ORIGAMETRY

Mathematical Methods in Paper Folding

THOMAS C. HULL

Origametry

Mathematical Methods in Paper Folding

THOMAS C. HULL

Western New England University

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108478724
DOI: 10.1017/9781108778633

First published 2021

Printed in the United Kingdom by TJ International Ltd, Padstow Cornwall

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Hull, Thomas, 1969– author.
Title: Origametry : mathematical methods in paper folding / Thomas Hull.
Description: New York : Cambridge University Press, 2020. | Includes index.
Identifiers: LCCN 2020022971 (print) | LCCN 2020022972 (ebook) |
ISBN 9781108478724 (hardback) | ISBN 9781108746113 (ebook)
Subjects: LCSH: Origami–Design. | Origami–Mathematics.
Classification: LCC TT870 .H85 2020 (print) | LCC TT870 (ebook) | DDC 736/.982–dc23
LC record available at https://lccn.loc.gov/2020022971
LC ebook record available at https://lccn.loc.gov/2020022972

ISBN 978-1-108-47872-4 Hardback
ISBN 978-1-108-74611-3 Paperback

Contents

Introduction

A Growing Interest

Origami is the art of paper folding, as is likely known by anyone picking up this book. Less known is how diverse origami is as an art form. Most people who have practiced origami have folded paper cranes, fishes, and frogs, or perhaps some of the many "playground" origami models that children teach to each other, like paper airplanes, fortune-tellers, and ninja throwing stars. Those who catch the paper-folding bug, however, learn how to fold dragons, insects, and octopi, each from an uncut square. Those who go further might see pictures like those in Figure 1, where two origami models, one rather simple and the other very complex, are shown with their respective **crease patterns** (the pattern of creases you would see if you were to unfold the model). Others might explore **origami tessellations**, which are origami models where the crease pattern forms a regular tiling of the plane. Others might become addicted to **modular origami**, where multiple, sometimes hundreds, of pieces of paper are all folded in the same way and then locked together to form beautiful polyhedral objects.

After staring at crease patterns and folding intricate geometric and representational origami models, one may start to suspect that there are mathematical rules at play in origami. Is there any inherent geometry to these crease patterns? Is there a way to predict into what shape they fold? Is there a limit to the complexity of shapes that can be folded? How would we make the previous question into a precise conjecture that could be proven?

On a somewhat different note, anyone interested in paper folding might have noticed a pronounced increase in the number of scientific news reports, viral web

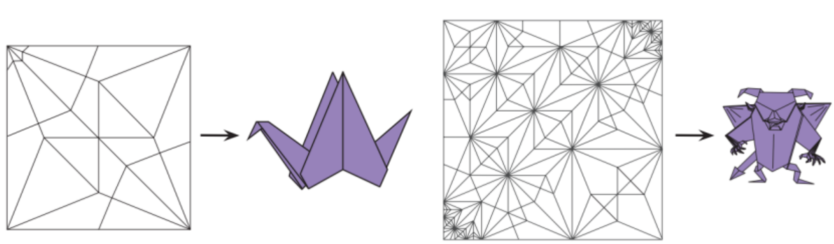

Figure 1 The classic flapping bird (crane) with its crease pattern, and Maekawa's Devil (Kasahara and Maekawa, 1983) with its crease pattern.

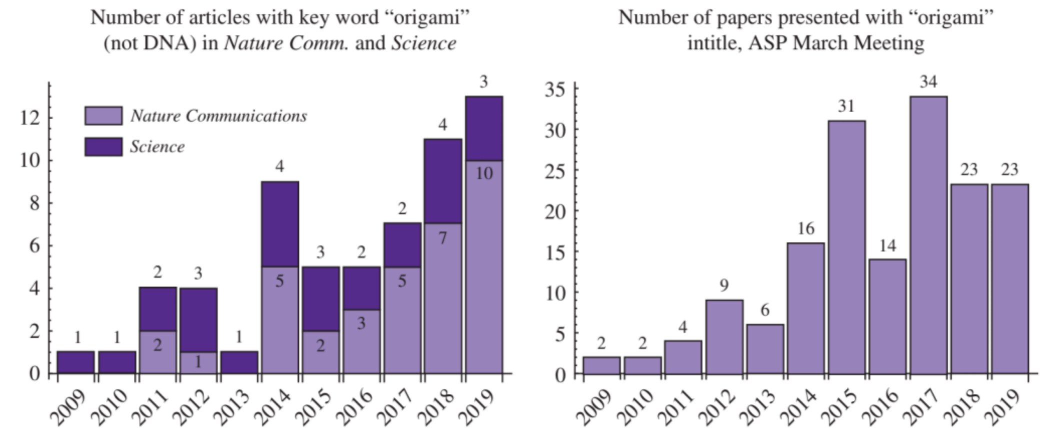

Figure 2 The rise of papers featuring origami applications in *Nature Communnications*, in *Science*, and at the American Physics Society's March Meeting.

videos, and research articles on the use of origami in science and engineering since the year 2013. The National Science Foundation (of the United States of America) offered about 12 large grants in each of the years 2012 and 2013 in their Emerging Frontiers in Research and Innovation program on "Origami Design for Integration of Self-assembling Systems for Engineering Innovation (ODISSEI)." Each of these four-year, approximately $2 million grants spurred a bevy of research in physics and engineering university departments across the country. Evidence of this growth in interest can be seen in the number of papers relating to origami applications in high-profile journals like *Science* and *Nature Communications*. (See Figure 2, where the numbers indicate papers with keyword "origami" excluding papers on DNA-origami, which has nothing to do with the folding of 2-dimensional sheets.)

At the same time, more and more sessions and papers at physics and engineering conferences have been devoted to origami applications since 2013 – far more than the NSF ODISSEI grants would generate on their own. See Figure 2 again for the number of papers with "origami" in their title from the American Physics Society's annual March Meeting over a period of ten years. Similar increases in origami applications in physics can be seen at engineering conferences. For example, at the ASME (American Society of Mechanical Engineers) 2019 Conference on Smart Materials, Adaptive Structures, and Intelligent Systems, there was one short course on origami mechanics, one keynote speaker with expertise in origami mechanics, one symposium of talks on "bioinspired adaptive origami systems," a best paper award given to Jakob Faber for the paper "Bioinspired spring origami," and 23 papers presented on origami research in mechanical engineering.

Why are so many scientists, including the National Science Foundation, excited about origami? Origami enthusiasts are not surprised; one very well-known origami model is called the **Miura-ori**, also known as the **Miura map fold**. It was invented by a Japanese astrophysicist named Koryo Miura who was searching for a way to send

As I continued to find mathematics papers that dealt with paper folding, I noticed two patterns emerging. One was that over and over again it seemed that researchers were reinventing the wheel, without knowledge of work in the area of mathematical paper folding that had come before. This is not too surprising. Before the days of the Internet, it was very hard (as I was learning while in graduate school!) to find each and every paper that had been published on a rather obscure topic, such as paper folding. But seeing this did make me think that someone needed to gather all these papers and unify the disparate work that had been done.

The second realization was cause for amazement. I was continually stunned by the number of different areas of mathematics – calculus, geometry, number theory, abstract algebra, differential topology, analysis – that had been applied to origami. Being a physical activity, it made sense that there would be many ways to model paper folding. But the sheer variety of different approaches one could take was surprising, exciting, and rather intimidating to a graduate student. As such, I found myself locating some papers on origami mathematics and thinking to myself, "I'm not ready to read this one, but perhaps in a few years I will be." For some papers, "a few years," turned out to be over a decade as the need to graduate, find a job, and learn how to be a professor took priority.

Yet the need was clear to me: a book needed to be written on the many mathematical methods that may be applied to the assorted aspects of origami. The diverse research that had been done needed to be unified, and the far-ranging approaches needed to be gathered in one place.

For me this became a labor of love. Discovering a paper like Stewart A. Robertson's "Isometric folding of Riemannian manifolds" (1977–1978) meant many joyous months of learning new results and connecting them with previously known (to me) work. Then, years later, stumbling upon a paper like Lawrence and Spingarn's "An intrinsic characterization of foldings of Euclidean space" (1989), which has so many similarities to Robertson's paper, meant more months of head-scratching as I tried to discern, reconcile, and then combine their two approaches.

Outline of the Book

The book you now hold in your hands is the result of such efforts. It took a long time to write. Much of Part I, on Geometric Constructions in origami, was first written in 2006 while on sabbatical leave from Merrimack College. Of course, in the intervening 14 years new research was done on this topic, and I've tried to keep those chapters updated. Still, in 2006 it was possible for me to contain most of everything known about origami geometric constructions in less then 100 pages. Now, if I wanted to include full proofs of all the important results in just this one aspect of origami mathematics, it would probably have to be a book of its own. Therefore I have had to cut some corners and refer readers to the literature for some of the more current results (such as Nishimura's proof (2013) that 2-fold origami can solve arbitrary quintic equations).

Part II is where I try to solidify an area of origami mathematics that I like to call the Combinatorial Geometry of Flat Origami. This concerns only origami models that lie in a flat plane when all the creases are folded. The previously mentioned theorems of Maeakawa and Kawasaki exemplify this area. Maekawa's Theorem states that the difference between the mountain and valley creases that meet at a vertex in a flat origami crease pattern must always be two. This is a combinatorics result, but its proof must at some level rely on the geometry (or, if viewed the right way, the topology) of the folded paper. Kawasaki's Theorem is a necessary and sufficient condition that determines if a collection of creases meeting at a vertex can fold flat, and this condition relies solely on the angles between the creases. Thus, this is a purely geometric result. However, Maekawa's and Kawasaki's Theorems are actually intertwined. We first see this in Chapter 5 when expanding our flat-folding perspective from planar paper to folding cones, and we see it again in Chapter 6 when looking through the lens of Justin's Theorem, a result on flat-foldable crease patterns with multiple vertices. The interplay of combinatorics and geometry is present in many aspects of mathematics, but origami provides a lovely example with enough results and depth to make it a separate sub-genre in origami mathematics.

