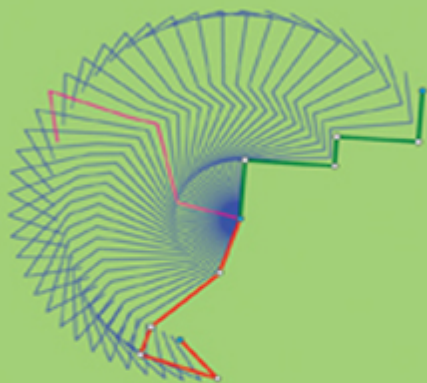
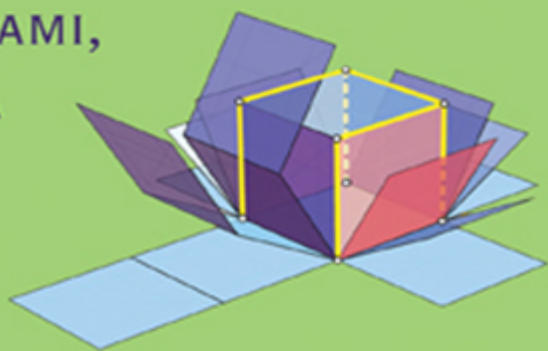


HOW TO FOLD IT

THE MATHEMATICS OF
LINKAGES, ORIGAMI,
AND POLYHEDRA



JOSEPH O'ROURKE

CAMBRIDGE

CAMBRIDGE

more information – www.cambridge.org/9780521767354

This page intentionally left blank

HOW TO FOLD IT

The Mathematics of Linkages, Origami, and Polyhedra

What do proteins and pop-up cards have in common? How is opening a grocery bag different from opening a gift box? How can you cut out the letters for a whole word all at once with one straight scissors cut? How many ways are there to flatten a cube?

You can answer these questions and more through the mathematics of folding and unfolding. From this book, you will discover new and old mathematical theorems by folding paper and find out how to reason toward proofs.

With the help of 200 color figures, author Joseph O'Rourke explains these fascinating folding problems starting from high school algebra and geometry and introducing more advanced concepts in tangible contexts as they arise. He shows how variations on these basic problems lead directly to the frontiers of current mathematical research and offers ten accessible unsolved problems for the enterprising reader. Before tackling these, you can test your skills on fifty exercises with complete solutions.

The book's web site, <http://www.howtofoldit.org>, has dynamic animations of many of the foldings and downloadable templates for readers to fold or cut out.

Joseph O'Rourke is Olin Professor and Chair of the Computer Science Department, a Professor of Mathematics, and Director of Arts and Technology at Smith College. His research is in computational geometry, developing algorithms for geometric computations. He has won several awards, including a Guggenheim Fellowship in 1987 and the NSF Director's Award for Distinguished Teaching Scholars in 2001. He has published more than 150 papers, more than 30 of which were coauthored with undergraduates. He has taught folding and unfolding to students in grade school, middle school, high school, college, and graduate school, and to teachers – of grade school, middle school, and high school – professors, and researchers. This is his sixth book.

How to Fold It

The Mathematics of Linkages, Origami, and Polyhedra

JOSEPH O'ROURKE

Department of Computer Science, Smith College



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521145473

© Joseph O'Rourke 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2011

Printed in China by Everbest

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

Library of Congress Cataloging in Publication Data

ISBN 978-0-521-76735-4 Hardback
ISBN 978-0-521-14547-3 Paperback

Additional resources for this publication at <http://www.howtofoldit.org>

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for
external or third-party Internet Web sites referred to in this publication and does not guarantee
that any content on such Web sites is, or will remain, accurate or appropriate.

Contents

Preface *page ix*

PART I. Linkages **1**

1	Robot Arms	3
1.1	Annulus	5
1.2	Reaching Angles	15
1.3	Above & Beyond	20
2	Straight-Line Linkages and the Pantograph	24
2.1	Straight-Line Linkages	24
2.2	Pantograph	28
2.3	Above & Beyond	36
3	Protein Folding and Pop-Up Cards	39
3.1	Fixed-Angle Chains	39
3.2	Protein Backbones	40
3.3	Maximum Span	42
3.4	Alignment	44
3.5	Piercing	46
3.6	Pop-Up Spinner	48
3.7	Above & Beyond	52

PART II. Origami **55**

4	Flat Vertex Folds	57
4.1	Mountain and Valley Creases	57
4.2	Single-Vertex Flat Folds	58
4.3	The Maekawa-Justin Theorem	61
4.4	The Local Min Theorem	64
4.5	The Kawasaki-Justin Theorem	66
4.6	Above & Beyond	68

5	Fold and One-Cut	72
5.1	Examples	72
5.2	Fold and One-Cut Theorem	78
5.3	Above & Beyond	81
6	The Shopping Bag Theorem	84
6.1	Two Rigid Origami Examples	85
6.2	Dihedral Angle Constraints	89
6.3	The Shopping Bag Theorem	93
6.4	Above & Beyond	96
PART III. Polyhedra		99
7	Dürer's Problem: Edge Unfolding	101
7.1	Albrecht Dürer's Nets	101
7.2	Convex Polyhedra	103
7.3	The Open Problem	106
7.4	Spanning Cut Tree	109
7.5	Some Polyhedra with Nets	112
7.6	Above & Beyond	115
8	Unfolding Orthogonal Polyhedra	119
8.1	Orthogonal Polyhedra	119
8.2	Orthogonal Terrains	120
8.3	Grid Unfoldings	125
8.4	Above & Beyond	126
9	Folding Polygons to Convex Polyhedra	130
9.1	Questions	132
9.2	Alexandrov's Theorem	133
9.3	Folding Convex Polygons	135
9.4	The Foldings of the Latin Cross	138
9.5	Above & Beyond	140
10	Further Reading	142
	Glossary	147
	Answers to Exercises	151
	Chapter 1	151
	Chapter 2	155
	Chapter 3	156
	Chapter 4	158
	Chapter 5	161
	Chapter 6	162

Chapter 7	165
Chapter 8	168
Chapter 9	170
<i>Acknowledgments</i>	173
<i>Index</i>	175

Preface

Cutting out the paper-doll figures below requires 64 straight scissors cuts if done without folding the paper. However, folding the paper along the dashed vertical creases lets you cut out all four people by just cutting one outline in the folded paper. Then you only have to do one-quarter of the work – 16 straight snips of the scissors rather than 64. Noticing that each figure is symmetric about a vertical line through the center of its octagonal head (humans have bilateral symmetry!) and additionally folding along that line permits cutting out all four people with eight straight scissors cuts, now through eight layers of paper. Wouldn't it be nice if there were a way to fold the pattern so that you could get away with a single straight slice of the scissors? Well, believe it or not, there is such a folding!

This beautiful “Fold and One-Cut” result is the topic of Chapter 5 in this book. We will see in that chapter that it is already not so straightforward to cut out a single irregular triangle in the center of a piece of paper with just one scissors cut. But understanding the triangle is a big step toward understanding how to cut out the four paper dolls. Folding the paper in preparation for cutting out a triangle reveals in its creases a theorem we all learned as teenagers (and most of us forgot!): The three angle bisectors of any triangle meet at a single point. Seeing the angle bisector creases converging at a point makes this abstract theorem (proved by Euclid) concrete and unforgettable.

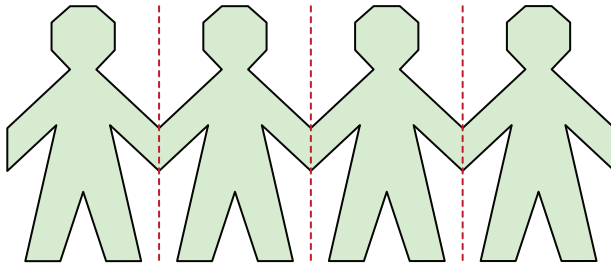


Figure 0.1. Paper-doll people for cut out.

This is the ideal at which I am aiming in this book. The nine chapters are unified by the notion of folding, but also unified by focusing on tangible constructions with rich mathematical content. I want you to be able to see the mathematical structure present in concrete, physical objects. The book is partitioned into three parts reflected in the title, which can be viewed loosely as concentrating on folding one-dimensional objects (linkages), two-dimensional paper (origami), and three-dimensional objects (polyhedra). (I will henceforth use 1D, 2D, and 3D as abbreviations.)

This book grew out of a monograph, *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*, aimed more at graduate students and researchers in computer science and in mathematics. Both my coauthor, Erik Demaine, and I have taught aspects of this material at various educational levels, from fifth-grade through high school, and found that the tangibility of the topics made them accessible through physical intuition. My goal in this book (which parallels the structure but not the content of *Geometric Folding Algorithms*) is to capitalize on the readers' physical intuition to introduce them to a variegated world of fascinating mathematics.

I assume only high-school mathematics: a little algebra, a little geometry, trigonometry only in a few marked exercises, no calculus or anything beyond. No computer science knowledge is presumed. Occasional boxed material explains technical terms and theorems that some readers will know but others will not (for example: vectors; the triangle inequality; convexity). Further terms are explained in the Glossary. Each chapter aims to reach one or a few mathematical gems. Because each topic is much larger than what I present, each chapter ends with an "Above & Beyond" section to explore more advanced results. I've avoided literature citations in the text, saving them for Chapter 10, "Further Reading."

Technical Terms and Symbols

I should explain two conventions from technical mathematical writing that may be unfamiliar to the reader. The first is that *technical terms* are italicized when introduced and defined, to alert the reader that a word or phrase is being given a special, usually technical meaning that may differ from its use in ordinary language. For example, in Chapter 1, I define the "shoulder" of a chain linkage in analogy with a human shoulder but with a specific meaning in the context of that chapter. The most important technical terms are gathered and defined in the Glossary. To avoid ambiguity, I underline for emphasis, reserving italics for technical definitions.

Second, there is a certain style of introducing symbols in mathematical writing to both shorten and make more precise the discussion. A typical example is, "Let x be a point on the polyhedron P in Figure 3." This means: Henceforth (for the duration of this discussion), we will use the symbol x to mean an arbitrary point on the polyhedron and the symbol P to mean the specific polyhedron illustrated in Figure 3. Sometimes the symbol introduction is flagged by "let ... be," and sometimes it is implicit, as in the case of P above.

What Is a Proof?

Many students learn “two-column” proofs in high school, and then never take any higher-level mathematics courses, or, if they do, those courses do not contain proofs. For example, many calculus courses focus almost solely on the “calculating” aspects of calculus. Two-column proofs are the exception rather than the rule in mathematics. They may be the norm for *formal proofs*, where every step is justified by reference to an axiom or some previously established theorem. But most proofs in mathematics are a mixture of prose and symbols, often supplemented by reference to figures labeled with those symbols. A proof is something like a legal brief. It is intended to convince an appropriately prepared reader that a formal proof *could* be formulated, even though it rarely is. To achieve this, a proof must cover all cases, delve into every logical corner, and provide cogent reasons why the reader should “see” that all claims in the proof must be true. A proof is generally written for a particular audience, which defines what is “an appropriately prepared reader.” The proofs that professional mathematicians write for one another would not be convincing for those without similar training.

This book contains many proofs, for I believe that proofs are the heart of mathematics. The audience member I am assuming for these proofs is an attentive reader who has taken (or is currently studying) standard high-school mathematics. I say “attentive” because I will describe a concept in one chapter and expect the reader to both master it and remember it in a later chapter (aided by a back-reference or the index). But the mastery will not require any background beyond high-school mathematics.

I also include several “proof sketches,” which are a cut below a proof in that they do not pretend to handle every logically possible case, or to be stand-alone convincing. Proof sketches are intended to give the reader a feel for how a full proof might go. Often they leave out messy details and ask the reader to believe a claim that those details can all be worked out (and in fact, have all been worked out in the professional literature).

Theory Versus Applications

Despite the tangible aspect of folding, the material in this book focuses on the theoretical, as the emphasis on proofs indicates. A parallel and quite interesting book could be written that instead emphasized the applications of folding, only citing the underlying theory when appropriate. My tack has been nearly the obverse: I plunge into the theory and only cite the applications. I do this for two reasons. First, the underlying mathematics is beautiful in itself, a beauty that can only be fully appreciated by immersion in the details, and conveying this is one of my primary goals. Second, time and again, advances in mathematics seemingly divorced from reality have proved to have significant applications, sometimes much later. To cite just one example unrelated to folding, the Austrian mathematician Johann Radon invented in 1917 what is now known as the “Radon transform.” Although his motivation was to extend the theoretical notion

of integral from calculus to a special situation, the Radon transform is now used daily in hospitals the world over to reconstruct images taken by Computer-Aided Tomography (CAT) scanners. So the mathematical theory is both beautiful and often surprisingly useful.

Exercises

Each chapter contains a number of exercises, with answers in the back of the book (Answers to Exercises). I have partitioned them into three types: *Practice* – Questions to affirm basic understanding of the immediately prior material, often only requiring a bit of calculation; *Understanding* – Questions that require a thorough grasp of the preceding material, often applied to a slightly new situation; and *Challenge* – Problems that ask for substantive extensions and/or significant investments of time. The lines between these three classes are not sharp, nor are those lines the same for all readers. In any case, I encourage the reader to read each exercise, work out as many as circumstances permit, and in any case, to please look at the answers, which often enrich the material.

Templates on the Web

At a number of junctures in the book, particularly in the origami chapters, the reader is invited to cut out or fold a particular illustrated diagram. Each such diagram is available on the book's Web site, <http://howtofoldit.org/>. Each can be downloaded and printed. The Web site contains other useful supplementary information.

Open Problems

This book includes many unsolved problems, usually called *open problems* in mathematics. These are clear statements that have not yet been settled as either TRUE or FALSE by a proof. Sometimes researchers are convinced a hypothesis is TRUE even though they cannot prove it. In such a case, the hypothesis is designated as a *conjecture*. Some open problems have resisted all attempts over many years. However, most progress in mathematics occurs not by settling these long-unresolved problems but rather by answering recently posed questions. So I have included a number of new problems (a few concocted while writing this book), which may be open primarily through lack of attention. Rarely are the frontiers of mathematical knowledge accessible to the amateur, but one attractive aspect of the topic of folding is that many of its unsolved problems are accessible to the novice and might be solved by just the right clever idea. Please let me know if you crack one of them! They are listed in the Index under “open problems” for ease of access.

PART I

Linkages

Linkages are mechanisms built from stiff, inflexible bars, which we will call *rigid links*, connected at freely rotating joints. You may have a linkage on your desk similar to the one depicted in Figure I.1. Many machines contain linkages for particular functions. Every car contains a crankshaft, a mechanism for converting the linear motion induced from the sparked explosion of gasoline in a piston chamber to the rotary motion turning the drive shaft. We'll explore three linkages, each with a clean mathematical story to tell, and each related to developments on the frontiers of mathematics and computer science today: robot arms, pantographs, and fixed-angle chains. We'll analyze the "reachability region" for robot arms viewed as a linear chains of links. The pantograph is a mechanical copying and enlarging mechanism with myriad uses, especially during the industrial revolution. Fixed-angled chains are superficially similar to robot arms, but are primarily of interest as models of protein backbones.

Although here we are emphasizing the relevance of these linkages, our focus will be on the mathematics behind their operation.

