folding any polyhedron*, 2010.
- [45] Peter Eades and Nicholas C. Wormald. Fixed edge-length graph drawing is NP-hard. *Discrete Applied Mathematics*, 28(2):111–134, August 1990.
- [46] Martin Held. On computing Voronoi diagrams of convex polyhedra by means of wave-front propagation. In *Proceedings of the 6th Canadian Conference on Computational Geometry (CCCG)*, pages 128–133, Saskatoon, Saskatchewan, Canada, August 1994.
- [47] David A. Huffman. Curvature and creases: A primer on paper. *IEEE Trans. Computers*, 25(10):1010–1019, 1976.
- [48] Thomas C. Hull and sarah-marie belcastro. Modelling the folding of paper into three dimensions using affine transformations. *Linear Algebra and its applications*, 348(1):273–282, 2002.
- [49] Jin-ichi Itoh, Chie Nara, and Costin Vilcu. Continuous flattening of convex polyhedra. In *Computational Geometry*, volume 7579 of *Lecture Notes in Computer Science*, pages 85–97. Springer, 2012.
- [50] Gyo Taek Jin. Polygon indices and superbridge indices of torus knots and links. *Journal of Knot Theory and its Ramifications*, 6(02):281–289, 1997.
- [51] D. Jordan and M. Steiner. Configuration spaces of mechanical linkages. *Discrete & Computational Geometry*, 22:297–315, 1999.
- [52] J. Justin. Towards a mathematical theory of origami. In *Proceedings of the 2nd International Meeting of Origami Science*, Otsu, Japan, 1994.

- [53] Daniel Kane, Gregory N. Price, and Erik D. Demaine. A pseudopolynomial algorithm for Alexandrov’s Theorem. In *Proceedings of the 11th Algorithms and Data Structures Symposium (WADS)*, volume 5664 of *Lecture Notes in Computer Science*, pages 435–446, 2009.
- [54] Michael Kapovich and John J. Millson. Universality theorems for configuration spaces of planar linkages. *Topology*, 41(6):1051–1107, 2002.
- [55] Toshikazu Kawasaki. On high dimensional flat origamis. In *Proceedings of the First International Meeting of Origami Science and Technology*, pages 131–141, 1989.
- [56] Toshikazu Kawasaki. On the relation between mountain-creases and valley-creases of a flat origami. In *Proceedings of the 1st International Meeting of Origami Science and Technology*, pages 229–237, 1989.
- [57] A. B. Kempe. On a general method of describing plane curves of the n^{th} degree by linkwork. *Proceedings of the London Mathematical Society*, 7:213–216, 1876.
- [58] A. B. Kempe. *How to Draw a Straight Line: A Lecture on Linkages*. Macmillan and co., London, 1877.
- [59] A. B. Kempe. On the geographical problem of the four colours. *American Journal of Mathematics*, 2:183–200, 1879.
- [60] Henry C. King. Planar linkages and algebraic sets. *Turkish Journal of Mathematics*, 23(1):33–56, 1999.
- [61] M. Kirszbraun. Über die zusammenziehende und lipschitzsche transformationen. *Fundamenta Mathematicae*, 22(1):77–108, 1934.
- [62] Pascal Koiran. The complexity of local dimensions for constructible sets. *Journal of Complexity*, 16(1):311–323, March 2000.
- [63] Kaori Kuribayashi and Zhong You. Deployable stent, June 13 2006. US Patent 7,060,092.
- [64] Robert J. Lang. Airbag folding. <http://www.langorigami.com/article/airbag-folding>. Accessed: September 10, 2016.
- [65] Robert J. Lang. Treemaker 4.0: A program for origami design, 1998.
- [66] Robert J. Lang and Alex Bateman. Every spider web has a simple flat twist tessellation. In *Origami⁵: Proceedings of the 5th International Conference on Origami in Science, Mathematics and Education (OSME)*, pages 449–454. A K Peters, Singapore, July 2010.
- [67] Joseph Y-T. Leung, Tommy W. Tam, C. S. Wong, Gilbert H. Young, and Francis Y. L. Chin. Packing squares into a square. *Journal of Parallel and Distributed Computing*, 10(3):271–275, 1990.
- [68] Shuhei Miyashita, Steven Guitron, Marvin Ludersdorfer, Cynthia R. Sung, and Daniela Rus. An untethered miniature origami robot that self-folds, walks, swims, and degrades. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 1490–1496. IEEE, 2015.