In Part III I chose to gather three aspects of origami mathematics that correspond to three mathematical fields that all students learn in graduate school: algebra, analysis, and topology. For algebra we find a delightful origami homomorphism that relates, using only the geometry of a flat-foldable crease pattern, the symmetry group of a crease pattern with the symmetry group of the flat-folded model. For topology we dive deeply into the work of Stewart Robertson (and Lawrence and Spingarn) on the folding of Riemannian manifolds, in which we find surprising and elegant generalizations of Kawasaki's and Maekawa's Theorems. For analysis we describe more recent work of Dacorognna, Marcelini, and Paolini that uses high-dimensional isometric foldings (the n-dimensional analog to flat origami) to solve certain Dirichlet partial differential equations.

Part IV turns to Non-flat Folding, in particular rigid origami. This is the natural setting for many applications of origami in physics and engineering. The mathematical models needed to prove that the Miura-ori fold can open and close with the crease pattern's faces remaining rigid are found here, as are many, many results on rigid origami folding angle relationships, complexity, and configuration spaces. Rigid origami is a very active research area, and this part of the book only represents my take on the subject, as a mathematician. Luckily, one of the world experts on rigid origami, professor Tomohiro Tachi of the University of Tokyo, is writing a full monograph on the subject (Tachi, 2020).

Features of the Book

Throughout the book I have tried to provide a thorough set of useful references, although as applied origami has grown more and more popular, including a complete bibliography of all the current literature has become impossible. When applicable, I have included historical notes as a way to provide context to the many, and sometimes confusing, reference points in the literature.

Diversions There are many places in this book where details of a certain aspect or problem of origami mathematics are omitted. This was done because if absolutely all details of every proof or idea were included, then this book would have been twice as long and rejected by the publisher. In addition, there are many delightful examples and excursions in the various approaches to mathematical origami. It really is a shame that I could not include them all. Instead, I sprinkled many **diversions** throughout the book. Some of these are interesting, fun exercises left for the reader to explore. Others are meant to be parts of proofs that are fairly straightforward and therefore can be left to the reader (often with references to the literature for full details).

This book is a monograph, not a textbook (although it could be used as a text for a graduate-level or reading course), and therefore it didn't seem appropriate to end each chapter with a list of exercises (nor would this be appropriate for many parts of the book). Spreading diversions through the chapters seemed to be a good compromise for including many of the juicy tidbits of origami mathematics without overburdening the exposition.

Open Problems I have also provided, at the end of all appropriate chapters, a list of some of the open problems in the related fields of origami mathematics. This, of course, runs the risk of dating the book in a few years, as many of the topics covered in this text are currently areas of active research. But hopefully these problems will encourage readers to join the fun that is origami mathematics.

I also need to point out that not all aspects of origami mathematics are covered in this book. The field has simply grown too large for one monograph to cover. The largest omission is the topic of origami using curved creases. While this topic is touched on very briefly in Section 4.5, it is only done in the context of origami geometric constructions. Modeling folded structures made from curved creases is possible using differential geometry. This is a very interesting subject and very much an active area of research. In fact, new work on curved crease origami is being developed so quickly as to make planning a monograph-level exposition on the topic somewhat impossible, as it could become painfully dated before the book saw print. Indeed, work is being done at the time of this writing by several leading researchers on mathematical and computational origami toward writing a stand-alone text on curved creases.

Acknowledgements

Finally, there are many people and organizations that need to be thanked who helped in the completion of this book. Large sections of this book were completed during sabbatical leaves, first from Merrimack College and then from Western New England University. The latter, in fact, was spent enjoying the hospitality of the University of Tokyo, where several graduate students, in particular Akito Adachi, test-read chapter drafts of this book. Parts of this book were written while under grant support from the National Science Foundation, namely grants EFRI ODISSEI-1240441 and

DMS-1906202. I am indebted to the following individuals for extensive conversations or correspondence on all matters of origami mathematics over the past 10+ years: Zach Abel, Hugo Akitaya, Roger Alperin, Hannah Alpert, Bryan Gin-ge Chen, David Cox, Erik Demaine, Martin Demaine, David Eppstein, Arthur Evans, Johnna Farnham, Tomoko Fuse, Jacques Justin, Toshikazu Kawasaki, Jason Ku, Robert J. Lang, Anna Lubiw, Jun Maekawa, Jeannine Mosely, Michelle Normand, Jessica Ginepro Notestine, Joseph O'Rourke, Emanuele Paolini, Aubrey Rumbolt, Christian Santangelo, and Tomohiro Tachi. Last, but not least, this book absolutely benefitted by my partnership with sarah-marie belcastro, who has been both endlessly supportive and helpfully critical.

Part I

Geometric Constructions

Diversion 1.1 Starting with a square sheet of paper, fold it to produce a square having three-fourths its area. Only five folds are allowed.

The puzzle is referenced as coming from a book called *Mathematical Brain Benders* by Stephen Barr (1982). This puzzle is especially fun for origami practicioners who immediately conjecture that they can do it in fewer than five folds.

This challenge is similar to the following: Starting with a square piece of paper, fold it into a perfect equilateral triangle. To accomplish this, one would need to construct a 60° angle, which could be done by folding the sides of a 30°-60°-90° triangle in the square. In other words, we would need to construct line segments of length 1, 2, and $\sqrt{3}$ in our paper. It is standard, however, to always assume that our starting square has side length 1, so it would be more feasible to create a 30°-60°-90° triangle with side lengths 1/2, 1, and $\sqrt{3}/2$. Constructing $\sqrt{3}/2$ is exactly what we would need for the 3/4-area puzzle as well.

There are many ways to fold a 30°-60°-90° triangle in a square. In fact, it is not hard to find explicit methods for doing this in origami instruction books, especially books on modular origami (like (Fuse, 1990)), although such books usually do not mention that they are performing such a construction.

Figure 1.1 shows a standard method for producing such a 30°-60°-90° triangle. This can be shown synthetically, or we can merely note that if the square has side length 1, then *PB* must also have length 1, since it is the image of a side of the square under the fold. By symmetry, *AP* must also have length 1, and thus we have that the points *APB* form an equilateral triangle. (This immediately gives us that the fold made in Figure 1.1 produced the desired angles.)

A variation on this challenge is to fold an equilateral triangle of maximum area within our square piece of paper. Utilizing analytic techniques to discover what triangle orientation gives the maximal area can be a good exercise for calculus students (see (Hull, 2012)), but developing a folding method is another matter. Figure 1.2 shows the standard method of doing this as presented by Emily Gingras (Merrimack College class of 2003). The first picture is her "proof without words" that the angle θ shown is 15°, which proves that the other pictures give the proper equilateral triangle.

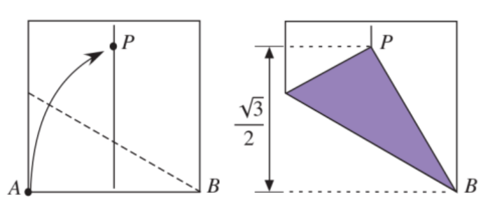

Figure 1.1 Producing a 30°-60°-90° triangle: First fold the square in half and unfold. Then fold the lower left corner up to the crease line, while making the crease go through the lower right corner.

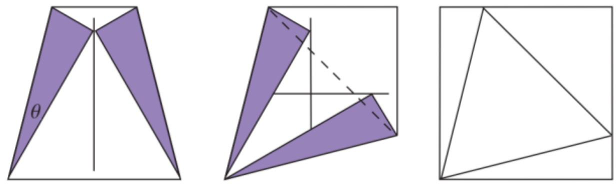

Figure 1.2 A "proof without words" for constructing the maximal equilateral triangle.

Both of these methods involve constructing a line segment whose length is an expression involving square roots, $\sqrt{3}/2$ in the first case and $2/\sqrt{2+\sqrt{3}}$ in the second. What kind of folding operation produced these lengths? In both cases we had a point being folded to a line (point A being folded to the half-way crease in Figure 1.1) where we also make sure that the crease passes through a second point (point B in Figure 1.1). This operation will be explored further in Chapter 2.

As an extra challenge, readers can try to use the method from Figure 1.2 to discover the classic method that paper folders use to fold a square into a regular hexagon with maximal area.

1.2 Dividing a Segment into 1/*n*ths

The problem of dividing the side of a piece of paper into n equal lengths is one which has been a favorite of origami geometry enthusiasts. References to it and various solutions for the cases where $n = 3$, 5, or 7 can be found in some origami books (see (Kasahara and Takahama, 1987; Kasahara, 1988)), on origami email lists, and on a variety of webpages. The challenge, of course, are the cases when n is odd, since folding lengths in half is simple and can generate all even numbers once the odds have been handled.

Such division methods have practical applications in origami as well. Many origami models start off by asking the folder to first divide the square into thirds or into a 5×5 grid. Interestingly, the most common method used by origamists to fold thirds is to use the method shown in Figure 1.3. The idea is to "eyeball" it by curving the paper into an S shape and easing the creases into their proper places. With practice this can

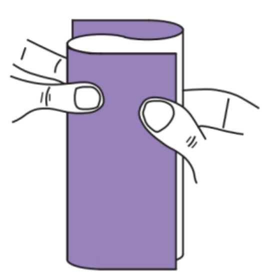

Figure 1.3 Folding thirds, the multifold method.