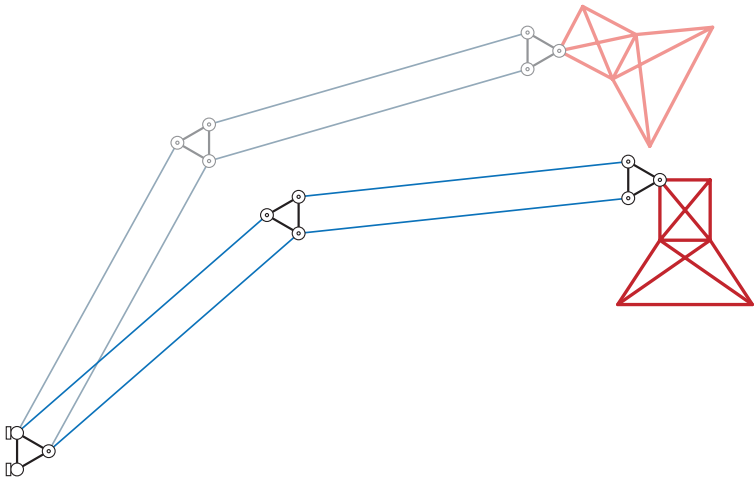


Figure I.1. A desk-lamp linkage. The linkage flexes at the circled joints, but is structurally rigid otherwise.

1

Robot Arms

Robot arms, despite their sophistication as machines, are particularly simple if you think of them as linkages. The arm in Figure 1.1(a), developed by a British robotics firm, is designed to apply adhesive tape to the edges of pieces of plate glass for protection. It has a fixed base (the *shoulder*) to which are attached three rigid links, corresponding roughly to upper arm, lower arm, and *hand*, or, in the technical jargon, the *end effector*. The rotation settings at the motorized joints determine the exact positioning of the hand as it performs its functions. The force dynamics and engineering aspects of robot arm design are quite interesting and challenging. However, we will focus on one simple question: determining what is called the *workspace* of the robot – the spots in space it can reach. We will pursue this question in almost absurd generality, permitting the arm to have an arbitrarily large number of links, each of an arbitrary length.

Model. First we need to reduce a complex physical robot arm to a simple mathematical model so that it can be analyzed. Typically, the initial abstraction chosen is crude, ignoring many physical details, and then, once analyzed, gradually made more realistic and complicated.

We reduce each robot arm piece to a straight-line *segment* of fixed length – a rigid link of mathematically zero thickness. Each joint motor is reduced to a mathematical point of zero extension joining the two *incident* links that it shares. So we have reduced the physicality of a real robot arm to segments and the endpoints of those segments; see Figure 1.1(b).

There are two more crucial physicalities to model: intersections and joint motions. Of course, no two distinct physical objects may occupy the same space at the same time, so the links should not be permitted to *intersect* – share points – except sharing the point at a common joint. However, we start our analysis with

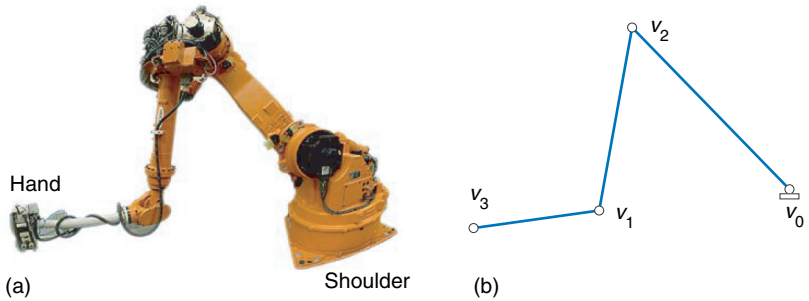


Figure 1.1. (a) A robot arm. (b) Arm modeled by linkage. v_0 labels the shoulder joint, and v_3 the hand.

the physically unrealistic assumption that intersections are ignored. Similarly, although most robot joints have physical constraints that prevent a full 360° rotation in two dimensions (2D), or free rotation in all directions in three dimensions (3D), we assume no joint constraint – so each is a *universal* joint, one that has total freedom of rotation. Later in exercises and in Chapter 3, we will constrain the joints.

So our mathematical model of a robot arm is a chain of n links, where n is some natural number $1, 2, 3, \dots$, each a fixed-length segment of some prespecified length, connected by universal joints. For the robot arm in Figure 1.1, $n = 3$: 3 links, 3 joints (including the motorized shoulder). The hand/end effector is not a joint, just a link endpoint. Indeed, the number of links and the number of joints is always the same, n , under this convention of viewing the shoulder, but not the hand, as a joint.

Now the question is: Under this model, what is the totality of locations in 3D space that an n -link robot arm can reach? This set is called the *reachability region* of the arm.

At this point, we invite the reader to guess the answer that this chapter will soon establish more formally. Reasoning from your own shoulder-to-hand linkage may be misleading, because humans have definite (and complex) joint constraints. Perhaps it will help intuition to imagine a specific example. Suppose we have 3-link arm whose link lengths are 10, 5, and 3. What is the region of space that the hand endpoint can reach? Hint: It is not a sphere of radius 18!

Box 1.1: Theorem

In mathematics, the term *theorem* is used for a concise statement of a central result, whereas a *lemma* is a result that is a stepping-stone on the way to a theorem. A *corollary* is a near-immediate consequence of a theorem. Although we will not use the term, a *proposition* is often used for a relatively straightforward theorem.

1.1 Annulus

Rather than keep the reader in suspense, let us immediately move to the answer to this question, which we encapsulate in a theorem (see Box 1.1):

Theorem 1.1

The reachability region of an n -link robot arm is an annulus.

Now we should explain the term *annulus*. In 2D, an annulus is the region between two circles with the same center but different radii. Such circles are called *concentric*. The 3D analog, the region between two concentric spheres of different radii, is generally called a “spherical shell,” but we opt to use “annulus” regardless of the dimension. See Figure 1.2(b). Right now we concentrate on 2D and consider 3D later (p. 19). For our 3-link example with link lengths 10, 5, and 3, the reachability region is an annulus with outer radius 18 and inner radius 2. That the inner radius is 2 is by no means obvious; it will be established later in Theorem 1.2.

There are two special cases that we further include under the term “annulus”: (1) If the radii of both circles are equal, the region reduces to just that circle itself; (2) if one radius is zero, the region is the entire disk enclosed by the circle. A *circle* can be viewed as a rim “wire” whereas a *disk* includes the points inside the wire.

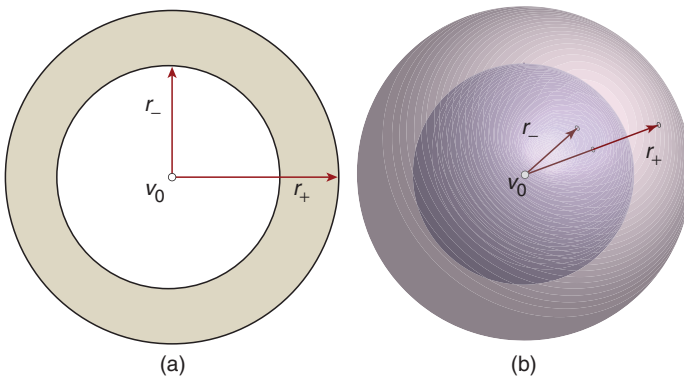


Figure 1.2. (a) 2D annulus: the region between two concentric circles. (b) 3D annulus (also known as a spherical shell): the region between two concentric spheres. Common centers are labeled v_0 .

Box 1.2: Induction

Induction is a proof technique that can be used to establish that some claim is true for all numbers $n = 1, 2, 3, \dots$. It is akin to climbing a ladder: If you know how to move from any one rung to the next, and you know how to reach the first rung, then you can climb to any rung, no matter how high. To reach the first rung, we only need prove the result holds for $n = 1$, the *base* of the induction. Moving from one rung to the next requires proving that if the theorem holds for $n - 1$ (you've reached that rung), then the theorem holds for n , where n is an arbitrary natural number. Then the theorem must be true for all n , "by induction," as they say: From $n = 1$, we can reach $n = 2$, and from there we can reach $n = 3$, and so on.

Annulus Proof. The proof of Theorem 1.1 uses a method known as *induction*; see Box 1.2.

The base case is straightforward: A 1-link arm can reach the points on a circle, and by our definition, a circle is an annulus. Now we could jump immediately to the general case using induction. But let's look at $n = 2$ to build intuition; say the two link lengths are r_1 and r_2 . This 2-link arm can reach all the points on a circle of radius r_2 centered on any of the points on a circle of radius r_1 . Figure 1.3 illustrates the idea. Imagine sweeping the red r_2 -circle around, centered on each point of the blue r_1 -circle. The swept pink region R_2 is an annulus.

Let us now consider the general case, an n -link arm, $n > 1$. Following the induction paradigm, we assume that we have established the theorem for arms up to $n - 1$ links. Then we know if we remove the last link of a given n -link arm

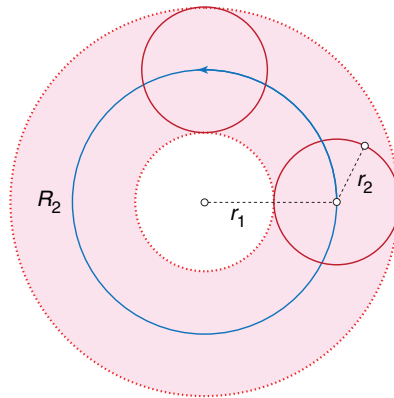


Figure 1.3. A 2-link arm can reach points in an annulus.

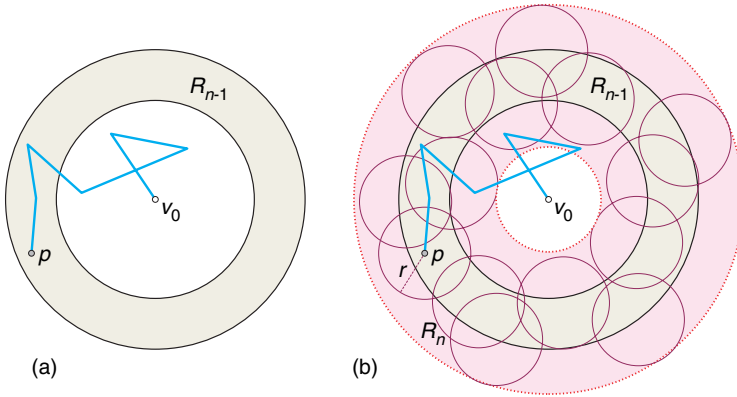


Figure 1.4. (a) R_{n-1} with one possible arm of $n-1$ links reaching a point p . (b) R_n is formed by adding the points on circles centered on every point p in R_{n-1} , with the radius r of these circles equal to the length of the last link of the arm.

(call it A_n), the shorter arm's reachability region is an annulus, because it has only $n-1$ links. (We have just employed the "induction hypothesis": $n-1$ link arms reach points in an annulus.) Let us call the shorter arm A_{n-1} and its region R_{n-1} . We seek to find R_n , the reachability region for A_n .

Let p be any point in R_{n-1} . We know that the hand of A_{n-1} can reach p , as in Figure 1.4(a). Now imagine adding the removed final link back to A_{n-1} . This permits A_n to reach all the points on a circle centered on p , where the circle's radius r is the length of that last link. So we can construct R_n by adding the points on a circle of radius r centered on every point p of R_{n-1} . See Figure 1.4(b).

Here I rely on the reader's intuition to see that R_n is again an annulus: Adding all these circles to an annulus results in a fatter annulus. Points p on the outer boundary of R_{n-1} reach out to a larger-radius circle bounding R_n , larger by r , and points on the inner boundary of R_{n-1} reach inward to a smaller-radius circle, smaller by r . Circles around points p in the interior of R_{n-1} fill out the remainder of the annulus. If r is enough to reach the center of R_{n-1} , then R_n becomes a disk, which we have defined as an annulus.

1.1.1 Radii

Our proof that the reachability region is an annulus does not directly yield the radii of the annulus. In particular, it would be useful to know under what conditions the reachability region is a disk, that is, when the hand can touch the shoulder. We now address this question.

Because the answer will depend on the arm's lengths, we will need some notation for those. Call the lengths of the n links $(\ell_1, \ell_2, \dots, \ell_n)$, and call the outer and inner radii of the annulus r_+ and r_- respectively. The outer radius is easy: The furthest reach of the arm is achieved by straightening each joint, stretching the arm out straight. Recalling our 3-link example with lengths $(10, 5, 3)$,

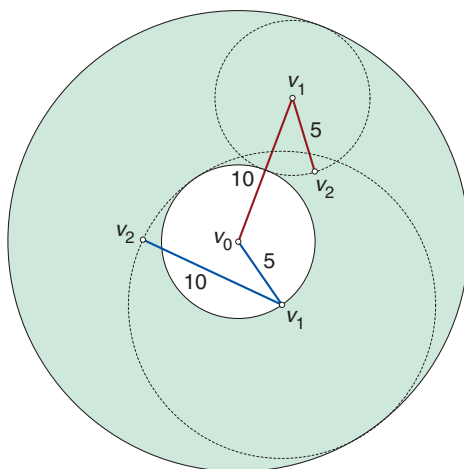


Figure 1.5. Annulus for the 2-link arm with lengths (5, 10) (red) is the same as for the arm with lengths (10, 5) (blue).

$r_+ = 10 + 5 + 3 = 18$. In general,

$$r_+ = \ell_1 + \ell_2 + \cdots + \ell_n.$$

Computing the inner radius r_- is less straightforward. A key idea that helps is hinted at by Figure 1.5, which shows that the reachability annulus for an arm consisting of two links of lengths 5 and 10 is independent of whether the longer or the shorter is the first link, incident to the shoulder. Somewhat surprisingly, this independence holds more generally:

Lemma 1.1

The reachability region of a robot arm is independent of the order of the link lengths: It only depends on the numerical values of those lengths, not the order in which they appear along the chain of links.

I will argue for this lemma before explaining its relevance to computing r_- . Let v_0 be the location of the shoulder joint of the arm, and $v_1, v_2, \dots, v_{n-1}, v_n$ the positions of the remaining joints, or, as they are commonly known in geometry, the *vertices* of the chain. (The singular is *vertex*.) The last vertex v_n is the position of the hand, not considered a joint (because there is nothing beyond that it joins). In any particular configuration of the arm, the vertices are at particular points in the plane. We take v_0 to be the origin of the coordinate system in which we express the points: $v_0 = (0, 0)$.

Box 1.3: Vectors

We illustrate the notion of a vector with the 3-link arm shown in Figure 1.6(a), whose shoulder is at $v_0 = (0, 0)$ and whose vertices are located at $v_1 = (1, 1)$, $v_2 = (1, 0)$, and $v_3 = (0, 3)$, with the shoulder at the origin $v_0 = (0, 0)$. The lengths of the links are $\ell_1 = \sqrt{1^2 + 1^2} = \sqrt{2}$, $\ell_2 = 1$, and $\ell_3 = \sqrt{1^2 + 3^2} = \sqrt{10}$. We can view each successive vertex as displaced from the previous one. So v_2 is obtained from v_1 by moving vertically down one unit, and v_3 is obtained from v_2 by one step left horizontally and three up vertically. These displacements are *vectors*, and can be computed by subtracting the points coordinate by coordinate. We will use uppercase letters with over-arrows to indicate vectors. So,

$$\vec{V}_1 = v_1 - v_0 = (1, 1) - (0, 0) = (1, 1)$$

corresponds to moving right and up 1,

$$\vec{V}_2 = v_2 - v_1 = (1, 0) - (1, 1) = (0, -1)$$

corresponds to moving down 1, and

$$\vec{V}_3 = v_3 - v_2 = (0, 3) - (1, 0) = (-1, 3)$$

corresponds to 1 left, 3 up. Because we chose $v_0 = (0, 0)$, $\vec{V}_1 = v_1 - v_0$ is the same as $v_1: (1, 1)$. The length of these vectors is exactly the link length which they “span,” for example, the length of \vec{V}_3 is $\sqrt{10}$.