- [69] N. E. Mnev. The universality theorems on the classification problem of configuration varieties and convex polytopes varieties. In Oleg Viro and Anatoly Vershik, editors, *Topology and Geometry—Rohlin Seminar*, volume 1346 of *Lecture Notes in Mathematics*, pages 527–543. Springer, 1988.
- [70] Mark Moll, David Schwarz, and Lydia E. Kavradi. Roadmap methods for protein folding. *Protein Structure Prediction*, pages 219–239, 2008.
- [71] Ares Ribó Mor. *Realization and counting problems for planar structures*. PhD thesis, Freie Universität Berlin, 2006.
- [72] Neil Robertson, Daniel Sanders, and Paul Seymour. The four-colour theorem. *Journal of Combinatorial Theory, Series B*, 70:2–44, 1997.
- [73] James B. Saxe. Embeddability of weighted graphs in k-space is strongly NP-hard. In *Proceedings of the 17th Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 1979.
- [74] Marcus Schaefer. Realizability of graphs and linkages. In János Pach, editor, *Thirty Essays on Geometric Graph Theory*, chapter 23. Springer, 2013.
- [75] Brigitte Servatius and Walter Whiteley. Constraining plane configurations in computer-aided design: Combinatorics of directions and lengths. *SIAM Journal on Discrete Mathematics*, 12(1):136–153, 1999.
- [76] G. C. Shephard. Convex polytopes with convex nets. *Mathematical Proceedings of the Cambridge Philosophical Society*, 78:389–403, 1975.
- [77] Peter W. Shor. Stretchability of pseudoline arrangements is NP-hard. *Applied Geometry and Discrete Mathematics: The Victor Klee Festschrift, DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, 4:531–554, 1991.
- [78] Ileana Streinu. A combinatorial approach to planar non-colliding robot arm motion planning. In *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science (FOCS)*, pages 443–453, Redondo Beach, California, November 2000. IEEE.
- [79] Ileana Streinu. Pseudo-triangulations, rigidity and motion planning. *Discrete and Computational Geometry*, 34(4):587–635, 2005.
- [80] Tomohiro Tachi. Generalization of rigid foldable quadrilateral mesh origami. In *Proceedings of the 50th Symposium of the International Association for Shell and Spatial Structures: Evolution and Trends in Design, Analysis and Construction of Shell and Spatial Structures*, València, Spain, 2009. Editorial Universitat Politècnica de València.
- [81] Tomohiro Tachi. Simulation of rigid origami. In *Origami*, volume 4, pages 175–187. A K Peters/CRC Press, 2009.
- [82] Tomohiro Tachi. Freeform rigid-foldable structure using bidirectionally flat-foldable planar quadrilateral mesh. *Advances in architectural geometry 2010*, pages 87–102, 2010.
- [83] Tomohiro Tachi. Geometric considerations for the design of rigid origami structures. In *Proceedings of the International Association for Shell and Spatial Structures Symposium (IASS)*, volume 12, pages 458–460, November 2010.

- [84] Walter Whiteley. Motions and stresses of projected polyhedra. *Structural Topology*, 7:13–38, 1982.
- [85] Shannon A. Zirbel, Robert J. Lang, Mark W. Thomson, Deborah A. Sigel, Phillip E. Walkemeyer, Brian P. Trease, Spencer P. Magleby, and Larry L. Howell. Accommodating thickness in origami-based deployable arrays. *Journal of Mechanical Design*, 135(11):111005, 2013.