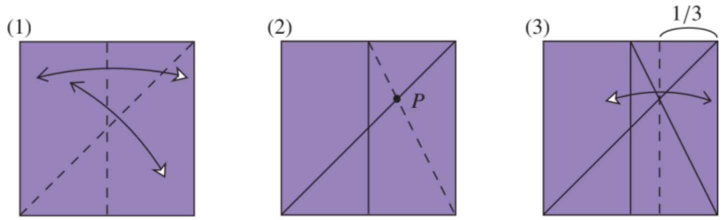

Figure 1.4 Folding thirds exactly. (1) Crease a diagonal and the 1/2 vertical crease. (2) Make a crease that connects the midpoint of the top edge and the lower right corner. Let this crease intersect the diagonal at P. (3) Make a crease at P perpendicular to the bottom and top sides. Then this last crease will be 1/3 from the right side.

be done very quickly and accurately, but it violates our rule of one fold at a time since it requires making two creases simultaneously (which is called a 2-fold or a multifold; these will be discussed in Chapter 4).

A mathematically precise, one-fold-at-a-time way to fold a square of paper into thirds is shown in Figure 1.4. While the origins of this method are unclear, Lang (1988) refers to it as the **crossing diagonals method**. The correctness of this method can be proven by similar triangles or by noticing that the point P is at the intersection of the lines $y = x$ and $y = -2x + 2$, where we assume that the square has side length 1 and lower left corner is at the origin. Thus $P = (2/3, 2/3)$.

This method can be generalized for arbitrary odd values of $n = 2k + 1$. Instead of making a vertical crease at the line $x = 1/2$, make it at $x = (2k - 1)/(2k)$. This is feasible because any odd factor of $2k$ is a smaller odd number than n. Thus, by induction, we can assume that dividing the side of our square into $1/(2k)$ths can be done. Then, using the same method as the 1/3 case, our point P would be at the intersection of the lines $y = x$ and $y = -2kx + 2k$, so $P = (2k/(2k + 1), 2k/(2k + 1))$, giving us a landmark for dividing the side into $1/n$ths.

The crossing diagonals method is sometimes used by origami designers; see John Montroll's Chess Board (Montroll, 1993), for example. However, since the method requires several crease lines to be made across the paper, it isn't viewed as ideal. A better method in this regard is based on **Haga's Theorem** (Kasahara and Takahama, 1987), which states that if we fold a corner of a square (or rectangular) sheet of paper to a point on a nonadjacent side, then several similar triangles can be found and the resulting crease can mark the sides of the paper at interesting lengths.

In particular, if we mark a point at $(1/(2k), 1)$ on the square and fold the lower left corner (the origin) to this point, as seen in Figure 1.5, then triangles A and B are similar. Also, triangle A is a right triangle and one leg, x, and the hypothenuse make up a side of the square, so the hypothenuse is $1 - x$. The Pythagorean Theorem then gives us that $x = (2k + 1)(2k - 1)/(8k^2)$. Letting the short leg of triangle B be y, the similarity relation gives us

$$\frac{y}{1/(2k)} = \frac{1 - 1/(2k)}{(2k + 1)(2k - 1)/(8k^2)},$$

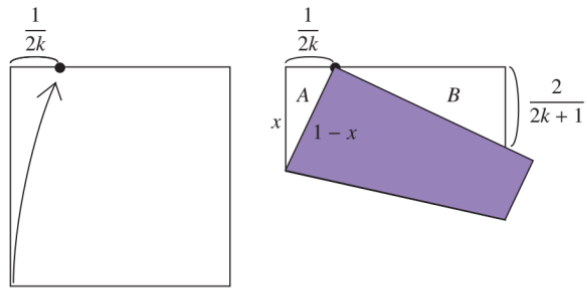

Figure 1.5 Haga's Theorem applied to the odd division problem.

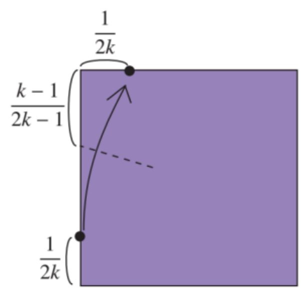

Figure 1.6 Noma's method.

which simplifies, amazingly enough, to $y = 2/(2k + 1)$. Thus if divisions of $1/n = 1/(2k + 1)$ are desired, constructing $1/(2k)$ and the one fold of Haga's Theorem will do the trick.

Diversion 1.2 (Geretschläger, 2002, 2008) Prove that the perimeter of triangle B in Figure 1.5 is always half the perimeter of the original square.

Haga's Theorem contains many other geometric morsels. See (Husimi and Husimi, 1979; Haga, 2002; Geretschläger, 2002) for more information.

However, it is possible to make any $1/n$ divisions along the side of a square without folding any creases all the way across the paper. The idea is to perform folds that only require making pinch marks on the perimeter of the paper. This would clearly be attractive for origami designers, making it possible to create any a/b mark on the perimeter without marring the paper's interior with extraneous creases.

Masamichi Noma (1992) developed such a method, and it is summarized in Figure 1.6. The idea is, if divisions of $1/n = 1/(2k - 1)$ are the goal, to make pinch marks at length $1/(2k)$ on the left side of the top edge of the square and at the bottom side of the left edge. This gives us two marked points on the paper's perimeter. If we fold these two points together, we can pinch the paper only on the left side, so as to avoid

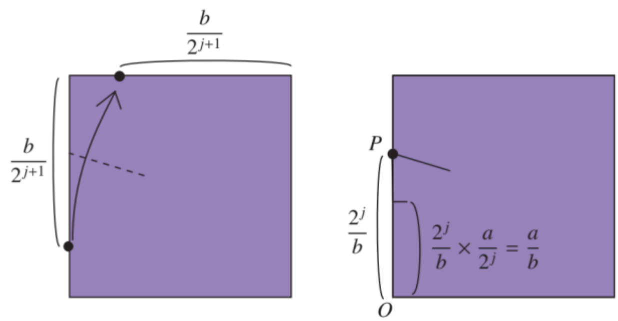

Figure 1.7 Noma's method used to construct a/b.

making a crease all the way across the square. This crease will intersect the left edge $(k-1)/(2k-1)$ from the top corner.

Diversion 1.3 Prove that Noma's method works.

Robert J. Lang has synthesized Noma's method, among others, to generate algorithms for producing folding sequences of pinch marks to create any rational length divisions. In (Lang, 2003) he suggests the following to apply Noma's method to create an arbitrary rational length a/b for integers $a < b$:

(1) Let 2^j be the largest power of 2 smaller than b.
(2) Construct lengths $b/2^{j+1}$ along the top and left sides of the square, as shown in Figure 1.7. (This is easy since the denomenators are just powers of 2.)
(3) Bring these two points together to make a crease pinch along the left side at point P.
(4) Then length OP (O being the lower left corner) will be $2^j/b$. (The same work needed in Figure 1.6 shows this.)
(5) Divide segment OP into $1/2^j$ths (which is easy). Taking a of these from O gives a length $(2^j/b)(a/2^j) = a/b$.

In all three of these division methods, none of the basic folding operations used are very complex. Each case involved only the "moves" of folding a crease between two existing points or folding one point onto another point. This is hardly surprising because only rational lengths were being constructed. But the variety and ingenuity of these methods are nonetheless a marvel.

1.3 Trisecting an Angle

The hallmark of origami geometric constructions has been the fact that paper folding can, fairly easily, trisect angles. The first known method for doing this was created by

(1) Let θ be an angle at a point O. Let P_1 be a point on one side of θ, and extend the line P_1O so that we can find a point P_2 on this line but on the other side of O so that $P_1O \cong P_2O$.
(2) Extend the other side of the angle θ to become the line L_1. Fold line L_2 to be perpendicular to L_1 at the point O. (This is a move we haven't seen before; it involves folding L_1 onto itself making the crease go through O.)
(3) Now fold P_1 onto L_1 and P_2 onto L_2 simultaneously to create line L_3.
(4) Finally, fold a crease perpendicular to L_3 that goes through point O. This line will make an angle of $\theta/3$ with one side of angle θ.

Diversion 1.4 Prove that Justin's trisection method works.

There is a lot of interesting mathematics to be explored in this "two points to two lines" origami move. Questions to ponder might include: Given any two points and two lines, can this operation always be performed? Does it always result in a unique crease? What kinds of numbers (segment lengths) is it constructing for us? We will address these questions in Chapters 2 and 3.

1.4 Folding a Regular Heptagon

The easiest regular polygons to fold from a square are, well, the square, regular octagon, 16-gon, and other 2^n-gons. We saw earlier that equilateral triangles are not too hard to fold from a square, and this admits regular hexagons and dodecagons relatively easily.

Diversion 1.5 Devise a way to fold a regular pentagon from a square piece of paper. (See (Morassi, 1989) for an analysis of approximate and exact methods for this.)

However, the smallest regular n-gon that origami can produce that SE&C cannot is the heptagon.