There is a certain ambiguity when we represent a point by its coordinates and a vector by its coordinate displacements, for they both look the same as pairs of numbers: Thus the point v_1 has the same coordinate representation as the vector \vec{V}_1 . But a point is a location in the plane, whereas a vector is a displacement in the plane. Every point in the plane can be viewed as a displacement from the origin – a viewpoint that is often convenient.

Two vectors are added by adding their displacements coordinate by coordinate. So the sum of the vectors $(1, 1)$ and $(0, -1)$ is $(1, 0)$, which, not surprisingly, is v_2 :

$$\vec{V}_1 + \vec{V}_2 = (v_1 - v_0) + (v_2 - v_1) = v_2 - v_0$$

which is v_2 because $v_0 = (0, 0)$.

The key to the proof of this lemma is to think of the vertices of the joints as reached by a series of *vector* displacements from the shoulder. Vectors are an important concept we will use in several chapters; see Box 1.3. Let us represent the vector displacement between adjacent vertices with the symbol \vec{V}_i , with $\vec{V}_i = v_i - v_{i-1}$, where the subscript i can take on any integer value between 1

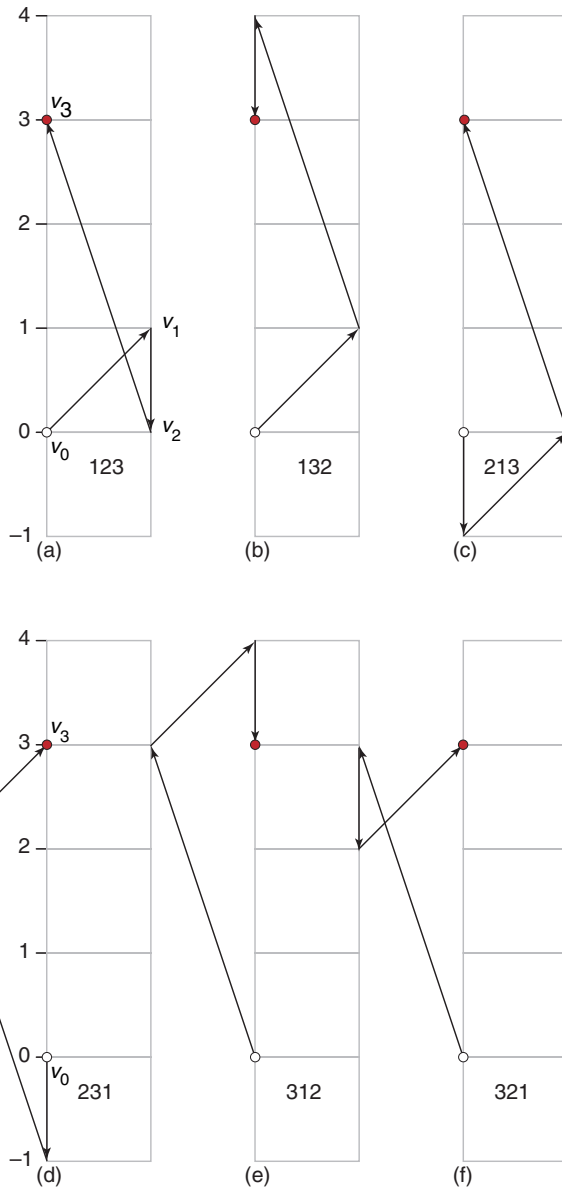


Figure 1.6. A 3-link arm reaching from $v_0 = (0,0)$ (white circle) to $v_3 = (0,3)$ (red circle). All six possible permutations (indicated below each figure) of adding the three vector displacements all reach to $(0,3)$.

and n . So the hand vertex v_n can be reached from the shoulder at the origin by adding up these vector displacements in sequence:

$$v_n = \vec{V}_1 + \vec{V}_2 + \cdots + \vec{V}_n.$$

Now, we all know that the addition of a series of numbers is *commutative*, that is, the result is independent of the order in which we sum them. For example, $1+2+3=3+2+1$. The order of summation can be “commuted” without altering the result. You may already be able to see that vector addition is also commutative, because we add vectors coordinate by coordinate, and coordinates are numbers for which the commutative property holds. Figure 1.6 shows that the other five different possible orders in which to add the three vectors for the example all reach the exact same point.

Exercise 1.1 (Practice) Vector Sum Commutativity. Let three displacement vectors be $\vec{V}_1 = (2,0)$, $\vec{V}_2 = (1,3)$, and $\vec{V}_3 = (-1,1)$. Starting from the shoulder at the origin $v_0 = (0,0)$, draw out, in one figure, all six different ways to add the three vectors, and show that they all reach the same point $\vec{V}_1 + \vec{V}_2 + \vec{V}_3 = (2,4)$.

So now we proved Lemma 1.1: Any point in the reachability region of an arm can be reached by shuffling the links, and therefore displacement vectors, of that arm.

Inner Radius. Now we return to the task of computing the inner radius r_- of the annulus. Let L be the length of a longest link among $\ell_1, \ell_2, \dots, \ell_n$; suppose $\ell_k = L$. We shuffle the links so that ℓ_k is first, knowing by Lemma 1.1 that the reachability region is unchanged. Thus, in Figure 1.5, we place link of length 10 before that of length 5. So the arm now consists of link lengths:

$$(\ell_k, \ell_1, \ell_2, \dots, \ell_{k-1}, \ell_{k+1}, \dots, \ell_n).$$

Let M be the sum of the lengths beyond the first, that is, excluding ℓ_k :

$$M = \ell_1 + \ell_2 + \cdots + \ell_{k-1} + \ell_{k+1} + \cdots + \ell_n.$$

Case 1: $M < L$. The reshuffling makes it clear that if $M < L$, then the hand cannot reach the shoulder, that is, v_n cannot reach v_0 because M is not sufficient to stretch back from v_1 to v_0 . The closest it can reach is $L - M$, and this is r_- ; see Figure 1.7.

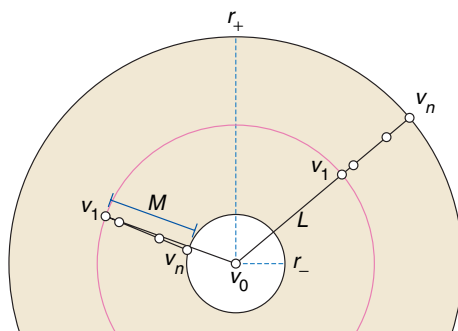


Figure 1.7. When $L > M$, the inner radius is $r_- = L - M$.

Box 1.4: The Triangle Inequality

The *triangle inequality* is a simple relationship that is surprisingly useful. For any triangle whose side lengths are A , B , and C (see Figure 1.8),

$$C \leq A + B. \quad (1.1)$$

This encapsulates in one form the fact that a straight line is the shortest distance between two points. If you are standing at one endpoint x of the length C and desire to reach the other endpoint y , you can do no better than follow the segment xy . Moving instead along two straight paths, which must form a triangle to reach from x to y , is always longer. See Figure 1.8(a).

Of course there is nothing privileged about C in Eq. 1.1. It must be that $A \leq B + C$ and $B \leq A + C$ as well.

Another inequality holds for any triangle:

$$|A - B| \leq C. \quad (1.2)$$

Here $|A - B|$ means the absolute difference between A and B : $A - B$ if A is larger, $B - A$ if B is larger, and zero if $A = B$. This relationship follows from Eq. 1.1. If $A \geq B$, then it follows from $A \leq B + C$, and if $B \geq A$, it follows from $B \leq A + C$.

A type of converse of the triangle inequality holds as well: To any three numbers $\{A, B, C\}$ satisfying

$$|A - B| \leq C \leq A + B \quad (1.3)$$

there corresponds a triangle with those side lengths. Figure 1.8(b) indicates why: When Eq. 1.3 holds, a circle centered at x of radius A intersects a circle centered at y of radius B .

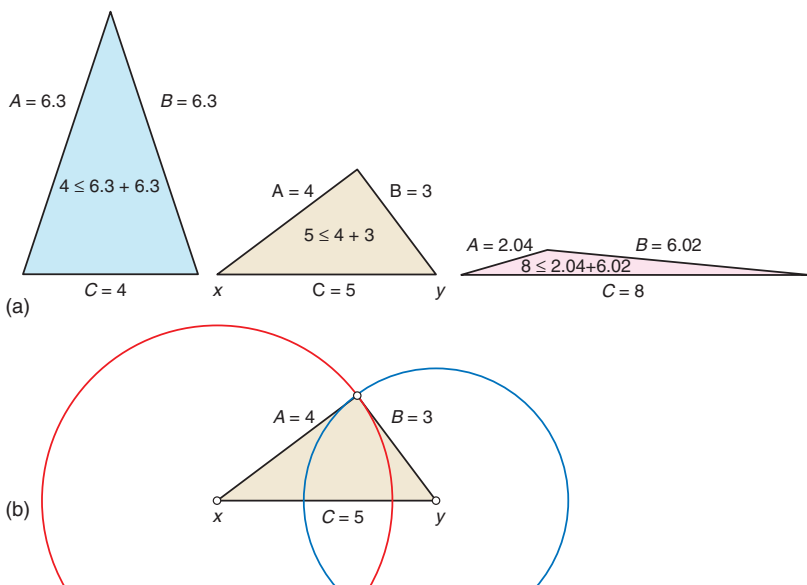


Figure 1.8. (a) The triangle inequality. (b) The converse of the triangle inequality.

Exercise 1.2 (Practice) Triangle Inequality. Which of these triples of numbers satisfy the triangle inequality? For those triples that do satisfy the inequality, sketch the corresponding triangle.

A	B	C	\triangle YES/NO
10	5	3	—
10	5	6	—
4	5	3	—
4	5	10	—

Case 2: $M \geq L$. When $M \geq L$, we'd like to claim that v_n can reach v_0 , which means that the annulus becomes a disk, and $r_- = 0$. This is intuitively plausible, as we have more than enough length M to reach from v_1 back to v_n . But the linkage is not infinitely flexible like a rope – it can only bend at its joints – so we need an argument to be certain. The proof we provide relies on the *triangle inequality*, described in Box 1.4.

Assume $M \geq L$, and partition the chain of length M from v_1 to v_n into two parts whose lengths M_1 and M_2 are as close to $\frac{1}{2}M$ as possible. The difference between M_1 and M_2 will not exceed L , because if it did, the gap could be narrowed by moving a link (necessarily of length $\leq L$ because L is the longest) from the larger to the smaller side of the partition. Now the three lengths $\{L, M_1, M_2\}$ satisfy the

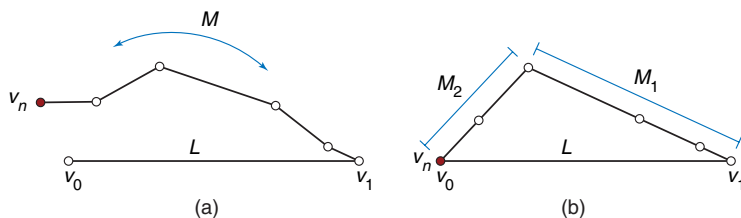


Figure 1.9. (a) $M \geq L$. (b) A triangle can be formed from lengths $\{L, M_1, M_2\}$.

triangle inequality:

$$|M_1 - M_2| \leq L \leq M_1 + M_2.$$

This implies that a triangle can be formed with those side lengths, as illustrated in Figure 1.9, which places v_n at v_0 , as claimed.

So we have finally settled the radii of the annulus, which we summarize in a theorem:

Theorem 1.2

The outer radius of the reachability annulus of an n -link arm is

$$r_+ = \ell_1 + \ell_2 + \cdots + \ell_n,$$

and the inner radius is

$$r_- = \begin{cases} L - M & \text{if } L > M \\ 0 & \text{if } L \leq M, \end{cases}$$

where L is the length of a longest link in the arm, and M the sum of the lengths of all the other links.

We have reached our first goal: a precise description of the reachability region of an n -link arm. Remarkably, all the analysis so far holds for 3D as well as 2D. In particular, Theorem 1.2 accurately describes the radii of the two spheres that define the 3D reachability annulus.

Exercise 1.3 (Practice) Reachability Radii. For each of the link lengths listed below, compute the inner and outer annulus radii r_- and r_+ .

Link Lengths	r_+	r_-
(1, 2, 3, 4, 5)	—	—
(1, 2, 10, 3, 4)	—	—
(1, 1, 1, 1, 5)	—	—
(12, 2, 5, 4)	—	—

1.2 Reaching Angles

Although we now know, given any point p , whether or not a given robot arm can reach p , we do not as yet know to which values the joint angles should be set in order to reach p . And this is of course crucial if the robot is to perform a specific task that requires its visiting a prescribed sequence of positions within its workspace. We now turn our attention to this question, first concentrating on finding joint angles that place v_n at p , and only later looking at how to reach that configuration.

So our question is:

Question 1. Given an n -link arm A_n , specified by its link lengths, $A_n = (\ell_1, \dots, \ell_n)$, and a point p within its reachability annulus R_n , which set of joint angles at v_1, v_2, \dots, v_{n-1} place the hand v_n at p ?

This is a more difficult question than finding the reachability region in the first place, requiring a more nuanced analysis. The first step toward simplifying the question is to rephrase it in terms of joint positions, ignoring the angles:

Question 2. Given an n -link arm A_n , and a point p in R_n , find positions for its joints $v_1, v_2, \dots, v_{n-1}, v_n$ such that $v_n = p$.

The reason Question 2 is equivalent to Question 1 is that, knowing the joint positions, we can compute the joint angles using trigonometry. The reason Question 2 is easier is that the joint positions are determined by circle intersections, so they can be found by geometric constructions. We will now answer Question 2, but via more of a “sketch” than a completely detailed analysis, riding a bit above some messy details. We progress through $n = 2$, $n = 3$, and $n > 3$.

2-Link Reachability Angles. We are given $A_2 = (\ell_1, \ell_2)$ and $p = v_2$, and we continue the convention that $v_0 = (0, 0)$. So it only remains to compute the position of the middle joint v_1 . Knowing positions for $\{v_0, v_1, v_2\}$ determines the joint angles.

We know the collection of points (the *locus* of points, as geometers say) reachable from v_0 by the first link only is a circle of radius ℓ_1 centered on the shoulder v_0 . Now we reverse the viewpoint and think of how the arm must look if it reaches out to p . With the second link at $v_2 = p$, the joint v_1 must lie somewhere on the circle of radius ℓ_2 centered on v_2 . So now we have two constraints on where v_1 must lie: It must lie both on an ℓ_1 -circle centered on v_0 , and on an ℓ_2 -circle centered on $v_2 = p$. So the v_1 points that place v_2 at p are points of intersection of these two circles! Thus we have a choice of two different arm positions that place v_2 at p . This is the crucial observation.

In general, two circles intersect in two distinct points; see Figure 1.10(a). This is the *generic* situation – the “typical” or “normal” situation. However, there are special cases, known as *degenerate* situations in technical mathematical

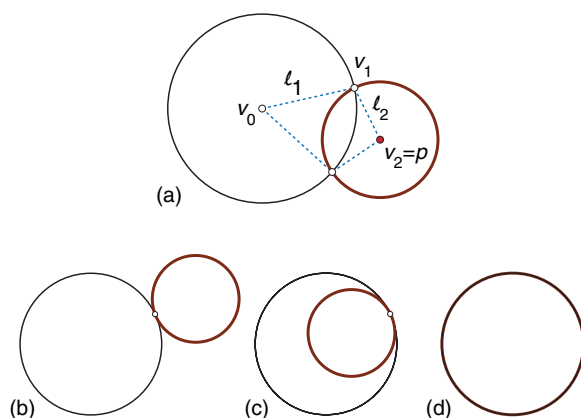


Figure 1.10. (a) Generic situation: two intersection points for v_1 ; (b–d) Degenerate cases: 1, 1, or infinitely many intersection points.

language, where two intersecting circles meet at just one point (Figure 1.10(b, c), or in an infinite number of points (d)). Of course two circles could have no intersection at all, but because we assumed p is within the reachability region of the arm, we know there must be some nonempty intersection. One way in which we are only sketching the computations is that we opt to ignore the degenerate situations, leaving it to the reader's intuition that the details could be worked out with sufficient patience. So, with that caveat, we have the solution to 2-link reachability: Intersect the two circles as in Figure 1.10(a), select one of the two intersection points for v_1 , and then calculate the angles using trigonometry. We will not actually descend into these trigonometric details; our only goal is to show that it could be done.