The first published instructions for folding a regular heptagon appears to be those of Scimemi (1989) and, independently, Geretschläger (1997b) (although Justin (1986b) also provides the basic ingredients for such a construction). Both of these methods are nearly identical, however, which is not surprising because both follow a classic algebraic approach to the heptagon problem, as can also be seen in the non-folding heptagon construction given by Gleason (1988). Gleason's construction assumes the tools of straightedge, compass, and an angle trisector. Since we know origami can trisect angles, Gleason's method could be used, as is, to fold a regular heptagon. Alperin (2002) did just this, but such a strategy produces a lengthy and inelegant folding

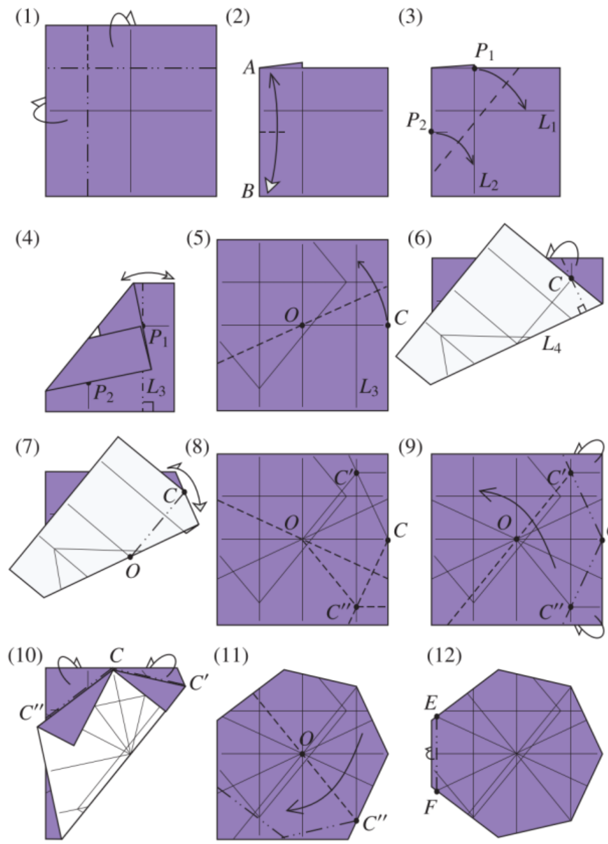

Figure 1.11 Folding a regular heptagon.

procedure. (We should note that mathematical purity demands that we find a folding procedure that is mathematically exact. However, when one physically makes a fold there will always be error present. So for practicality's sake it is always better to find folding sequences that help minimize error, either by being short or by encompassing folds that are easy to perform.)

Figure 1.11 shows a more cleaned-up way to fold a regular heptagon than those previously given. (Scimemi (1989) doesn't give an explicit folding sequence, and Geretschläger (1997b) has more folds than are necessary.) Our procedure is as follows:

(1) First, crease the paper in half from top to bottom and left to right. Then fold the top 1/4 behind and then the left 1/4 behind.
(2) Make a pinch crease on the left side by bringing points A and B together.
(3) Now we're ready for the fold that does the "magic." Fold point P_1 onto line L_1 and point P_2 onto line L_2 at the same time.

(4) Notice where P_1 went after step (3). Mountain fold a vertical crease, perpendicular to the bottom edge, on the underneath layer of paper, creating line L_3. Crease sharply and then unfold everything.
(5) Notice where L_3 is on the unfolded sheet. Fold C, the midpoint on the right side, to line L_3 so that the crease goes through the center O of the paper.
(6) Step (5) created the folded edge L_4. Fold the right flap of paper behind, making the crease go through C while being perpendicular to L_4.
(7) Now fold and unfold line OC. (This crease already exists, but you want it to be made through **all** layers of paper.) Then unfold everything.
(8) Line CC' is one side of our heptagon. (C' is the image of C under the fold in step (5).) Repeat steps (5)–(7) on the bottom half of the paper, creating point C''.
(9) Fold CC' and CC'' behind. Then valley fold OC', extending it across the paper.
(10) Use the images of CC' and CC'' to fold two more sides of our heptagon. Then unfold OC'.
(11) Repeat steps (9)–(10) on the bottom half.
(12) Fold the left side behind with crease EF to complete the heptagon.

To see why this works, let us set up a coordinate system for the paper as follows: let O, the center of the paper, be the origin, and let the side of the square be of length 4. (We choose these coordinates to more easily illustrate the connection with Gleason's analysis in (Gleason, 1988).) Our goal is to show that the point C' in Figure 1.11 has coordinates $(2\cos(2\pi/7), 2\sin(2\pi/7))$, and thus points C, C', and C'' form three vertices of a heptagon of radius 2. Since these points are used to generate the other vertices in a logical way, this would prove the folded heptagon's validity.

The points in step (3) of Figure 1.11 are $P_1 = (0, 1)$ and $P_2 = (-1, -1/2)$, where L_1 is the x-axis and L_2 is the y-axis. Suppose that P_1 gets folded to the point $P_1' = (t, 0)$ on L_1 and P_2 gets folded to the point $P_2' = (0, s)$ on L_2.

The segment P_1P_1' has slope $-1/t$, and the crease line in step (3) must be the perpendicular bisector to this segment. So the slope of the crease line must be t and pass through the midpoint of P_1P_1', which is $(t/2, 1/2)$. Thus one formula for the crease line in step (3) is

$$y = tx - \frac{t^2}{2} + \frac{1}{2}.$$

On the other hand, segment P_2P_2' has slope $(2s+1)/2$ and midpoint $(-1/2, (2s-1)/4)$. Thus, another formula for our crease line is

$$y = \frac{-2}{2s+1}x - \frac{1}{2s+1} + \frac{2s-1}{4}.$$

Our aim is to find the value of t (the x-coordinate of P_1'), since this determines the location of line L_3 in step (5) and thus the location of C'. Equating the slopes of our two line equations gives $s = -(t+2)/(2t)$. This can then be substituted into the equation we get by equating the constant terms of our two line equations, resulting in a single equation in t. After simplifying, this becomes

$$t^3 + t^2 - 2t - 1 = 0. \tag{1.1}$$

Sure enough, $t = 2\cos(2\pi/7)$ satisfies this equation, proving that L_3 is in the proper place. (The other roots of Equation (1.1) are real and negative, and thus are not values of t that would make the fold in step (3) of Figure 1.11 work.) For readers who do not immediately believe our claims as to the solutions of Equation (1.1), we present an argument from (Gleason, 1988).

Consider the vertices of a regular heptagon as the seventh roots of unity in the complex plane, that is, the complex solutions of $z^7 - 1 = 0$. Factoring out the obvious $z - 1$ term for the $z = 1$ corner, we get the equation for the remaining six corners: $z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0$. Let $A = \cos(2\pi/7) + i\sin(2\pi/7)$ (the principle seventh root of 1). Since the reciprocals of complex numbers on the unit circle are the same as the complex conjugates, we have that $1/A = A^6$, $1/A^2 = A^5$, and $1/A^3 = A^4$. Plugging these into our factored heptagon equation, we see that A satisfies

$$A^3 + A^2 + A + 1 + \frac{1}{A} + \frac{1}{A^2} + \frac{1}{A^3} = 0. \tag{1.2}$$

But we also have that $A + 1/A = A + \overline{A} = 2\cos(2\pi/7)$. Furthermore, notice that

$$A^2 + \frac{1}{A^2} = \left(A + \frac{1}{A}\right)^2 - 2 \quad \text{and} \quad A^3 + \frac{1}{A^3} = \left(A + \frac{1}{A}\right)^3 - 3\left(A + \frac{1}{A}\right).$$

Substituting these into Equation (1.2), we get

$$\left(A + \frac{1}{A}\right)^3 + \left(A + \frac{1}{A}\right)^2 - 2\left(A + \frac{1}{A}\right) - 1 = 0.$$

Therefore, $A + 1/A = 2\cos(2\pi/7)$ is a solution to Equation (1.1). Similar machinations show that other two roots of Equation (1.1) are $2\cos(4\pi/7)$ and $2\cos(6\pi/7)$.

Figure 1.12 provides a geometric interpretation of what's going on. If A is the principle root of $(z^7 - 1)/(z - 1)$, then $1/A$ is just A reflected about the real axis, and $(A + 1/A)/2$ is the midpoint of the segment connecting these two points. Since this midpoint is on the real axis, it's just the real part of A, which is $\cos(2\pi/7)$. The same thing holds for the other roots A^2 and A^3. Therefore the roots of $(z^7 - 1)/(z - 1)$ after the substitution $z = A + 1/A$ will just be twice the real parts of the seventh roots of unity, excluding $z = 1$. There are only three such numbers, and the equation after substitution simply becomes Equation (1.1).

In fact, Equation (1.1) is the standard equation one tries to solve when confronting the regular heptagon construction problem. (See (Martin, 1998), for example.) In a very real sense, this equation is being solved in step (3) of the folding sequence. It is interesting to note that both the folding methods of Scimemi (1989) and of Geretschläger (1997b) incorporate basically the same fold seen in step (3) of Figure 1.11 to create a crease with the proper slope for the heptagon construction. That two independent researchers came upon the same fold to solve Equation (1.1) is not a coincidence – they were both trying to solve the same equation via folding. We will see in Chapter 2 how to go about solving general cubic equations with such folding operations.

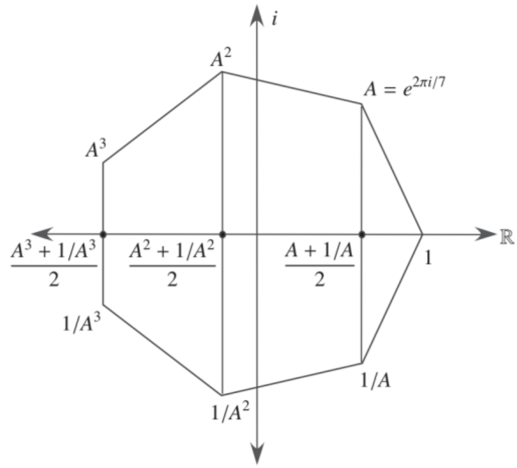

Figure 1.12 Geometric interpretation of $A + 1/A$ and the like.

1.5 The Basic Origami Operations

Now that we have seen several examples of origami constructions, we are in a better position to consider classifying the **basic origami operations**, or **BOOs** for short. This turns out to be more problematic than one might expect.

As seen at the beginning of this chapter, it is easy to classify what operations are possible under SE&C because we know exactly what our tools can do. The examples we've seen show that origami admits many different types of operations. When trying to make a list of them, it is not clear if one is really a special case of another, or whether we have found them all. For example, in the 1980s Huzita and Scimemi (1989) developed the following list of operations for origami:

O1: Given two points P_1 and P_2, we can fold a crease line connecting them.
O2: Given two lines, we can locate their point of intersection, if it exists.
O3: Given two points P_1 and P_2, we can fold the point P_1 onto P_2 (perpendicular bisector).
O4: Given two lines L_1 and L_2, we can fold the line L_1 onto the line L_2 (angle bisector).
O5: Given a point P and a line L, we can make a fold line perpendicular to L passing through the point P (perpendicular through a point).
O6: Given two points P_1 and P_2 and a line L, we can, whenever possible, make a fold that places P_1 onto the line L and passes through the point P_2.
O7: Given two points P_1 and P_2 and two lines L_1 and L_2, we can, whenever possible, make a fold that places P_1 onto L_1 and also places P_2 onto L_2.

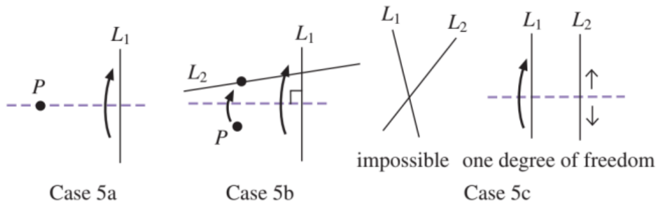

Figure 1.16 The subcases of folding a line L_1 to itself.

Case 3c: Also fold another point P_2 to another line L_2. This is operation O7.
We can now see that we've covered all seven of our BOOs. To make sure there are no more, however, we need to look at the last case.

Case 5: *Fold a line L_1 to itself.* Folding a line L_1 to itself is really just folding a crease perpendicular to L_1. But since this crease could intersect L_1 at any point, this gives us a degree of freedom. Combining with other operations gives the following (as shown in Figure 1.16):

Case 5a: Also fold a point P to itself. This gives us operation O5 again.

Case 5b: Also fold a point P to another line L_2. This gives us O8 again.

Case 5c: Also fold another line L_2 to itself. If these two lines intersect, then this is impossible, since the resulting crease would have to be perpendicular to both. Thus the two lines would have to be parallel, but this still results in one degree of freedom. Thus this combination is redundant—we would still need another operation to specify a unique crease, bringing us back to Case 5a or 5b.

This exhausts all the possibilities of folding points and lines to points and lines, completing the proof. □

Operations O1–O8 encompass everything that straight-crease, single-fold origami can do. This list does contain redundancies, however, and to eliminate them we need to be more specific about what is given to us at the start of our constructions.

For example, Alperin (2000) assumes the paper to be the entire complex plane with the given constructed points 0 and 1. He then chooses operations O1–O4, O6, and O7 to be his list of construction operations. (Actually, Alperin and several other writers use the word "axiom" to refer to allowed folds. But since O6 and O7 are not always possible, "operations" seems a more appropriate term.) One could equivalently assume that the four points $(\pm 1, \pm 1)$ are given, which, when the lines $y = \pm 1$ and $x = \pm 1$ are folded using O1, simulate the boundary of a square piece of paper.

Certainly the given points we start with and the operations O1–O8 guarantee that every crease line that is constructed will have a constructed point on it somewhere, as well as every point having a constructed line passing through it. This observation has led several people, including Martin (1998) and Hatori (2003) to prove that all we really need to characterize origami constructions are operations O2 and O7.

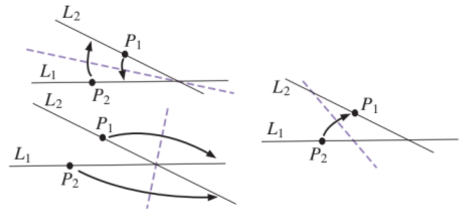

Figure 1.17 Generating O3 and O4 from O7.

Theorem 1.2 *Assuming that we are given at least two constructed points contained in nonparallel constructed lines (which may be identical), then any straight-crease, single-fold origami construction from this starting set can be completely described by combinations of operations O2 and O7.*

Proof By Theorem 1.1, we know that operations O1–O8 are all we need to consider. We obviously need to keep O2, but the others can be shown to be special cases of O7 where either we have $L_1 = L_2$ or some of the points P_1, P_2 lie on the lines L_1, L_2. This needs to be done carefully.

For the operations O3 and O4, let P_1 be on L_2 and P_2 be on L_1. (We know that all constructed lines contain constructed points and vice versa, so we can assume this from the premises of O3 and O4.) Then there will be at most three ways in which P_1 and P_2 can be folded onto L_1 and L_2, respectively, as shown in Figure 1.17. The first two cases will amount to folding L_1 onto L_2, or bisecting one of the angles made at their intersection, producing O4. (If L_1 and L_2 are parallel, then there will be only one way to do this.) The other case folds P_1 onto P_2, giving us O3.

In O6, we are given points P_1, P_2 and line L_1. Let L_2 be any line through P_2. So long as O6 is possible, we can have O7 fold P_2 to itself on line L_2 and fold P_1 onto L_1. (Note that we are not concerned here with when O6 is possible – we will take that up in the next chapter.)

For O1 and O5, let L_1 and L_2 be lines containing P_1 and P_2, respectively. Then, assuming L_1 and L_2 are not parallel, there are at most three different folds we could make that will leave P_1 on L_1 and P_2 on L_2: (a) making the crease pass through P_1 and P_2, (b) making the crease pass through P_1 and be perpendicular to L_2, and (c) making the crease through P_2 and perpendicular to L_1. While some of these may be identical, case (a) is operation O1. Cases (b) and (c) would give us O5. If L_1 and L_2 do happen to be parallel, then P_1 and P_2 could not have been the original two points in the construction, so there exist other points and lines that we can use to construct a different line through one of P_1 or P_2.

For O8 we can take an arbitrary point P_2 on L_2 and use O7 to fold P_1 onto L_1 while folding P_2 to a (probably) different place on L_2 to ensure that the crease line will be perpendicular to L_2. This covers all the operations O1–O8. □

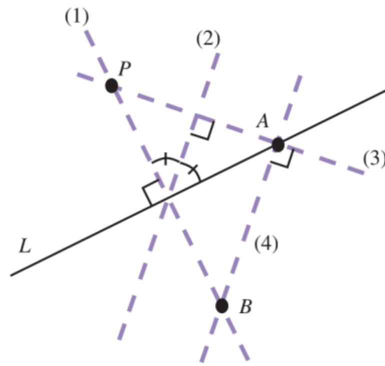

Figure 1.18 Constructing the reflection of P about L.

Remark 1.3 A common concern regarding the list of BOOs O1–O8 is whether or not we unfold the paper after each operation. Geometrically, all we care about is that each BOO creates a new line in the plane and that the intersection of lines creates new points. From this view, it does seem that every time we perform a BOO, we unfold the paper immediately to consider the new line formed in the paper (plane).

However, in practice origamists often leave the paper folded in order to perform an operation using the folded layers of paper as a guide. This was done several times in the heptagon construction in Section 1.4, for example. It seems conceivable that repeated applications of a BOO, especially O7, on a piece of paper without unfolding it might lead to a construction that could not be achieved by performing one BOO at a time, unfolding after each.

Yet this is not the case. The locations of any points or lines that are moved in the process of folding the paper flat with a BOO can be reconstructed if we immediately unfold the paper. (And therefore there is no need to keep the paper folded.) To prove this, suppose that L is a crease line constructed by some BOO. All we need to do is show that for any point P or other line L', we can construct the reflections of P and L' about L.

One method for constructing the reflection of P about L is shown in Figure 1.18. First use O5 to crease a line (1) perpendicular to L passing through P. Then use O4 to fold this crease line onto L, bisecting the angle between them to make the crease (2). Then use O5 again with P and the crease (2), labeling A as the point where this crease (3) intersects L. Finally, perform O5 again with the crease line (3) and the point A to make the crease (4). Where this crease intersects the crease line (1), called B in Figure 1.18, is the reflection of P about L.

Constructing the reflection of a line L' about L can be handled similarly. All we need is a point or two constructed on L'; by reflecting these about L, we can then create the reflection of L' by using O1.

This does, of course, use the convention that we can think of our sheet of paper as being as large as we wish and that the boundary lines of the paper are just constructed lines like any other. In any case, we conclude that while in practice it is often much more efficient to leave the paper folded after performing an origami operation, we can keep our list of BOOs simply to O1–O8 (or O2 and O7, if we want to be really efficient) by unfolding the paper after each step without losing any origami construction power.

Some of the subtleties in the work of this section, especially Theorem 1.1, seem to be ignored by much of the literature in this area. Several researchers (see (Alperin, 2000), for example) create definitions of what it means to be origami constructible (as we will in Chapter 3) referring to the basic origami operations mentioned here, but they do so with no argument as to whether more operations might be possible. Other papers, for example (Auckly and Cleveland, 1995) and parts of (Alperin, 2000), ask what can be constructed by a deliberately reduced set of origami operations. Investigators in the origami community were very concerned with the question of whether more operations existed (see (Hull, 1996)), whereby Lang's proof of Theorem 1.1 is viewed as a breakthrough.

In some sense it doesn't matter which of the moves O1–O8 we choose for an official list of basic origami operations. As we will see in the next chapter, it is O7 that separates origami from SE&C constructions, and being able to reference the other operations makes for very convenient notation.