Exercise 1.4 (Practice) Number of Ways to Reach. Suppose a 2-link arm is given by lengths $\ell_1 = 3$ and $\ell_2 = 1$, that is, $A_2 = (3, 1)$. With $v_0 = (0, 0)$, how many distinct ways are there to reach $v_2 = (3, 1)$? How many to reach $v_2 = (0, 2)$?

3-Link Reachability Angles. Our strategy is to reduce the question for a 3-link arm $A_3 = (\ell_1, \ell_2, \ell_3)$ to one of several 2-link questions, which we've just seen how to handle. Again we are given a point p in the reachability region R_3 of this arm A_3 . As in the 2-link case, the locus of possible positions for v_1 is the circle of radius ℓ_1 centered on v_0 . Now we know that the reachability region of the 2-link arm $A_2 = (\ell_2, \ell_3)$ is an annulus centered on v_1 . Let us call this annulus Q_2 . We can imagine v_1 rotating continuously around its circle of possible locations, and Q_2 rotating along with its moving center v_1 . Q_2 sweeps out the reachability annulus R_3 , as depicted in Figure 1.11(a). Because we know p is in R_3 , we know at some position(s) of v_1 , p must be inside Q_2 ; otherwise p would not be reachable by the arm A_3 .

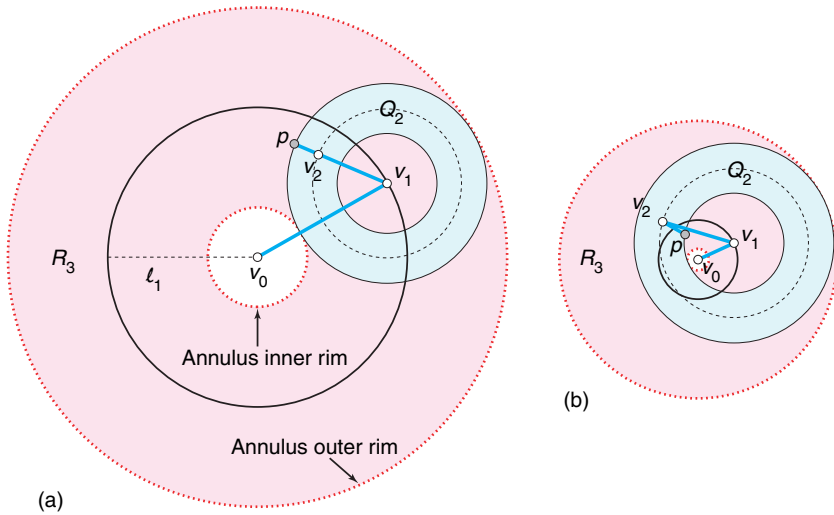


Figure 1.11. The black circles of radius ℓ_1 are possible positions for v_1 . (a) $p = v_3$ is reached by aligning ℓ_2 with ℓ_3 . (b) $p = v_3$ is reached by anti-aligning ℓ_2 and ℓ_3 .

Now there are only three logical possibilities:

Case 1. At some v_1 position in this imagined rotational sweep of the annulus Q_2 , p lies on the outer boundary of Q_2 , either just entering or exiting Q_2 . This is the situation illustrated in Figure 1.11(a).

Case 2. At some v_1 position, p lies on the inner boundary of Q_2 , as illustrated in (b) of the figure.

Case 3. p remains inside Q_2 throughout the entire sweep: It never touches the inner or outer boundary of Q_2 . This occurs, for example, when $\ell_1 = \ell_2 = \ell_3 = 1$ and $p = (\frac{1}{2}, \frac{1}{2})$: Q_2 , a radius-2 disk, covers p for every position of its center v_1 .

In Case 1, the positions for v_1 and v_2 can be found by solving the 2-link problem $(\ell_1, \ell_2 + \ell_3)$. In Case 2, the joint positions can be found by solving the 2-link problems with lengths $(\ell_1, |\ell_2 - \ell_3|)$, where $|\ell_2 - \ell_3|$ is the positive difference between the second and third link lengths: $\ell_2 - \ell_3$ when ℓ_2 is longer than ℓ_3 as in Figure 1.11(b), and $\ell_3 - \ell_2$ when ℓ_3 is longer. In Case 3, every position of v_1 leads to a solution. So we can pick one – say, v_1 on the horizontal line through v_0 – and then solve the 2-link problem with lengths (ℓ_2, ℓ_3) centered on this position of v_1 .

So we have arrived at a method for solving a 3-link problem: solve three 2-link problems corresponding to the above three cases, and at least one of them must have a solution, which then leads to positions for the joints of A_3 . Again knowing these joint positions permits computing the joint angles.

Exercise 1.5 (Practice) *3-link Reaching.* For a 3-link arm given by lengths $A_3 = (3, 1, 2)$, find at least two distinct ways to reach the point $v_3 = (2, 2)$. (Note A_3 is an extension of A_2 from Exercise 1.4.)

Exercise 1.6 (Understanding) *3-link Unique Reaching.* For the 3-link arm with link lengths $A_3 = (2, 1, 2)$, describe the sets of points in the plane for which there is only one way for A_3 to reach. A point p is “uniquely reached” if there is only one setting of the three joint angles (at v_0 , v_1 , and v_2) to place v_3 at p .

n -Link Reachability Angles. We now jump to the general case, an n -link arm A_n (still in 2D), $n > 3$. There are several possible routes here to a solution, but I will emphasize one via the surprising “Two Kinks” Theorem, surprising to me at the time of its discovery by a college student, John Kutcher.

Define a *kink* as any joint angle that is not straightened by aligning the two links it connects, that is, a kink angle is any different from 180° . Of course any 3-link arm can reach a point in its reachability region with (at most) two kinks, because it only has two interior joints (v_1 and v_2). The Two Kinks Theorem says that it never need be more complicated than this for longer arms: An n -link arm can reach any point in its reachability region with at most two kinks! See Figure 1.12 for an illustration. Perhaps this is not so surprising in light of the argument we used (p. 14) to show that when $M \geq L$, the hand v_n can reach the shoulder v_0 , for that construction incidentally showed it can reach the shoulder with just two kinks; see again Figure 1.9. We will not argue further for the Two Kinks Theorem, but just proceed assuming it is true. As in Figure 1.9, the theorem also pinpoints which joints might need to be kinked and which can be aligned.

The Two Kinks Theorem reduces solving an n -link arm problem to solving a 3-link arm, which we reduced to solving 2-link arms. So, in the end, the entire calculation reduces to intersecting two circles!

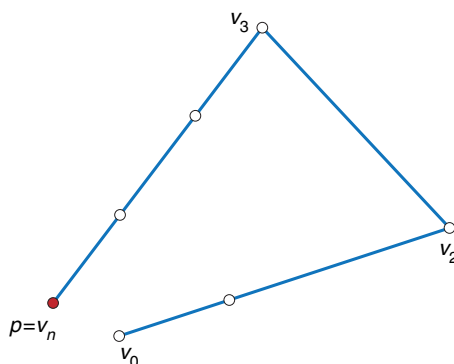


Figure 1.12. Any point p within the reachability region of an n -link arm can be reached with at most two kinks at either end of one particular link, in this case, at v_2 and v_3 .

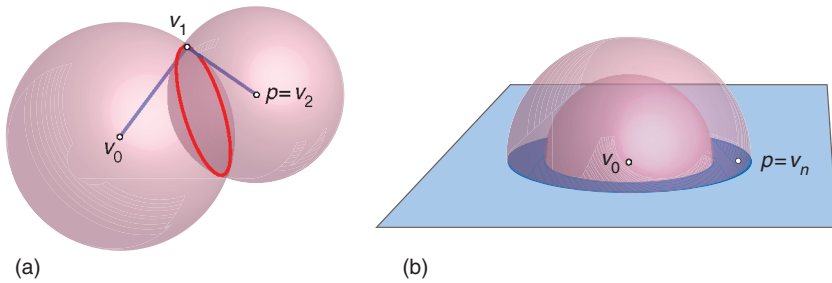


Figure 1.13. (a) The intersection of two spheres is (generically) a circle. (b) The intersection of a 3D annulus with a plane through the center/shoulder v_0 is a 2D annulus.

3D Reaching Angles. We mentioned there is no difference in computing the reachability region in 2D or in 3D. This is not the case with computing the angles, but the difference is minor. Consider first a 2-link arm in 3D, and follow the same logic we used previously. The shoulder is fixed at v_0 , the hand at $v_2 = p$. The middle joint v_1 must lie on the sphere of radius ℓ_1 centered on v_0 , and simultaneously on the sphere of radius ℓ_2 centered on v_2 . So v_1 lies on the intersection of these two spheres, which is (generically) a circle, as illustrated in Figure 1.13(a). This is another “abundance of riches” situation: Rather than choosing from among the two solutions in 2D, here we have an infinite number of solutions from which to choose. All we need is a strategy to reduce our options.

There is an easy resolution. Given an n -link arm and a point p in its 3D annulus R_n (the region between the spheres), slice R_n by a plane that contains v_0 and p . As Figure 1.13(b) illustrates, this plane intersects R_n in a 2D annulus. We can simply solve the 3D problem within this plane, and use those angles. Thus 3D reduces to 2D.

Dynamic Reconfiguration. Given initial and final angles at the joints of an n -link arm, the easiest way to move continuously between the two configurations is to simply *interpolate* the angles. We imagine a clock ticking in small increments from $t = 0$ to $t = 1$ between initial and final angles. If the initial angle at a joint is α and the final angle β , as measured, say, counterclockwise from the horizontal, then at time t , the angle is $\alpha + t(\beta - \alpha)$. This expression evaluates to α at $t = 0$ and to β at $t = 1$. In between, it ramps up linearly between those values. See Figure 1.14. When interpolating angles, it generally makes sense to take the shorter of the two routes around the circle of angles from initial to final angle. So in Figure 1.14, we move the angle at v_1 forward from its initial -45° (the angle of $v_1 v_2$ with respect to the horizontal) to its final 90° , a turn of 135° , while we move the angle at v_3 from its initial -45° to its final -150° by turning backward 105° (rather than forward by 255°).

Now here, as throughout, we have ignored self-intersection of the arm. If that were taken into consideration, this simple-minded interpolation would not

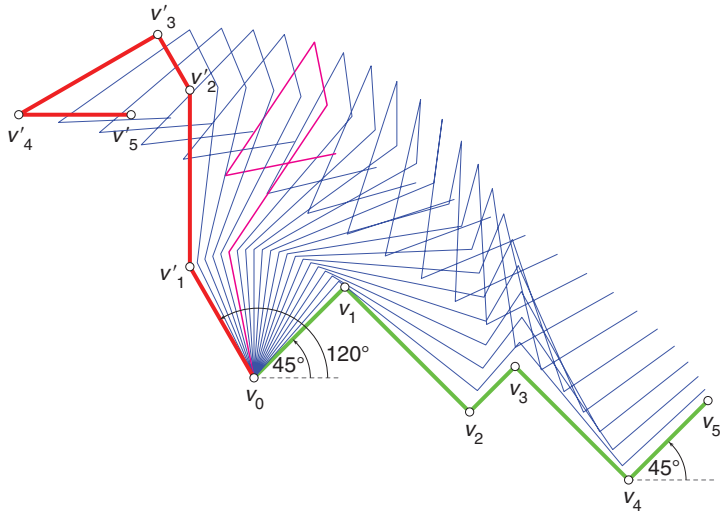


Figure 1.14. Initial angles (green) with respect to horizontal: $(45^\circ, -45^\circ, 45^\circ, -45^\circ, 45^\circ)$; Final angles (red): $(120^\circ, 90^\circ, 120^\circ, -150^\circ, 0^\circ)$. One self-crossing intermediate configuration is highlighted.

suffice, as intermediate configurations self-cross. We touch on this issue briefly in the next section.

Exercise 1.7 (Understanding) Noncrossing Motion. Suppose that instead of measuring angles with respect to the horizontal, angles are measured with respect to the previous link, with the convention that the angle at v_0 is measured with respect to the $+x$ axis (the horizontal to the right). So the initial angles in Figure 1.14 are $(45^\circ, -90^\circ, 90^\circ, -90^\circ, 90^\circ)$ and the final angles $(120^\circ, -30^\circ, 30^\circ, 90^\circ, 150^\circ)$. If these angles are linearly interpolated, will the chain self-cross at some intermediate configuration?

1.3

Above & Beyond

In this section, we look at more realistic and complex robot arm tasks. Beyond the simple reachability questions we have explored, matters become significantly more complicated.

Stowing a Robot Arm (Ruler Folding). Suppose we want to stow an n -link robot arm into a small space. Such stowing is needed, for example, for the robot arms used on the Space Shuttle and International Space Station. The natural method is to fold it flat, perhaps alternating the joint angles between fully turned clockwise to fully turned counterclockwise, aligned and anti-aligned links. However, this only produces a compact configuration if the links are all about the same length. For an arbitrary n -link arm, it is less clear how to fold it flat compactly.

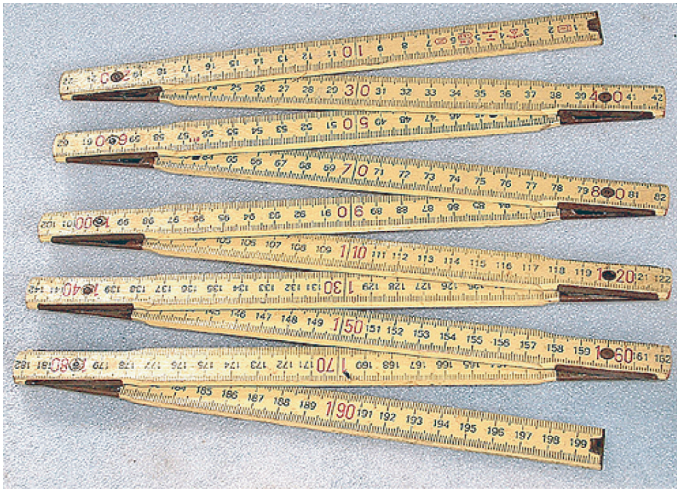


Figure 1.15. A Carpenter's ruler.

An alternative formulation of the problem is obtained by viewing the arm as a strange carpenter's ruler (Figure 1.15), with measuring segments of differing lengths. We want to fold the ruler flat so that, end to end, it has the smallest total length possible for its link lengths. We are insisting that it be completely flat, having effectively no thickness.

This is not a reachability question, which only specifies where the hand v_n is to be located. Instead, here we specify this particular, flat, minimal-length configuration as the goal. See Figure 1.16 for a challenging example.

It turns out this problem is extremely difficult to solve in general. There is a technical term for this level of difficulty: The problem is said to be *NP-complete*. Effectively this means that, for sufficiently large n , no one knows a practical method for deciding whether or not a given arm/ruler can fold to less than a prescribed length. There is a method – try all possible foldings – but for large enough n , this is infeasible. For example, if we had a computer spend only one nanosecond (one billionth of a second) trying each possibility, it would run for more than a trillion years before yielding an answer for $n = 100$, for it must test all $2^n = 2^{100} \approx 10^{30}$ foldings. This is not to say that every instance of the ruler-folding problem is difficult: Particular instances may be easy to solve. But no one knows of an efficient method for solving an arbitrary instance.