1.6 Historical Remarks

The first person to seriously analyze origami geometric constructions seems to have been T. Sundara Row (1901). The first person to introduce operation O7 seems to have been Margherita Beloch (1936). (See the remarks in Section 2.5 for more information.) However, none of these early researchers made a formal list of possible origami construction operations. The first such list to see print seems to have been in Jacques Justin's 1986 paper (Justin, 1986b). Justin states that his list was inspired by an unpublished list created by Peter Messer (1984). Messer's list contains the operations O1–O7, but not O8. Justin's list contains all O1–O8. It appears that Scimemi independently developed a list of origami construction operations, which became those listed in (Huzita and Scimemi, 1989) and only included O1–O7.

Complicating things further, George Martin published a paper in 1985 (Martin, 1985) that defines origami constructions using only operations O2 and O7, and he seems to have developed this without any knowledge of Beloch's work. Martin cites Row and a publication of Yates, but he also cites (Dayoub and Lott, 1977), which describes how one can use a Mira, a geometric construction tool that allows one to reflect points about a line in the same way origami does, to trisect an arbitrary angle. In doing so Dayoub and Lott use the Mira to perform an operation very similar to O7 in origami. Martin then refines this method for the Mira, publishing his own paper on Mira constructions that contains exactly the form of operation O7 where the two given lines are perpendicular, whereupon Martin proves that the Mira can be used to construct cube roots (Martin, 1979). So it could be that Martin was inspired by the Mira to devise his version of operation O7 for origami.

Both Justin's paper (Justin, 1986b) and the Huzita–Scimemi paper (Huzita and Scimemi, 1989) were published in hard-to-find publications, or at least in periodicals that weren't indexed in the standard mathematical abstracts. This is likely why subsequent work, like (Auckly and Cleveland, 1995; Geretschläger, 1997a; Alperin, 2000; Hatori, 2003) make no mention of Messer, Justin, or Scimemi (or Beloch, for that matter).

This will not have a solution when the discriminant is less than zero: $x^2/4 - y < 0$. The boundary of this region is our parabola $y = x^2/4$, and all points above this curve are "bad" choices for P_2.

Seeing that operation O6 generates tangents to a parabola, one naturally wonders if other conic section tangents can be folded. The answer is yes, but before describing how, we'll give an informal motivation.

One can obtain a parabola from an ellipse in the following way: Recall that an ellipse is the set of all points P such that the sum of the distances between P and two fixed foci equals a constant called the major axis of the ellipse. Fix a focus of the ellipse as well as the vertex point closest to that focus and let the other focus slide off to infinity. The result will be a parabola. This process is equivalent to altering the angle at which a plane intersects a cone to go from an elliptical intersection to a parabolic one. (This fact was known as far back as Kepler (1604). See also (Hilbert and Cohn-Vossen, 1956, pp. 3–4).)

Now consider an ellipse with foci F_1 and F_2 determined by the constant distance (major axis) r. Draw a circle C with center F_2 and radius r. This circle has some useful properties, described in the following theorem.

Theorem 2.2 *Using the ellipse and circle described previously, let P be a point on the ellipse and Q be the point where the extended ray F_2P meets the circle C. Then*

(i) the shortest distance between P and a point on the circle C is given by the segment PQ, and
(ii) the tangent line to the ellipse at P reflects F_1 onto the point Q.

Proof For (i), suppose that a different point $Q' \neq Q$ on C has a length PQ' less than that of PQ. Then we would have $|F_2P| + |PQ'| < r = |F_2Q|$, which violates the triangle inequality.

For (ii), notice that since $|F_2Q| = |F_2P| + |F_1P| = r$, we have $|F_1P| = |PQ|$. Also, a property of tangents to an ellipse is that they form equal angles with the lines from the point of tangency to the two foci (see (Hilbert and Cohn-Vossen, 1956, p. 4)). In the notation of Figure 2.2, this means that $\angle F_1PT_1 = \angle F_2PT_2$. But $\angle F_2PT_2$ and $\angle T_1PQ$ are vertical angles and thus are equal. Therefore $\angle F_1PT_1 = \angle T_1PQ$,

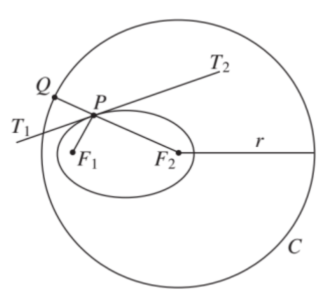

Figure 2.2 An ellipse and its folding circle.

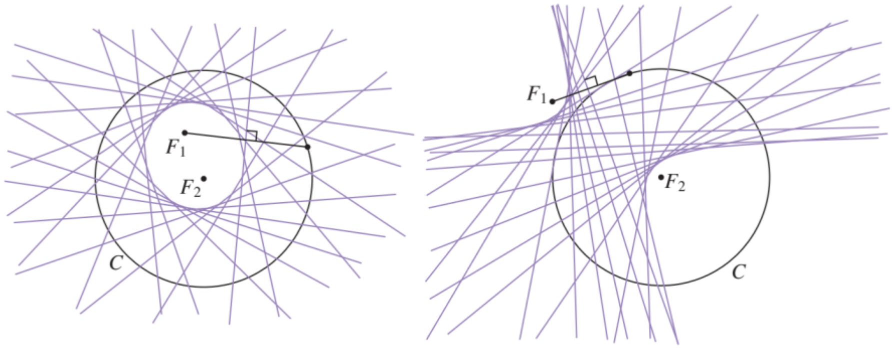

Figure 2.3 Folding tangents to an ellipse and a hyperbola.

which means that the tangent at P bisects $\angle F_1PQ$. This combined with $|F_1P| = |PQ|$ settles (ii). □

We now ask what happens to the circle C as we drag the focus F_2 to infinity, leaving the focus F_1 as well as the ellipse vertex closest to F_1 fixed. As the major axis of the ellipse grows, so will the radius of the circle, but one point of the circle, the point closest to F_1 where the circle intersects the line containing the major axis, will also remain fixed. Thus as F_2 moves farther and farther away, the circle will appear more and more like a straight line near the focus F_1. In the limit, we'll have that F_1 will be the focus of a parabola and the circle C will have turned into the directrix of that parabola.

In other words, and in light of Theorem 2.2, the circle C and focus F_1 play that same role for the ellipse as the directrix and focus play when folding tangents to a parabola. Thus if we draw a circle C with center F_2 on our paper, and pick a point F_1 inside C, then repeatedly folding F_1 onto C will generate the tangents to the ellipse with foci F_1 and F_2 and major axis constant equal to the radius of C, as shown in Figure 2.3 on the left. (In practice this is easier to do if one starts with a circular piece of paper, picks a point inside at random, and then folds the paper's circular boundary to the point over and over again.)

Similarly, the tangents to a hyperbola can be folded if we choose the focus F_1 to be **outside** the circle C. See the right side of Figure 2.3. In fact, a hyperbolic version of Theorem 2.2 can be proven using exactly the same argument as for the ellipse.

Diversion 2.1 Derive the equations for an ellipse or hyperbola generated by these folding methods. For example, prove that if we fold the point $(1, 0)$ over and over again to the circle $(x + 1)^2 + y^2 = 16$, then the crease lines will be tangent to an ellipse (and find the equation of this ellipse). Hint: Parameterize the circle and then use the substitutions $\sin\theta = \frac{2t}{1+t^2}$ and $\cos\theta = \frac{1-t^2}{1+t^2}$ to obtain a trigonometry-less parameterization.

We have seen how folding a point to a line generates tangents whose envelope is a parabola and folding a point to a circle generates tangents whose envelope is an ellipse or hyperbola (or circle, if the point is the circle's center). In general, we can let C be a curve in the plane and ask what curve will be the envelope of the crease lines made when folding a fixed point to C repeatedly. Rupp (1924) addressed this question and showed that this process generates a copy of the negative pedal of the curve C scaled down by a factor of 1/2.

The **negative pedal** of a curve C is determined as follows: Given a fixed point O (called the **pedal point**) and a point P on the curve, let L_P be the line through P that is perpendicular to OP. The envelope of the lines L_P over all $P \in C$ is the negative pedal curve of C (Lockwood, 1967). To see why this relates to our operation of folding a point O repeatedly to a curve C, notice that if $P \in C$ is the point to which we fold O, the locus of the midpoints M of OP (over all $P \in C$) forms a copy of the curve C dilated toward the point O by a factor of 1/2. The crease lines we fold would be perpendicular to the segments OM, and thus their envelope would form the negative pedal curve of this smaller copy of C. Therefore all the classic results for negative pedal curves would apply (Lockwood, 1967, pp. 156–159). For example, if C were a cardioid and O the cusp point, then folding O to C repeatedly will create creases tangent to a circle.

However, the major result from all this is that operation O6 is generating tangent lines to a parabola. That is, when we fold a point P to a line L, the resulting crease will be tangent to the parabola with focus P and directrix L. This hints at a connection between operation O6 and solving quadratic equations, which will be discussed in Section 2.2.

This also gives us a way to interpret origami operation O7, where we simultaneously fold two points to two lines. Each point folding to a line determines the focus and directrix of a parabola, and the crease line made will be tangent to both. Thus operation O7 is finding a common tangent to two parabolas. The implications of this will be explored in Section 2.3.

We end this section with some comments on the literature available for folding conic tangents. T. Sundara Row (1901, Note 235) seems to be the first to call attention to the fact that folding a point to a line produces a parabola tangent, although the proof given there is not the most elegant. Virtually all of the later publications, however, cite Row's work as their key influence. Lotka (1907) expanded Row's discovery to the case of folding a point to a circle, providing a very nice analytical method that simultaneously proves that the resulting envelopes are ellipses and hyperbolas. (His approach is similar to the one given here for the parabola, where we used the discriminant of a well-chosen quadratic to determine which points will not be hit by any crease lines.) Rupp (1924) seems to be the first to successfully analyze what happens when we fold a point to a general curve. Yates (1943), citing Row's and Lotka's work, provides proofs for the parabola, ellipse, and hyperbola that are the most concise and elegant that the author has seen.