Exercise 1.8 (Understanding) Ruler Folding. Suppose a 5-link ruler has lengths $A_5 = (23, 15, 16, 17, 9)$. What is the overall smallest length into which it can be folded?

Obstacles. Generally there are obstacles within the workspace of the arm that need to be avoided; for example, the very object on which the arm is working. Even if we continue to let the arm self-cross, computing reachability

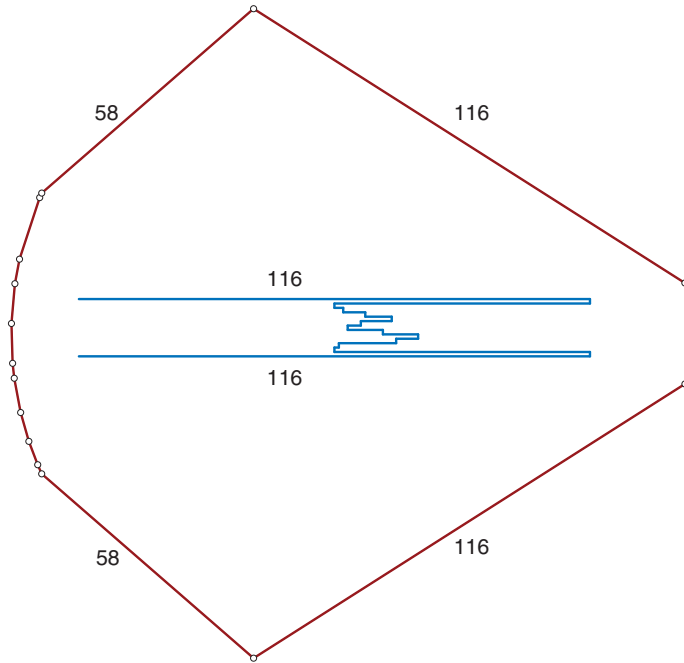


Figure 1.16. An unfolded arm/ruler of $n = 14$ links (red), of lengths (116, 58, 1, 13, 5, 8, 8, 3, 7, 6, 5, 2, 58, 116) counterclockwise. It can fold flat (blue) to be only 116 units long by carefully arranging the short middle ten links.

while avoiding collision with obstacles is again extremely difficult, reaching an even higher level of difficulty than NP-complete, known to the experts as *PSPACE-complete*, another intractability (measure of hardness) classification. The consequence under this level of difficulty is essentially the same as under NP-complete: There is no known practical method for solving a reachability question for n -link arms where n is large.

The same holds true if we have no obstacles but disallow self-intersection: Reachability in 3D is again PSPACE-complete, although the technical difficulty of this task in 2D remains unresolved today.

However, all these negative results hold only for arbitrary n -link arms for large n . And most robot arms have n at most 10. (The relevant number in 3D is not so much the number of links, but rather the number of *degrees of freedom*. The robot arm in Figure 1.1 has 3 links but 6 degrees of freedom.) For such “small- n ” arms, a rich collection of ad hoc techniques have been developed that permit solving all the problems we have considered. Solving them is challenging, but not out of reach. In fact these problems are being solved on a routine basis every day by robot design manufacturers.

You might wonder why researchers care about pinning down the precise difficulty of various tasks for n -link arms for arbitrarily large n when usually $n \leq 10$.

There are at least two reasons. The first is to gain a deep understanding of a theoretical problem for its own aesthetic sake, and also because, as I emphasized in the Preface, time and again theoretical knowledge has later found practical application. The second is that there are robot arm-like structures for which n is indeed large. One example is snake or serpentine robots, used for search-and-rescue missions or for surgery, which might have $n = 30$. A rather different and quite important example is the backbone of a protein molecule, where values of $n = 10,000$ are reached. The intractability results then force researchers to use clever approximations. We will touch on this topic in Section 3.2.

Exercise 1.9 (Challenge) *2D Angle-Limited Linkages: One Constraint.* Let a 2-link arm have link lengths $(1, r)$, where r is some fixed number greater than 0. Suppose the first link may only turn within a 90° range, but the second link is free to rotate the full 360° . To be specific, let us say that the first link makes an angle at the shoulder v_0 to the $+x$ -axis of between 0° (horizontal) and 90° (vertical). Draw and describe the reachability region of the endpoint v_2 of the second link, under three conditions: (1) $r \leq \sqrt{2}/2$, (2) $\sqrt{2}/2 < r \leq 1$, and (3) $r > 1$.

Exercise 1.10 (Challenge) *2D Angle-Limited Linkages: Two Constraints.* Continuing the previous problem, also constrain the v_1 joint to only turn within a 90° range. To be specific, the angle “ $v_0v_1v_2$ ” is between 90° (perpendicular to v_0v_1) and 180° (aligned with v_0v_1). Again draw and describe the reachability region of the v_2 endpoint. Are there critical values of r at which the structure of the reachability region changes?

2

Straight-Line Linkages and the Pantograph

The robot arm / polygonal chain we studied in Chapter 1 is among the simplest of linkages. It is fundamentally linear – one link after another. Creating cycles (loops) in a linkage constrains its possible movements and simultaneously renders it more useful and more difficult to analyze. More useful because constrained movements can form the basis of a variety of mechanisms, as we will soon see. More difficult to analyze because the motions of joints are determined by several interacting constraining equations. In this chapter, we recount the pursuit of a linkage that can “draw a straight line,” and analyze the simple but useful pantograph linkage using vectors. There are fewer theorems and more “stories” in this chapter – something of an interlude for the reader between the heavy lifting in the previous chapter and that to come in the next chapter.

2.1

Straight-Line Linkages

Although linkages have been used since medieval times – in saw mills, in mechanical clocks, in looms, in printing presses – their golden age was the eighteenth century, driven by the demands of the steam engine, which powered the Industrial Revolution. One particular need was for a mechanism to constrain a piston rod to move along a straight-line path within the steam-pressured hydraulic cylinder; see Figure 2.1. Any deviation from straight-line motion induces lateral forces that quickly wear down the rubbing parts. This led James Watt, after whom the unit of power (wattage) for a light bulb is named, to patent in 1784 the clever 3-bar linkage shown in Figure 2.2(a).

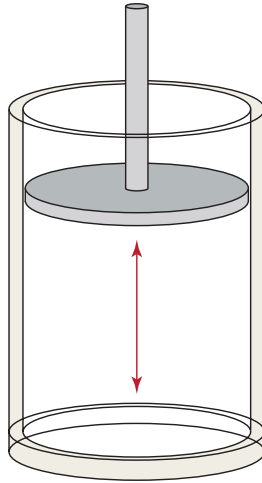


Figure 2.1. Piston and steam cylinder. Ideally the piston rod should move along a vertical line.

The joints x and y are *pinned* – fixed to the plane but free to rotate about their fixed centers. Joints a and b move on equal-radii circles centered on x and y respectively, whereas point c (not a joint), the midpoint of ab , moves passively along between. It is point c that moves nearly on a straight line.

This makes some intuitive sense, in that joint a is pulled leftward and b rightward on their circles, with c staying approximately balanced between. However, the actual motion of c is rather intricate: c follows a figure-8 curve known as a “lemniscate,” which, in this instance, has two long, nearly straight sections, shown in (b) of the figure. Positioning the joints x and y so that the rotations about them are not too large ensures that c remains on just one of those sections.

Watt was aware that the central motion was not precisely straight, but it sufficed for his purposes, and was employed successfully in steam engine designs. Although today we celebrate Watt for his pioneering improvements to the steam engine, his own assessment valued this “parallel motion” linkage above all his other accomplishments: “I am more proud of the parallel motion than of any other mechanical invention I have ever made.”

Exercise 2.1 (Understanding) *Watt Linkage Angle Range.* Suppose the three bars of a Watt linkage are all the same unit length, $|xa| = |ab| = |by| = 1$; here the notation $|xa|$ represents the length of the segment whose endpoints are x and a . (Earlier we used $||$ for the absolute value of an enclosed number. Both uses are common in mathematics.) And suppose the circle centers are placed as illustrated in Figure 2.2: separated horizontally by $|xa| + |by| = 2$ and vertically by $|ab| = 1$. Compute the angle range for joint a on its circle: How far can

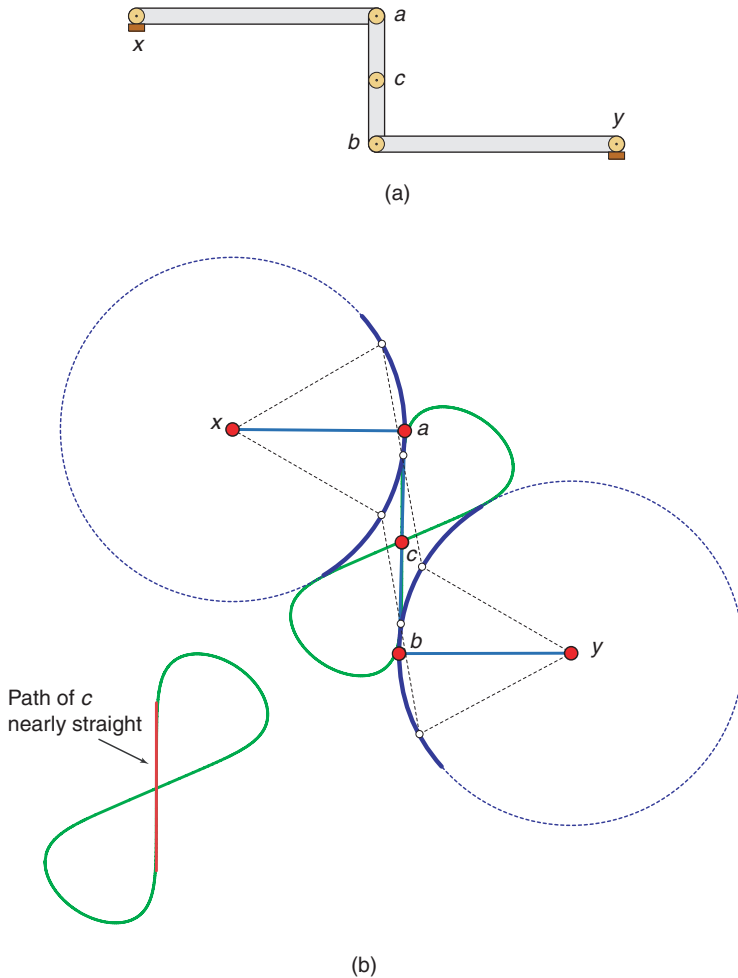


Figure 2.2. (a) Basic Watt linkage. (b) Dynamics of point c .

the link xa turn counterclockwise about x , and how far clockwise? [Requires trigonometry.]

Although Watt's linkage addressed the practical need for straight-line motion, the quest for a linkage that achieved exact straight-line motion continued. The reason this is so difficult to achieve is that linkage motion is fundamentally a combination of circular motions: Each link endpoint follows a circle centered on the joint at its other end. To achieve straight-line motion requires somehow playing off several moving circular motions against one another to miraculously result in some point on the linkage moving in a perfectly straight line.

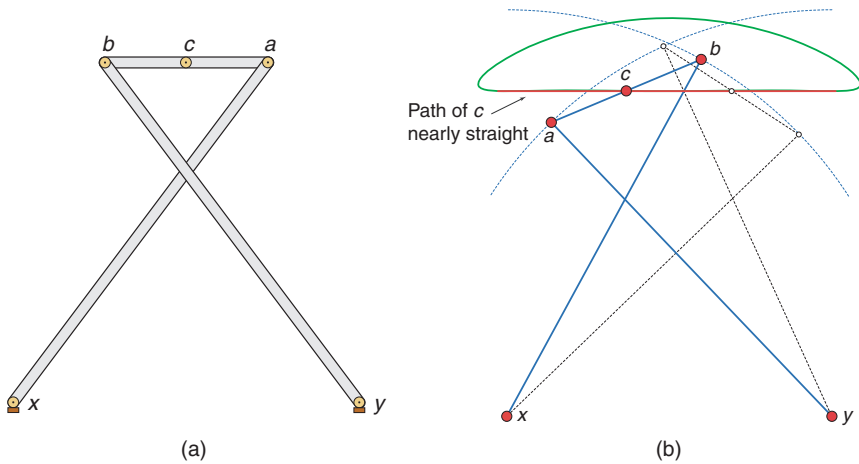


Figure 2.3. (a) Chebyshev linkage: $|xa| = |yb| = 5$, $|ab| = 2$; (b) The trace of point c is nearly straight within a portion of its locus.

The next significant step was taken by the Russian mathematician Pafnuty Lvovich Chebyshev, who was fascinated by linkages. In the 1850s, he invented another approximate straight-line drawing linkage, shown in Figure 2.3(a). It is essentially Watt's linkage crossed, but requires specific lengths to function appropriately, and it behaves rather differently, as shown in (b) of the figure. It achieves an even closer approximation to a straight line than does Watt's linkage. Chebyshev used his linkage in a "Foot-Stepping Machine" that was displayed at the 1878 World's Fair in Paris. His mathematical work toward straight-line linkages led to what are now known as "the Chebyshev polynomials."

The elusive goal of exact straight-line motion was finally achieved in 1864 by Charles-Nicolas Peaucellier, a captain in the French army. He received a prize, the Prix Montyon, from the Institute of France for his discovery. A student of Chebyshev, Lipman Lipkin, independently discovered the same mechanism in 1871, and so it is sometimes called the Peaucellier-Lipkin linkage.

This remarkable linkage is illustrated in Figure 2.4(a). Joints x and y are pinned, and rods xa and xb keep joints a and b on circle C_x , and rod yc keeps joint c on circle C_y , which also includes x (so $|xy| = |yc|$). The mechanism is "driven" by moving c on C_y , which causes the rhombus $abcd$ to expand and narrow symmetrically about the diagonal ab . As the cell $abcd$ is rotated clockwise, the gap between the circles C_x and C_y narrows, squeezing the cell, amazingly just enough so that joint d stays precisely on a vertical line L ! Establishing this mathematically is a bit involved, and we will not attempt it.

When Lord Kelvin (after whom our low-temperature unit is named) worked a model of the device, he is reputed to have remarked that "it was the most beautiful thing he had ever seen."

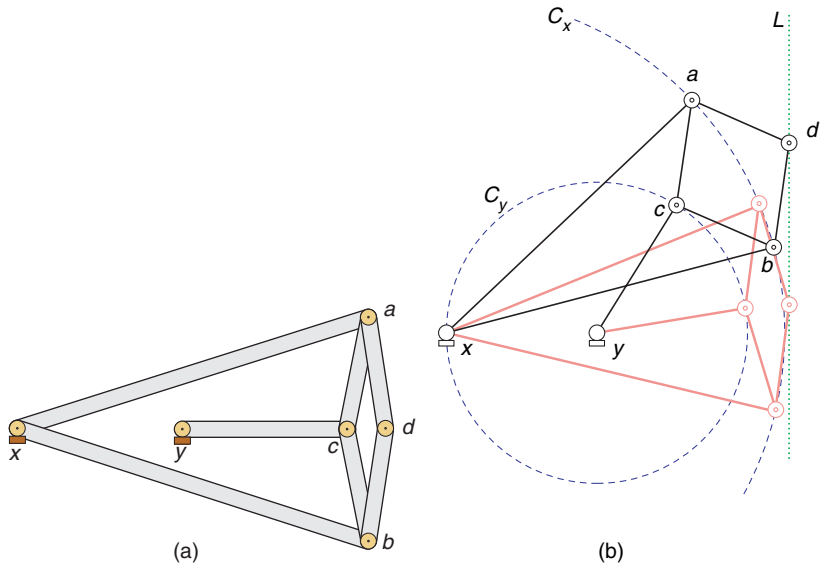


Figure 2.4. (a) Peaucellier linkage. (b) Geometry of linkage. x and y are pinned; d moves along vertical line L .

2.2 Pantograph

From the 100-year struggle to find a linkage that draws a straight line, the reader may have drawn two conclusions: (1) the motion of even simple linkages is subtle and difficult to analyze mathematically; and (2) this is all ancient history. The first conclusion is correct, but the second is not. Although there is no question that the need for linkages has diminished with the rise of electricity and then electronics, linkages still abound. Besides the automotive crankshaft mentioned in the Preface, every car contains a clever 4-bar linkage (Figure 2.5) to convert the rotary motion of a motor to the familiar slap-slap back-and-forth arc of the windshield wipers. Many other everyday examples may be noticed by the discerning eye.

Returning to the first conclusion, perhaps the most useful linkage whose motion is simple enough to permit a clean and thorough mathematical understanding is the *pantograph*. This word has two meanings today: an instrument for copying drawings at reduced or enlarged scale, or a linkage with similar structure and a variety of uses. An example of the latter is the mechanism on top of an electric train designed to maintain electrical contact with the overhead power wires; see Figure 2.6. But here we concentrate on the former meaning, a copying linkage whose functioning is more straightforward than the straight-line linkages we've seen in the previous section.

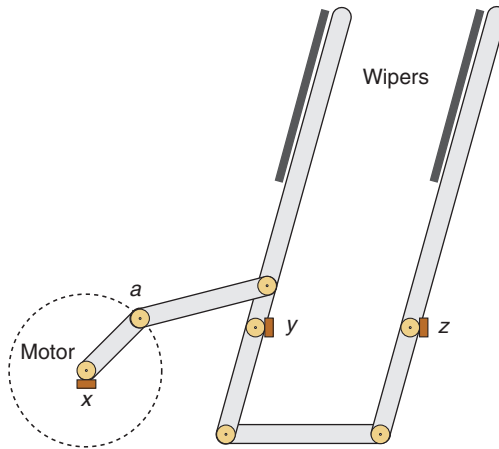


Figure 2.5. Windshield wiper linkage. Joints x , y , and z are pinned. Joint a is driven around a circle by a motor at x , causing the wipers to rock back and forth in parallel.



Figure 2.6. Train pantograph.

A pantograph consists of four bars, two long and two short. We start by assuming both long bars are of length 2 and both short bars of length 1. Let xc and cz be the two long bars, as in Figure 2.7. The short bars are attached with swivel joints to the midpoints a and b of the long bars, and arranged so that $acby$ is a parallelogram (which happens to be a rhombus in this particular figure). If joint x is pinned to the plane, this linkage has the property that whatever curve is traced by point y is duplicated at twice the scale by the movement of point z .

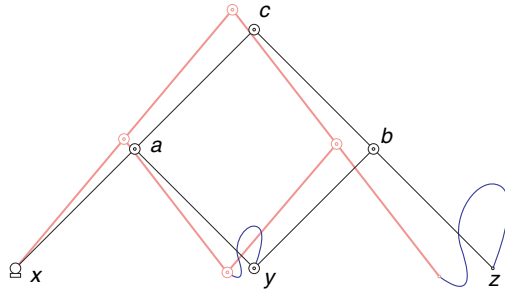


Figure 2.7. A pantograph. Joint x is pinned. The movement of joint y is duplicated and doubled by point z .

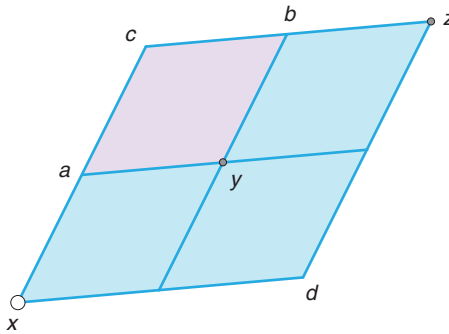


Figure 2.8. Extending the pantograph to nested parallelograms.

Thus it is an enlarge-by- $2\times$ mechanism, or, reversing the logic (letting point z drive y), a diminish-by- $\frac{1}{2}\times$ device.

Later we will answer the question of why anyone would need such a device in the age of copiers with magnification buttons. For now we seek a detailed understanding of how it works.

A basic intuition can be attained by adding two more long bars to complete the outer parallelogram, as shown in Figure 2.8. Because the inner parallelogram $acby$ is half the size of the outer parallelogram $xczd$, and because they move in concert, it is natural that the inner movement is at half the scale of the outer movement. To convert this intuition into a more precise statement, we turn again to vectors (the reader may want to revisit Box 1.3).

Let \vec{A} and \vec{B} be the two vectors along the linkage which, placed head-to-tail starting at x , reach y . As in Chapter 1, these vector displacements can be calculated by subtracting the coordinates of the relevant points (Figure 2.9):

$$\vec{A} = a - x$$

$$\vec{B} = y - a$$

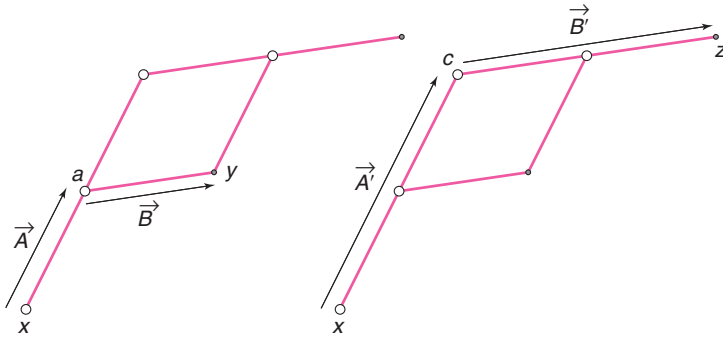


Figure 2.9. $\vec{A} + \vec{B} = y - x$ and $\vec{A}' + \vec{B}' = z - x$.

$$\begin{aligned} \vec{A} + \vec{B} &= a - x + y - a \\ &= y - x \end{aligned}$$

Let \vec{A}' and \vec{B}' be the similar vectors that reach z :

$$\begin{aligned} \vec{A}' &= c - x \\ \vec{B}' &= z - c \\ \vec{A}' + \vec{B}' &= c - x + z - c \\ &= z - x \end{aligned}$$

We stipulated that the length of the long and short links were 2 and 1 respectively, so we know that:

$$\begin{aligned} \vec{A}' &= 2\vec{A} \\ \vec{B}' &= 2\vec{B} \end{aligned}$$

Now putting together these three relationships yields:

$$z - x = \vec{A}' + \vec{B}' = 2\vec{A} + 2\vec{B} = 2(\vec{A} + \vec{B}) = 2(y - x).$$

If we place the origin of a coordinate system at $x = (0,0)$, then $z - x = z$ and $y - x = y$. Then the above equation says that the coordinates of z are always exactly double the coordinates of y in this coordinate system. So whatever drawing is traced by y is traced at twice the size by z .

This analysis easily leads to the alterations needed for the pantograph to achieve different scale factors: To achieve a factor of 3, we want $\vec{A}' = 3\vec{A}$ and $\vec{B}' = 3\vec{B}$, as in Figure 2.10(a); to achieve a factor of $1\frac{1}{2}$, we want $\vec{A}' = \frac{3}{2}\vec{A}$ and $\vec{B}' = \frac{3}{2}\vec{B}$, as in (b). Commercial pantographs have marked holes to make it easy to reconfigure the links for different magnifications.

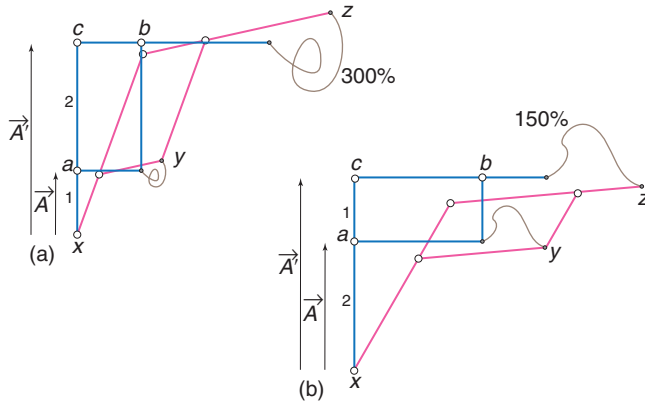


Figure 2.10. Pantographs with different magnifications. (a) $\vec{A}' = 3\vec{A}$: scale factor $\times 3$ (300%). (b) $\vec{A}' = \frac{3}{2}\vec{A}$: scale factor $\times \frac{3}{2}$ (150%).

Exercise 2.2 (Practice) Fivefold Pantograph. Design a pantograph to achieve fivefold magnification.

Exercise 2.3 (Understanding) Two-Thirds Pantograph. Design a pantograph to achieve two-thirds reduction.

Another consequence of our analysis is that points x , y , and z always lie on a common line: They are *collinear*. This follows because $z = ky$, where k is the scale factor. Suppose $k = 3$ and, for specificity, let $y = (1, 2)$ and $z = 3y = (3, 6)$. Then the slope of the line through $x = (0, 0)$ and y is $\frac{1}{2}$ (the ratio of vertical over horizontal displacement), and the slope of the line through x and z is $\frac{3}{6} = \frac{1}{2}$. Clearly these slopes will always be the same, because the coordinates of z are just k times the coordinates of y . Therefore, y and z lie on the same line through x , and so all three are on a common line at all times.

Let us summarize our analysis in a theorem:

Theorem 2.1 (Pantograph)

For a pantograph whose joints are labeled as in Figures 2.7 and 2.10, when x is pinned to the plane, point z traces a scaled version of the path followed by joint y . The scale factor is $|xc|/|xa|$, the ratio of the lengths of the long links to that of the short links.

Although pantographs are no longer needed for enlarging drawings, they are still used in machining parts, and in particular for etching and engraving. See, for example, Figure 2.11.

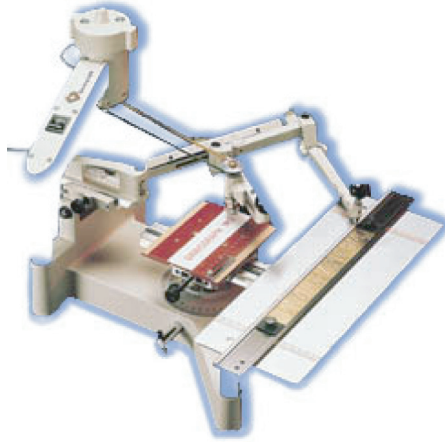


Figure 2.11. Industrial pantograph built by Gravograph Ltd. Model IM3.

The fine lettering on the inside of a wedding ring is almost invariably produced by a pantograph-tracing stylus that runs over the text at a large scale, linked to the engraving tool that follows at a much reduced scale.

Although pantographs remain in use today, some of the most interesting applications lie in the past. For example, Thomas Jefferson used a pantograph to make copies of his extensive correspondence. Figure 2.12 shows the “polygraph” on display at Monticello, his home in Virginia. Here the goal is not magnification, just duplication, which slides toward the more generic use of the word “pantograph.” We can view duplication as a limiting case of the pantograph as presented in Figure 2.10 as the magnification scale factor k gets smaller and smaller, approaching 1 from above. In Figure 2.13(a), the scale factor is slightly larger than 1, and an extension of y to y' separates the controlling point y' from the follower point z . The limit of this process, achieving exact duplication, is simply a parallelogram (b), echoes of which are visible in Jefferson’s pantograph.

A most unusual application of a pantograph is provided by the 18th-century chess-playing automaton known as “The Turk” (Figure 2.14). The inventor, Wolfgang von Kempelen, managed to astonish the courts of Europe, many of whose members were only too ready to believe that a clockwork mechanism could achieve such wizardry. Others were convinced it was a trick, but no one guessed the precise mechanism in over twenty years of touring. Among the skeptics was Charles Babbage, arguably the inventor of the first computer, whose encounter with (and loss to) the Turk helped trigger a train of thought that eventually led to his notion of a truly general-purpose computer. As the skeptics suspected, there was an operator hidden inside, one (among the several who served in that role) a Viennese chess master. But even among those convinced of fraud, the mechanism of deceit eluded them. The heart of the mechanism was a pantograph that permitted the hidden operator to move a pointer on a small

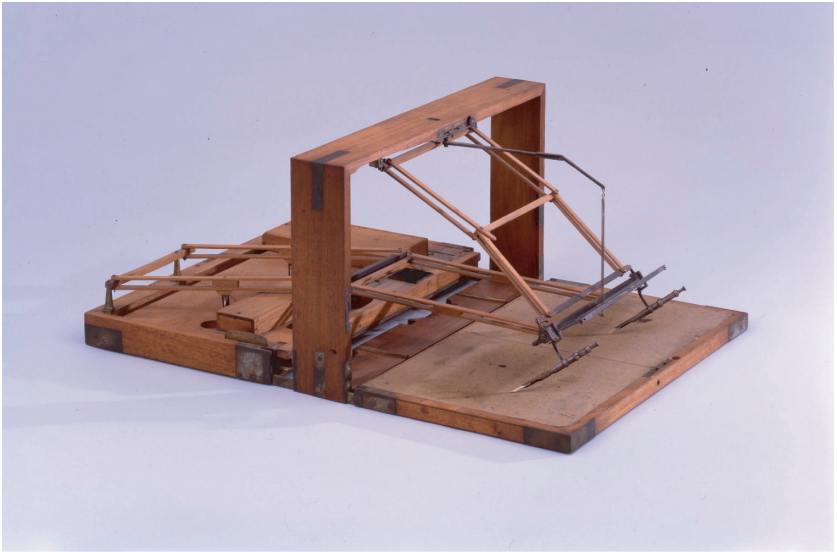


Figure 2.12. Thomas Jefferson Polygraph (i.e., pantograph), 1804. Slanted pen holders are visible, as are parallelograms.

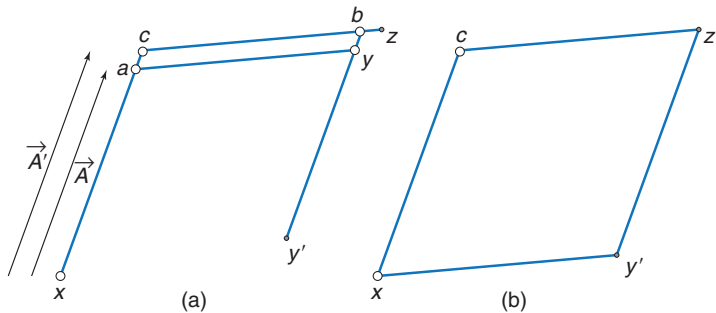


Figure 2.13. (a) A pantograph with scale factor $k = 1.1$; (b) The limiting parallelogram. The bottom bar xy' serves to brace the structure into two pairs of parallel bars, effectively replacing the bracing provided by ay and by in (a). Now z and y' trace exact duplicate curves.

interior chess board, and have his or her moves duplicated at slightly larger scale by the Turk above via a pantograph traveling up his sleeve. A twist below closed the gloved fingers of the Turk's hand on a piece, a counter twist released his grip. The effect was apparently mesmerizing.