Other publications have emphasized the educational potential of the origami-conic connection. The NCTM booklet *Paper Folding for the Mathematics Class*

(Johnson, 1957) presents minimal details and incomplete proofs. (Fehlen, 1975) presents the material through pedagogy in a style that leaves the proofs incomplete. (Bruckheimer and Hershkowitz, 1977) gives a very straightforward proof of the parabola case. (Scher, 1996) gives incomplete proofs but calls attention to how these folding activities can be dramatically modeled with computer software like Geometer's Sketchpad or GeoGebra. (Smith, 2003) offers alternate analytic proofs but doesn't cite any previous work. The author tried his hand at specifically framing this material for the classroom (Hull, 2012, Activity 6).

There are undoubtedly other references on operation O6 (we chose to not mention references that only state the folding exercise, without any attempt at a proof). In fact, it seems that numerous times people have completely re-invented the wheel on this topic. For example, during the years 1970–2000 there were at least four articles in *Mathematics Teacher* on folding tangents to conics, none of which reference each other or the previous work listed above. Despite the inherent simplicity and charm of exploring conics via operation O6, this has been, as Yates puts it, "unfortunately relegated to the limbo" (Yates, 1943, p. 230).

2.2 Solving Second Degree Equations

At this point it is fairly evident that origami should be able to solve second degree equations. The previous section provides good evidence for this. Also, the operations of a straightedge and compass seem to be mimicked by origami. Origami operation O1 simulates a straightedge, and O6 can be thought of as doing the work of a compass as follows: since the point P_2 lies on the crease line, it can be thought of as remaining fixed, like the center of a circle. Then P_1 is folded onto the line L, and in doing so the length of the segment P_1P_2 is preserved as P_1 is moved to its new location. Thus the image of P_1 under folding operation O6 is the same as finding a point of intersection between the line L and a circle centered at P_2 with radius $|P_1P_2|$. This provides the intuition, at least, that origami constructions should be able to do everything that SE&C can do, and since the field of numbers constructible under SE&C is precisely those that are solutions to quadratic equations, origami should be able to find such solutions as well.

Actually doing this is another matter. As we shall see in Chapter 3, where we construct the field of origami constructible numbers, it is relatively easy to, given origami constructible nonzero lengths a and b, construct any combination of these lengths using the operations addition, subtraction, multiplication, division, and square roots. Therefore the output of the quadratic formula can be constructed, giving us any real solutions to quadratics with rational coefficients.

However, in light of our work in Section 2.1, the fact that we can fold tangents to conics is too tempting to not apply to this problem.

Suppose that we want to find roots, if they exist, of $y = x^2 + ax + b$, where $a, b \in \mathbb{Q}$. By locating the focus and directrix of this parabola, we can try to pick a well-chosen tangent to fold and solve our problem. After enacting a change of coordinates to translate the parabola's vertex to the origin, we can use standard methods to see that the focus will be at the point $P_1 = (-a/2, b - a^2/4 + 1/4)$ and the directrix line L will be $y = b - a^2/4 - 1/4$. By the same methods we used in Sections 1.4 and 2.1, we find that if we fold P_1 to an arbitrary point $(t, b - a^2/4 - 1/4)$ on L, then the resulting crease line will have equation $y = (2t + a)x - t^2 + b$. If we take $t = (-a + \sqrt{a^2 - 4b})/2$, then the crease will be tangent to the parabola at a root, and the intersection of the crease and the x-axis (which we assume is a constructed line) will locate the root for us. Substituting this value of t into our line equation gives us the rather horrendous line equation

$$y = \sqrt{a^2 - 4b}x + \frac{a}{2}\sqrt{a^2 - 4b} + 2b - \frac{a^2}{2}.$$

But now we see that if we let $x = -a/2$ on this line, the square root terms will cancel and give us $y = 2b - a^2/2$. Therefore we can pick $P_2 = (-a/2, 2b - a^2/2)$, which is easy to construct from the coefficients of our original quadratic, and perform origami operation O6 to fold P_1 onto L making the crease go through P_2; then, the intersection of the resulting crease and the x-axis will be a root.

The other root could have been found by taking $t = (-a - \sqrt{a^2 - 4b})/2$. Of course, if the vertex of the parabola, $(-a/2, b - a^2/4)$, is above the x-axis, then no real roots exist and the results of this origami method will be meaningless.

Figure 2.4 shows how this works to find a root of $x^2 - x - 1 = 0$. We pick $P_1 = (1/2, -1)$, $P_2 = (1/2, -5/2)$, and L the line $y = -3/2$ and perform operation O6.

Diversion 2.2 Use the example in Figure 2.4 to solve Diversion 1.5 (constructing a regular pentagon via origami). The fact that $\cos(2\pi/5) = (1/2)/((1 + \sqrt{5})/2)$ might help.

Many other folding methods for solving quadratics exist, and readers are encouraged to discover their own. In fact, one way to develop origami constructions

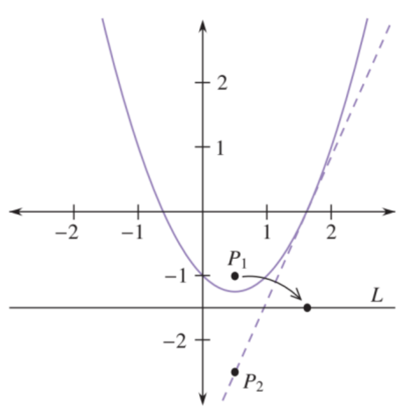

Figure 2.4 Folding to find a root of $y = x^2 - x - 1$.

We then construct an $(n-1)$-order path beginning at O and ending at T as follows: Position our turtle back at O, and this time point it at a yet-to-be-determined angle $-90^\circ < \theta < 90^\circ$ from the previous a_n line. Let the turtle march in this direction until it comes to the infinite line containing the a_{n-1} line segment. (Note that depending on θ, the turtle might not hit the actual a_{n-1} segment, so the extended line may be needed.) Then rotate the turtle by $\pm 90^\circ$ so that it faces the next a_{n-2} line. Let it march again until it hits the line containing a_{n-2}. Then rotate by a right angle again and repeat this process. In the end the turtle will hit the line containing the a_0 segment, and now we get to see what our choice of θ should have been. Our aim is to choose a value θ that makes the turtle arrive in the end at point T. (This is illustrated by the dotted line paths in Figure 2.6.) If such a value of θ exists, then it will lead us to a solution of our original polynomial (2.2).

Theorem 2.3 (Lill, 1867) *In the above process, the value* $x = -\tan\theta$ *will be a root of* $a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$.

Proof Since the turtle turns 90° at each corner, the n-order path and the $(n-1)$-order path form a sequence of right triangles all of which have an angle θ in common and are thus similar. Label the corners of the n-order path $P_n, P_{n-1}, \ldots, P_1$ and the corners of the $(n-1)$-order path $Q_{n-1}, Q_{n-2}, \ldots, Q_1$. An example where $n = 5$ is shown in Figure 2.7.

Let $x = -\tan\theta$ and consider the length of the side opposite θ in each of these similar triangles, P_iQ_{i-1} for $i = n, \ldots, 2$ and P_1T. We obtain

$$|P_nQ_{n-1}| = -a_n x,\ |P_{n-1}Q_{n-2}| = -x(a_{n-1} - |P_nQ_{n-1}|) = -x(a_{n-1} + a_n x),$$
$$|P_{n-2}Q_{n-3}| = -x(a_{n-2} - |P_{n-1}Q_{n-2}|) = -x(a_{n-2} + x(a_{n-1} + a_n x)), \ldots$$
$$|P_1T| = -x(a_1 + x(a_2 + x(a_3 + \cdots + x(a_{n-1} + a_n x)\cdots))).$$

Since $|P_1T| = a_0$, we have that $a_0 = -a_1x - a_2x^2 - a_3x^3 - \cdots - a_{n-1}x^{n-1} - a_nx^n$. Therefore $x = -\tan\theta$ is a solution to Equation (2.2). □

Finding the proper angle θ to make this work is certainly not something that normal construction tools (compass, marked straightedge, what-have-you) can handle in general. As seen in Figure 2.5 in the previous section, the case where $n = 2$ can easily be modified for SE&C, but Lill did offer another suggestion for handling the cases

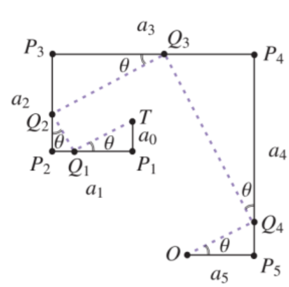

Figure 2.7 Proving Lill's method works in a case where $n = 5$.

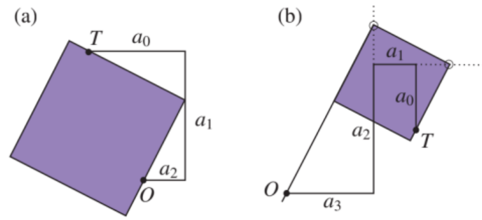

Figure 2.8 Lill's use of transparent graph paper to find a solution square for the (a) $n = 2$ and (b) $n = 3$ cases.

where $n = 2$ or 3. He proposed, after drawing the initial 2- or 3-order path, overlaying it with translucent graph paper (with a fine mesh) and rotating it until the points O and T both lie on sides of a square (or on lines that contain a side of the square) and some of this square's corners lie on lines that contain a segment a_i. For the $n = 2$ case, we would want one corner of the square to lie on the a_1 segment (or a line that contains a_1); see Figure 2.8(a). In the cubic $n = 3$ case we would want one corner of the square to lie on the a_1 line and an adjacent corner to lie on the a_2 line. We'd also want, for the cubic case, the points O and T to lie on lines containing opposite sides of the square. See Figure 2.8(b) for an example. We call this square, in the proper position in the plane, the **solution square** for Lill's method in the $n = 2$ or $n = 3$ cases.