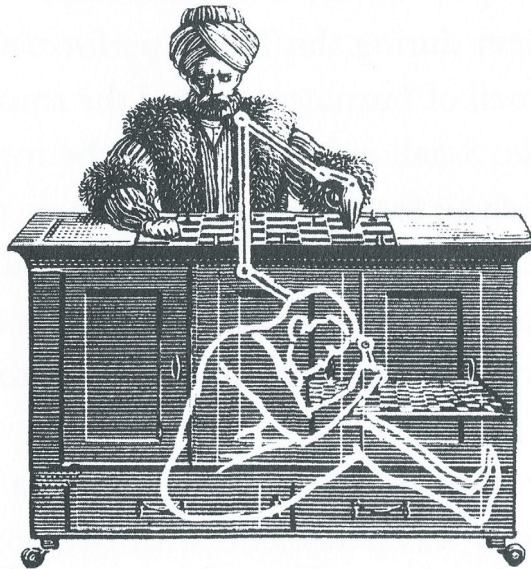


Figure 2.14. The Turk, with pantograph linkage traveling down his left arm.

2.3 Above & Beyond

In light of the difficulty of designing a linkage to draw a straight line, it was audacious of the barrister and amateur mathematician Alfred Bray Kempe to explore in 1876 whether there is a linkage that will trace *any* given curve. The goal is to have one driver joint (like y in the pantograph) and one joint identified as the writing stylus (like z in the pantograph). The curve to be drawn is arbitrary, except that it must be bounded within a finite region of the plane (an obvious necessary condition), and he stipulated that it must be definable by a collection of algebraic equations. Restricting to algebraic curves excludes $y = \log x$ and its ilk, but admits a wide variety of complex and beautiful curves, including those that approximate nonalgebraic curves arbitrarily closely.

Kempe provided a beautiful and intricate solution, but over time it was realized that there were technical flaws in the proof, which took more than a century to completely resolve. Today Kempe's result is accepted as a theorem, with several independent proofs, including a patched version of his original proof. Because the algebraic curve may have the cusps and multiple pieces characteristic of handwriting, the result is often phrased today as:

Theorem 2.2

There is a linkage that signs your name! (Figure 2.15).

The proofs of “Kempe’s Universality Theorem” are constructive but would produce such complex linkages, that I doubt if they have ever been applied to actually create a signing linkage. To give a hint of the possibilities, Figure 2.16 shows a linkage that signs a crude ‘J’, hardly the ‘J’ of John Hancock!

So far we have not made a distinction between linkages whose links never cross one another and those where they might. The former are *planar linkages*, in that they may be viewed as lying in a plane. Watt’s linkage (Figure 2.2) is planar, whereas Chebyshev’s linkage (Figure 2.3) is not. Chebyshev’s linkage must be built on two levels so that the yb link can slide over the xa link. Similarly, the Peaucellier linkage (Figure 2.4) needs two levels to accommodate



Figure 2.15. There is a linkage that traces a thin version of this collection of curves.

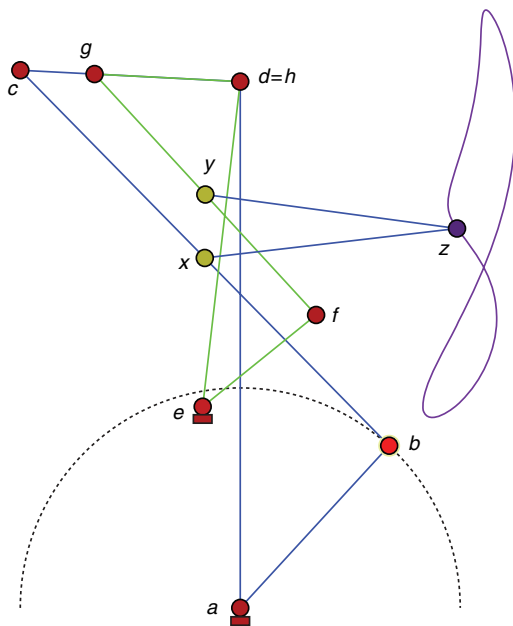


Figure 2.16. Joints a and e are pinned to the plane. Joint b is the driver, moving around a circle centered on a . Both $abcd$ and $efgh$ are *contraparallelograms*, parallelograms with one corner flipped over the diagonal. x and y are midpoints of bc and fg respectively. Point z is the writing stylus.

links xa and xb sliding over link yc . The complex constructions lying behind the signs-your-name theorem (Theorem 2.2) generally produce nonplanar linkages. It is unknown if the same generality can be achieved solely with planar linkages:

OPEN PROBLEM: *Planar Signing (General Case)*
Is there a planar linkage that signs your name?

As this may be a challenging problem, let me pose a simpler, specific, hopefully more accessible version, which I label an open “subproblem” because it is only a small piece of the larger issue:

OPEN SUBPROBLEM: *Planar Signing Digits*
Are there a planar linkages that draw representations of each of the ten digits, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9?

We could say that the Peaucellier linkage solves the problem for the digit 1 by restricting its range of motion to avoid link crossings. The Watt linkage has the digit 8's underlying shape as the locus of point c , but c cannot reach all parts of that locus, as Exercise 2.1 showed. There is a linkage invented by Frans van Schooten, a 17th-century Dutch mathematician, that draws an ellipse, which we could interpret as the digit 0, but the linkage seems fundamentally nonplanar. So, with only the digit 1 solved, there are nine challenges remaining!

3

Protein Folding and Pop-Up Cards

We conclude our exploration of linkages by returning to the polygonal chains that were the focus of Chapter 1, but now with angle constraints. If you struggled through Exercises 1.9 and 1.10, you know that angle constraints greatly complicate the possible motions of the chain. But in many applications, there are angle constraints, so they must be confronted. We consider two applications in this chapter, which are, surprisingly, related: protein folding and a certain pop-up card. Despite the whimsical chapter title, the real focus will be the “maxspan of 90° -chains,” the mathematical structure shared between the two applications. Several techniques and ideas from the previous chapters will resurface here, including induction and the triangle inequality.

3.1 Fixed-Angle Chains

In both the robot-arm polygonal chains we talked about in Chapter 1 and the linkages in Chapter 2, all joints are “universal,” meaning that there is complete rotational freedom at each joint. In this chapter, we explore open polygonal chains with the angle fixed at each joint, which are called *fixed-angle chains*. You might think that fixing the angle is the opposite extreme of allowing universal motion, and initially might seem to totally rigidify the chain into one fixed configuration. But in fact, motion is still possible in 3D: What is called *revolute* or *dihedral* motion.

Look at the 3-link chain in Figure 3.1(a), with both joint angles fixed to 90° . We’ll call this a 90° -*chain*. The first and third links can revolve about the middle link while still maintaining 90° at the joints, as shown in (b) of the figure. You

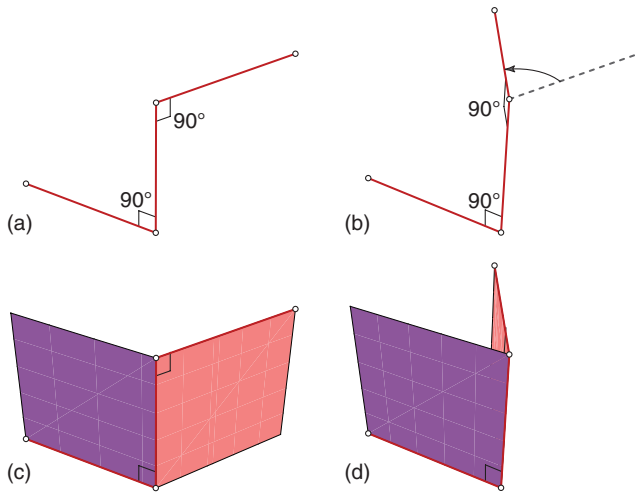


Figure 3.1. Dihedral motion of a 3-link 90° -chain. The 90° angle between adjacent links remains fixed, while the two planes swivel on the shared middle link.

can think of this revolute motion as the swinging motion of two doors sharing a common hinge along the middle link, say a front door and a swinging screen door; see (c-d) of the figure. This is called “dihedral motion” by molecular biologists, who view it as the relative motion of the two planes containing the doors: two (*di-*) planes (*-hedral*). The angle between the planes/doors is not fixed, but the angle at the joints between the two links remains 90° . The reason molecular biologists care about fixed-angle chains is that they serve as a model of protein backbones, to which we now turn.

3.2 Protein Backbones

A protein molecule is a long chain of atoms with short, attached *side chains* consisting of clusters of several atoms each. The central *backbone* consists of repeated copies of the common 3-atom core of an amino acid. Amino acids are fundamental building blocks of life; they come in twenty different varieties but all sharing a common core. The sequence of atoms along the backbone consists of nitrogen (N) and carbon (C) atoms, three per amino acid: NCC NCC NCC... . A typical protein is constructed from between 100 and 1,000 amino acids, although some (e.g., the muscle protein titin) contain as many as 30,000 amino acids.

Proteins fold into complex 3D shapes, as you can see from Figure 3.2, which shows a small protein of 36 amino acid bonds strung along the highlighted backbone. Although we could imagine many shapes into which a given protein might fold, the chemical properties and biological environment conspire to ensure that each folds nearly unerringly into one particular shape, its so-called *native state*.

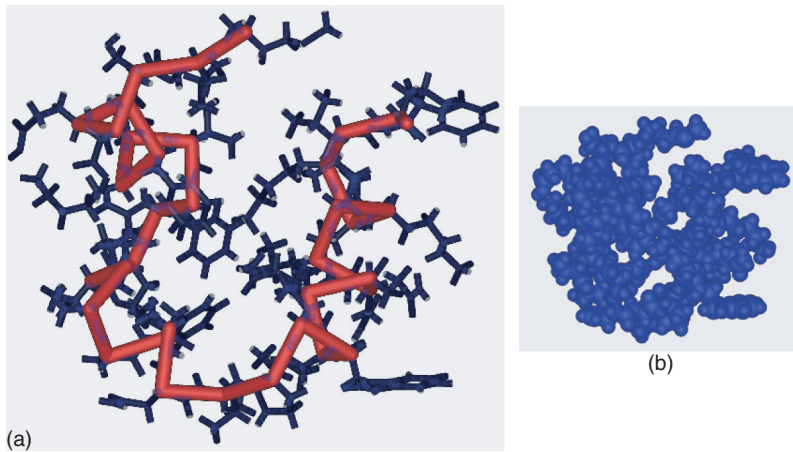


Figure 3.2. A small structural protein, a variant of the “villin headpiece.” [2PPZ in the Protein Data Base]. (This protein “manages” the actin molecule, the engine of our muscles.) (a) Backbone of 36 amino acid cores highlighted (red), each consisting of three atoms, CCN, surrounded by bonds indicating the side chains. (b) Sphere model of the same molecule, with each atom a sphere.

Misfoldings are rare and are at the root of certain diseases, such as Mad Cow disease. These misfoldings aside, the 3D native state is completely determined by the sequence of different amino acids along the chain. The great unsolved “protein folding problem” is to predict the 3D native state from the one-dimensional sequence of amino acids.

The reality of protein folding is extremely complicated, not only geometrically, but also in terms of the chemical and electrostatic bonds, the biology of the cell environment, and the physics of molecular dynamics. A gross simplification of this messy situation will lead us to the purely geometric motion of a fixed-angle chain. The simplification occurs in four steps. First, we will ignore all side chains, leaving only the backbone of the protein, a linear chain of 3-atom cores of amino acids. Comparing Figure 3.2(a) and (b) shows what a significant (and unrealistic) simplification this is. Second, we will treat the bond angles between adjacent atoms along the backbone as fixed. This is nearly true. Third, we assume the chain permits free dihedral motion about each of its bonds. This is definitely not true, because one bond per amino acid (the so-called “peptide” bond) only permits two dihedral angles, 0° and 180° . Fourth, we will ignore all chemical and electrostatic forces, leaving only the geometry of dihedral motions and the restriction that the chain cannot pass through itself.

These four simplifications reduce the protein in Figure 3.2 to a chain of $3 \times 36 = 108$ links: three links per amino acid core. Figure 3.3 shows a closeup of a small portion of the backbone and the corresponding chain of links.

Now we have left messy reality so far behind that fixed-angle chains hardly deserve to be called “models” for protein molecules. At most we can say they

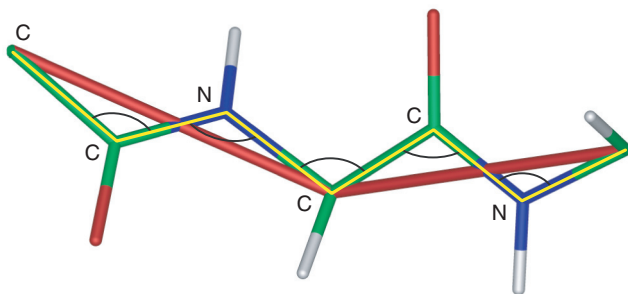


Figure 3.3. Two CCN amino acid cores along the backbone (red), with most side-chain atoms suppressed. The chain and bond angles are superimposed.

are inspired by proteins. And we will see that their possible motions are already quite intricate.

These four simplifications lead us to study fixed-angle polygonal chains. Two further simplifications focus in on a particularly interesting class of these chains. First, we assume the fixed angle is 90° at every joint. The actual bond angles along a protein backbone lie roughly in the range of 109° to 122° , so this is a rough approximation of the reality of protein chains. Second, we assume that the link lengths are all the same, which we take to be 1. The actual bond lengths between atoms along the backbone of a protein vary a bit, but not much, roughly between 1.33\AA and 1.52\AA .

So we have arrived at *unit 90° -chains* – each link of length 1, each joint angle fixed at 90° – and we will analyze these throughout the remainder of the chapter. We will occasionally contrast these unit 90° -chains with variants without these two final assumptions – allowing different link lengths, or angles different from 90° , but the unit 90° -chains will remain our focus.

3.3 Maximum Span

Chemists have long been interested in the statistical distribution of the end-to-end lengths of polymers, a class of chain-like molecules that includes proteins and plastics. To understand the distribution, they need to know the maximum possible end-to-end distance, known as the *maximum span*, or *maxspan* for short, of the chain. Our goal is now to understand which configurations of a 90° unit chain achieve the maxspan.

If we didn't assume that all angles are 90° or that all lengths are the same, the problem is complex, but now almost completely understood through recent advances (see Further Readings, p. 143). Complications arise because the maxspan is generally achieved by a fundamentally 3D configuration, consisting of

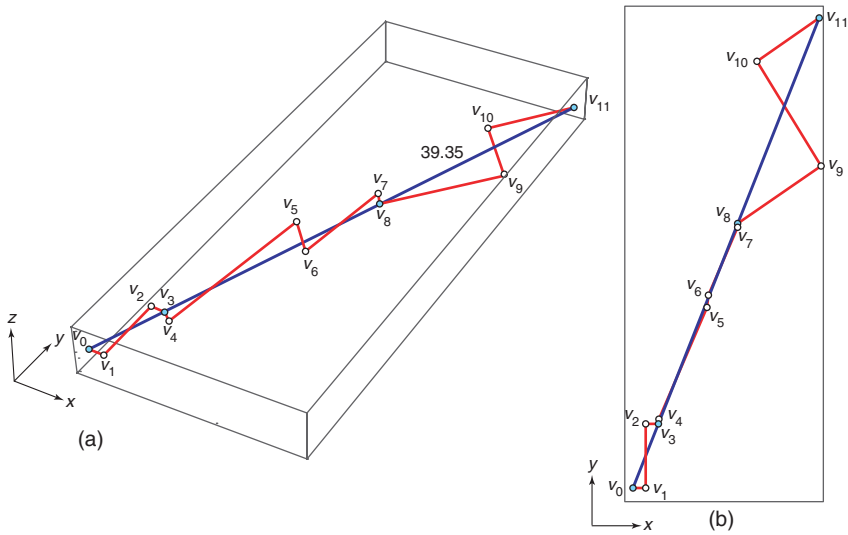


Figure 3.4. An 11-chain in maxspan configuration. Its link lengths are (1,5,1,1,10,3,6,1,8,10,6), and its span is 39.35. Vertices on the span segment $s = v_0v_{11}$ are marked. (a) Oblique view; (b) Overhead view.

twisted sections aligned along the line determined by the two end vertices of the chain.