Of course, Lill's method won't always work. If no real solutions exist for the polynomial, then no angle θ will work, and no turning or positioning of our translucent graph paper will produce a solution square.

Diversion 2.4 Prove that the $(n - 1)$-order path generated by Lill's method to solve Equation (2.2) is a scalar multiple of the path made by the degree-$(n - 1)$ polynomial obtained by factoring $(x + \tan\theta)$ from Equation (2.2). In this way, Lill's method represents geometrically the process of successively factoring out the real roots from a polynomial.

It was the cubic case of Lill's method that Beloch discovered could be performed by paper folding. She states in (Beloch, 1936) that she developed this work while teaching a course on geometry, where she no doubt presented Lill's method, at the University of Ferarra, Italy, in the 1933–34 academic year. Beloch's technique is illustrated in Figure 2.9 and proceeds as follows: Given a cubic $a_3x^3 + a_2x^2 + a_1x + a_0$ with rational coefficients, we construct the 3-order path according to Lill's method, and our aim is to find the proper angle θ, if it exists, to give us a 2-order path connecting O and T. First construct a line D_1 at $x = 2a_3$ and another line D_2 at $y = a_2 + a_0$.

Theorem 2.4 (Beloch, 1936) *Using the above notation, if the origami operation O7, folding the point O to the line D_1 and the point T to the line D_2, is possible, then the line segment that the extended a_2 and a_1 lines cut out of the resulting crease will be one side of the solution square for Lill's method.*

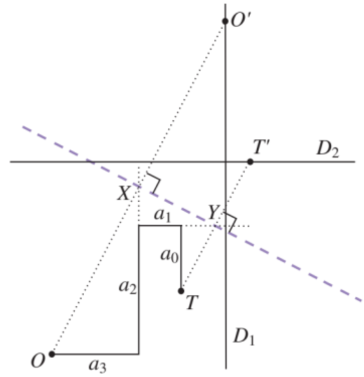

Figure 2.9 Beloch's use of operation O7 to solve a cubic.

Proof The proof is very simple once one understands the reasoning behind the location of the lines D_1 and D_2. Suppose that we wanted to consider a parabola with focus $O = (0, 0)$ and vertex $P_3 = (a_3, 0)$. Then the directrix of this parabola would be the line $x = 2a_3$, which is D_1. Therefore, any fold that places O onto D_1 will be tangent to this parabola. Furthermore, if O' is the image of O on D_1 after the folding, then the midpoint X of OO' will be on the line containing the a_2 segment.

Similarly, consider the parabola with focus $T = (a_3 - a_1, a_2 - a_0)$ and vertex $P_1 = (a_3 - a_1, a_2)$. The directrix of this parabola will be the line $y = a_2 + a_0$, which is D_2. When we fold T onto a point T' on line D_2, the resulting crease will be tangent to this second parabola, and the midpoint Y of TT' will lie on the line containing the a_1 segment. See Figure 2.9.

When we fold O onto D_1 and T onto D_2 using operation O7, we have that the points X and Y will be on this crease line. Thus the segment XY is perpendicular to OX and TY, and thus XY is the side of a square with two adjacent corners (the points X and Y) on the a_2 and a_1 lines. Two (extended) opposite sides of this square contain the points O and T, and therefore this is a solution square for Lill's method. □

Diversion 2.5 Determine the operation O7 fold needed, and the 3-order path it generates for Lill's method, to find one of the three real solutions of $x^3 - 7x - 6 = 0$ using Beloch's approach. (Note that the lack of an a_2 term doesn't change the procedure; the a_2 segment will merely have zero length and the "line containing a_2" will still be perpendicular to the a_3 and a_1 lines. See (Riaz, 1962) or (Hull, 2012) for hints.)

It seems that Beloch's work on origami constructions laid in obscurity for many years. While attention was drawn to it in the early years of the origami mathematics community (see (Huzita and Scimemi, 1989) in particular), the methods of trisecting angles by Abe and Justin and Messer's cube root of two construction (Messer, 1986) were done without knowledge of Beloch's work. Furthermore, the reference

(Huzita and Scimemi, 1989) was in a very hard-to-find publication, and the more recent constructions of (Alperin, 2000) and (Hatori, 2003) for solving general cubics were made without any knowledge of Beloch. It is fascinating, then, to learn that both of these modern solutions are virtually identical to Beloch's approach, although without the reference to Lill.

In Hatori's construction (Hatori, 2003) we begin with an arbitrary cubic $x^3 + ax^2 + bx + c = 0$ and consider the parabola with focus $P_1 = (a, 1)$ and directrix $L_1 : \ y = -1$ and the parabola with focus $P_2 = (c, b)$ and directrix $L_2 : \ x = -c$.

Diversion 2.6 Prove that in Hatori's construction the slope of the crease line we obtain (if it exists) by folding P_1 onto L_1 and P_2 onto L_2 using origami operation O7 will be a real solution to $x^3 + ax^2 + bx + c = 0$.

Figure 2.10 shows how Hatori's construction can be interpreted as equivalent to Beloch's. Indeed, if we reflect this picture about the line $y = -x$ and translate P_1 to the origin, it is exactly what Beloch's application of Lill's method would generate.

Alperin's construction (Alperin, 2000) begins with the equivalent form of a general cubic $x^3 + ax + b = 0$. (See (Hull, 2012, Activity 6) or (Weisstein, n.d.) for an explanation.) He then asks us to consider the parabolas

$$y = \frac{1}{2}x^2 \quad \text{and} \quad \left(y - \frac{1}{2}a\right)^2 = 2bx.$$

The first one has focus $P_1 = (0, 1/2)$ and directrix $L_1 : \ y = -1/2$. The second has focus $P_2 = (b/2, a/2)$ and directrix $L_2 : \ x = -a/2$. These foci are the same that Beloch's method would give, except that they are rotated 90° and scaled down by 1/2, which is the equivalent of multiplying the cubic equation by 1/2. Figure 2.11 illustrates this, but the application of Lill's method can be hard to discern due to the lack of an x^2 term in the polynomial. One should think of the "line segment of length a_2" as being there, parallel to (actually, along) the x-axis but of zero length. (Thus the line "containing" the a_2 segment is the x-axis.)

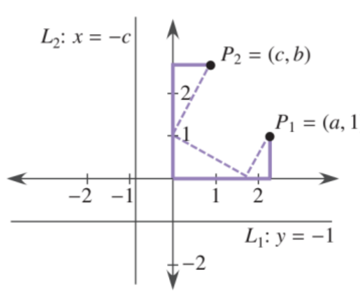

Figure 2.10 Hatori's cubic solution viewed through the lens of Beloch/Lill.

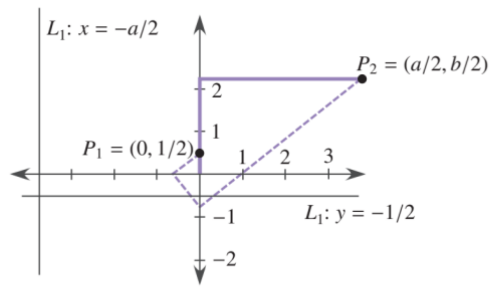

Figure 2.11 Alperin's cubic solution viewed through the lens of Beloch/Lill.

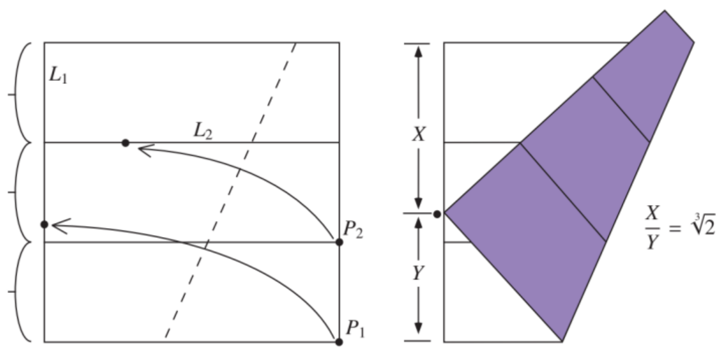

Figure 2.12 Peter Messer's cube root of two construction.

Diversion 2.7 Prove that in Alperin's construction the slope of the crease line we obtain (if it exists) by folding P_1 onto L_1 and P_2 onto L_2 using origami operation O7 will be a real solution to $x^3 + ax + b = 0$.

Diversion 2.8 Use Beloch's method to devise a paper-folding way to construct $\sqrt[3]{2}$. (You're likely to come up with a similar folding procedure to that of Hatori (2003).)

Given how all these methods of solving cubics via origami are really the same at heart, it is surprising to then find one that is different. Peter Messer's cube root of two construction (Messer, 1986) does not fit as easily into the Beloch/Lill strategy; see Figure 2.12. Messer does utilize two parabolas whose directrices are perpendicular, as Beloch does. But if we were to use Beloch's method to solve $x^3 - 2 = 0$, our 3-order path would start with a horizontal segment of length 1, then two segments (vertical then horizontal) of length 0, and then a vertical segment of length 2. There is no way the points O and T would share the same x- or y-coordinate, as do Messer's points P_1 and P_2. However, Messer's method finds two lengths whose **ratios** give $\sqrt[3]{2}$, as opposed to a length obtained by a slope of the turtle path from Lill's method.