Figure 3.4 shows an example that drops the unit-length assumption but retains the 90° -angle assumption. This is an 11-link 90° -chain in a maxspan configuration, whose link lengths vary from 1 to 10. Let its vertices be labeled $(v_0, v_1, \dots, v_{11})$. The span is the segment $s = v_0v_{11}$, in this case of length about 39.35. The vertices v_3 and v_8 lie directly on s , partitioning the chain into three *planar* (i.e., flat) sections. That the sections are planar is more evident in the overhead view shown in (b) of the figure: the two end sections $\{v_0, v_1, v_2, v_3\}$ and $\{v_8, v_9, v_{10}, v_{11}\}$ lie in a plane parallel to the xy -plane, and the middle section $\{v_3, v_4, v_5, v_6, v_7, v_8\}$ lies in a vertical plane parallel the the z -axis.

Exercise 3.1 (Practice) 3-Link MaxSpan. What is the maxspan of the 3-link 90° -chain with link lengths (1,2,3), if all three links lie in the same plane?

Exercise 3.2 (Challenge) 3-Link MaxSpan. Can you increase the span of the (1,2,3) linkage from the previous exercise by rotating the 3rd link out of the plane containing the first two links?

A natural alternative to the twisted configuration in Figure 3.4 is the planar *staircase configuration*, which simply alternates left and right turns at each joint, maintaining the chain in a plane. Figure 3.5 illustrates this for the same 11-link

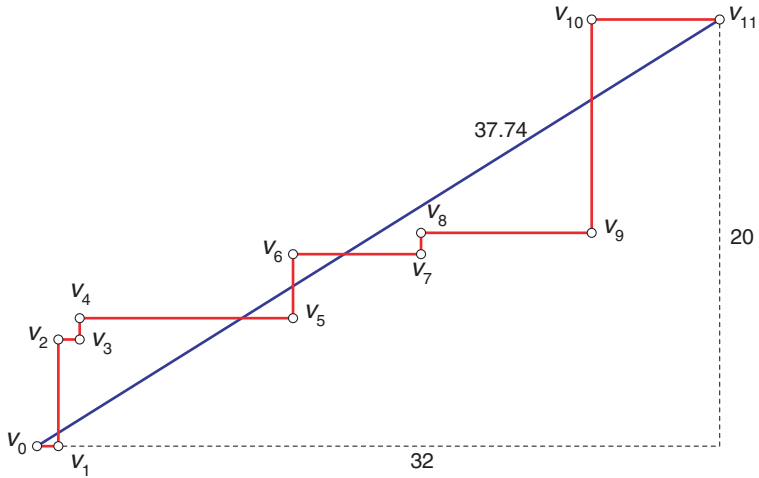


Figure 3.5. The chain from Figure 3.4 laid out in the staircase configuration has a shorter span: $\sqrt{32^2 + 20^2} \approx 37.74$.

chain, and reveals that this planar configuration has a smaller span: 37.74 vs. 39.35 for that in Figure 3.4.

These complications disappear if we reinstate both the 90° assumption and the unit-length assumption, for the maxspan configuration of a unit 90° -chain is always flat:

Theorem 3.1 (Unit-90)

The maxspan configuration of a unit 90° -chain of n links is achieved by the planar staircase configuration.

Figure 3.6 illustrates staircase configurations for unit 90° -chains.

The fact that the claim of this theorem is FALSE if the links are not the same length shows that there must be some subtleties here. Fortunately those subtleties can be skirted for these specific chains. We now sketch a proof of the Unit- 90° Theorem 3.1, first discovered and proved by a college student, Nadia Benbernou. The proof takes a path through two concepts, both illustrated in Figure 3.4: alignment and piercing.

3.4 Alignment

We start with a lemma on alignment, which says, essentially, that if we align two chains end to end, each of which is in maxspan configuration, then the combined

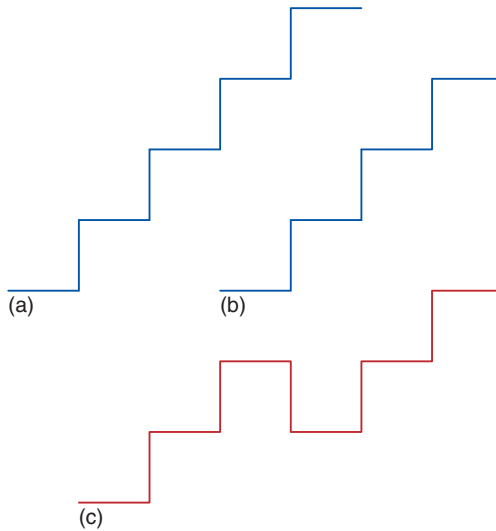


Figure 3.6. (a,b) Staircases; (c) Not a staircase, because the turns do not alternate.

chain is in maxspan configuration. This lemma is very general, because it does not rely on either the 90° assumption nor on the unit-length assumption.

Lemma 3.1 (Alignment)

If C_1 and C_2 are both in maxspan configuration, and if joining their spans end-to-end along the same line meets the joint-angle constraint at the join, then the new chain C is in maxspan configuration.

Proof: We opt here for aiming for the intuitive essence of the proof: the *triangle inequality* that we employed in Chapter 1 (Box 1.4). This states that the length of one side of a triangle is at most the sum of the lengths of the other two sides. If the three triangle side lengths are a , b , and c , the claim of the triangle inequality is that $c \leq a + b$.

Now think of a and b as the maxspans of C_1 and C_2 , and c the span of their join C . By the triangle inequality, $c \leq a + b$. So the largest c could ever be is $a + b$. And that $a + b$ span of C is achieved by aligning the largest spans of C_1 and C_2 . If, when the maxspans of C_1 and C_2 are aligned, the angle at the join vertex is the required fixed angle, then we have achieved a valid configuration of C , realizing the maxspan c . Λ

Alignment is nicely illustrated in Figure 3.4, as it turns out the the three flat subchains (v_0 to v_3 , v_3 to v_8 , v_8 to v_{11}) are each in maxspan configuration, and

the alignment shown manages to maintain the 90° angle at the joints v_3 and v_8 , so the lemma says the whole chain must achieve the maxspan.

3.5 Piercing

We will reach our goal of proving the Unit- 90° Theorem 3.1 by establishing a stronger theorem that will easily imply Theorem 3.1 as a more specific, special case. This relies on a notion that is again nicely illustrated by the 11-link chain in Figure 3.4. The end-to-end spanning segment in that figure, $s = v_0v_{11}$, pierces each link of the chain in order, where *pierce* here means “passes through a point of,” perhaps just an endpoint. So s touches the first link at v_0 (which counts as piercing), passes through the midpoint of the second link, touches the third and fourth links at their shared vertex v_3 , and so on. Let $C = (v_0, \dots, v_n)$ be a chain with vertices v_i . We’ll say that a chain is in a *piercing configuration* if the spanning segment v_0v_n meets every link of the chain in order. The staircase chain in Figure 3.5 is not in piercing configuration, because several links are not pierced by the spanning segment.

Theorem 3.2 (Piercing)

If a chain C is in a piercing configuration, then C is a maxspan configuration.

We will settle for a sketch of the proof to avoid unrevealing details. The proof is by induction on n , where n is the number of links in the chain. (Induction was introduced in Box 1.2, p. 6.) Both 1-link and 2-link chains are automatically piercing, and for each, their only configuration is the maxspan configuration. So the theorem holds for the base of the induction, $n = 1$ and $n = 2$.

Now, following the induction paradigm, we assume the theorem has been established for $n = 1, 2, \dots, n - 1$ links, and using this knowledge, we seek to establish it for n links. If we are successful, we establish the theorem for all n .

Let $C = (v_0, \dots, v_n)$ be some piercing configuration of the n -link chain, with spanning segment $s = v_0v_n$. We aim to show that C must be a maxspan configuration. Because the configuration C is piercing, we know that s pierces the penultimate link $v_{n-2}v_{n-1}$ of C . Now there are several cases depending on where s pierces this link. However, looking at the two cases illustrated in Figure 3.7 should suffice to convince the reader.

In the first case (a), s pierces $v_{n-2}v_{n-1}$ at some interior point x , where x is neither endpoint v_{n-2} nor v_{n-1} . Let $C_1 = (v_0, \dots, v_{n-2}, x)$ be the chain up to x , and $C_2 = (x, v_{n-1}, v_n)$ be the 2-link remainder of C . Then C_1 has $n - 1$ links (it is missing the last link $v_{n-1}v_n$) and is piercing because C is piercing and $s_1 = v_0x$ is a subsegment of $s = v_0v_n$. So we can apply the induction hypothesis to C_1 and conclude that it is in maxspan configuration. C_2 falls into the $n = 2$ base case,

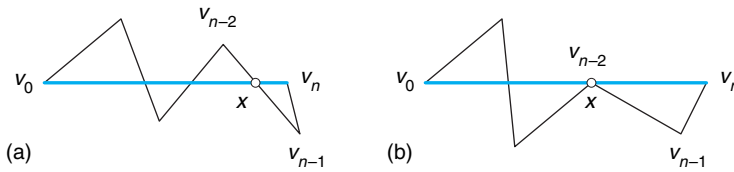


Figure 3.7. Piercing proof cases. (a) $s = v_0v_n$ pierces $v_{n-2}v_{n-1}$ at an interior point x ; (b) s pierces $v_{n-2}v_{n-1}$ at its v_{n-2} endpoint.

and so is in maxspan configuration. Finally, C_1 is aligned with C_2 , in that their individual spans line up along s . To apply the Alignment Lemma (Lemma 3.1), we only need to verify that the fixed-angle requirement at x is satisfied. One can view the angle at x fixed to be 180° , satisfied because x is an interior point of that $v_{n-2}v_{n-1}$ link. So indeed the Alignment Lemma applies and leads to the conclusion that the joined chain C is in maxspan configuration, and we have settled case (a).

Case (b), when s pierces $v_{n-2}v_{n-1}$ at the endpoint v_{n-2} , differs in that C_1 has only $n - 2$ links, but the induction hypothesis covers all n up to $n - 1$, so this is settled by a similar argument. In this case, the joint angle at x must satisfy the fixed-angle constraint because we assumed that C is a legal configuration of the chain in the first place, and again the Alignment Lemma applies.

Other cases occur when the last link, or the last two links, lie right along s , but they present no new issues. It is here that we are passing over details that would be needed in a formal proof. In all cases, induction applies, establishing Theorem 3.2. This ends the proof sketch.

Notice that the statement of this theorem makes no assumptions on link lengths or on joint angles or on planarity. So, like the Alignment Lemma, it is quite general. Now we apply the Piercing Theorem to our unit 90° -chains. Recall the claim of the Unit- 90° Theorem (Theorem 3.1) is that the staircase configuration achieves the maxspan.

If we have staircase configuration of a unit 90° -chain of n links, there are only two fundamentally different configurations, one for even n and one for odd n . Both configurations are obviously piercing; see Figure 3.8. When n is odd, the spanning segment pierces each but the first and last links at an interior point. When n is even, the spanning segment passes through every other vertex, and so every link is pierced at one of its endpoints. So the Piercing Theorem (Theorem 3.2) applies and shows those configurations must achieve the maxspan. This is precisely the claim of the Unit- 90° Theorem (Theorem 3.1).

Exercise 3.3 (Practice) Staircase Span. What is the maxspan of the two chains shown in Figure 3.8? Find general formulas for the maxspan of unit 90° -chains of n links, distinguishing between n even and n odd.

The generality of the Piercing Theorem (Theorem 3.2) permits a strengthening of Theorem 3.1 to any unit chain with the same fixed angle at every joint

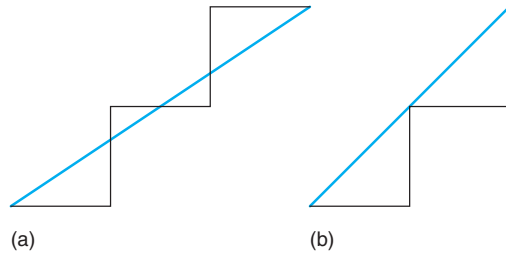


Figure 3.8. Unit 90° -chains in staircase configuration are piercing: (a) $n = 5$; (b) $n = 4$.

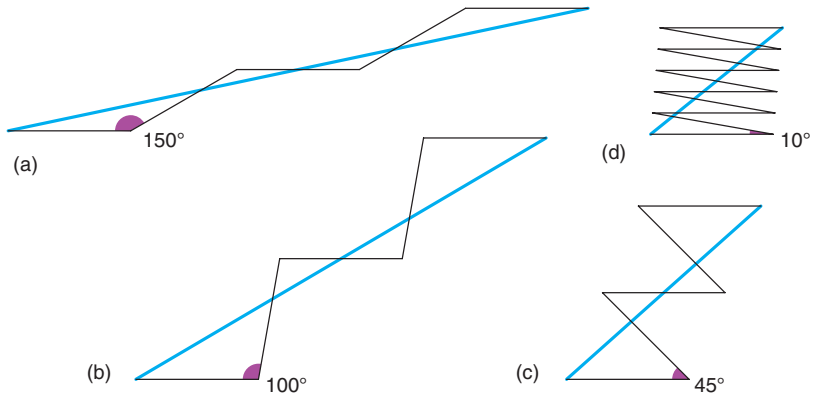


Figure 3.9. Unit α -chains are piercing, for any fixed angle α . Here all n are odd.

(see Figure 3.9), but we will not pursue this and other generalizations. Instead we turn now to a surprising application of the Unit- 90° Theorem (Theorem 3.1).

3.6 Pop-Up Spinner

Pop-up books and cards have been around since the 18th century and recently have seen a surge in popularity through the elaborate designs of pop-up masters like Robert Sabuda and Matthew Reinhart. But the most stunning and elegant pop-up effect I have encountered is the *pop-up spinner* card, invented in Japan by an unknown student at Musashino Art University in 1988. Its nested diamond-frames spin about a central axis as the card opens, with the inner frames spinning faster than the outer frames, creating a dazzling dynamic effect, as (tepidly) depicted in Figure 3.10. When the card is fully opened flat, one sees a simple pattern of nested diamonds (squares tilted at 45°) cut directly into the card – unlike other impressive pop-up effects, this one does not rely on complex attachments.

The engine behind all pop-up design is conversion of the card/page opening motion to pop up the parts of the construction and drive their motion. The

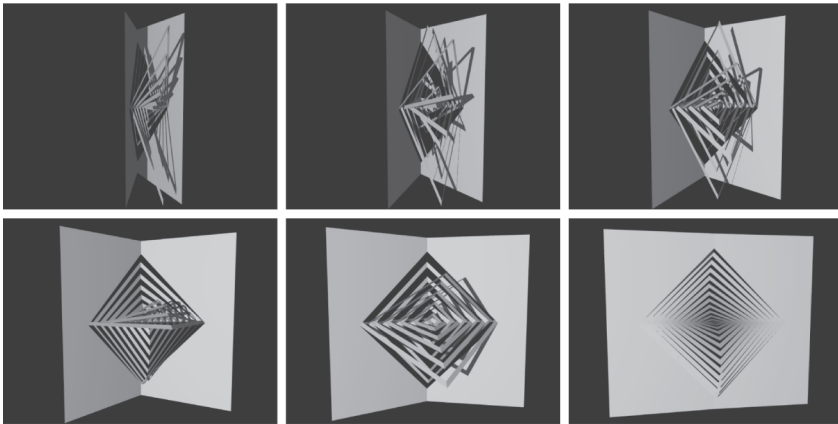


Figure 3.10. Pop-up spinner card opening animation, by Akira Nishihara.

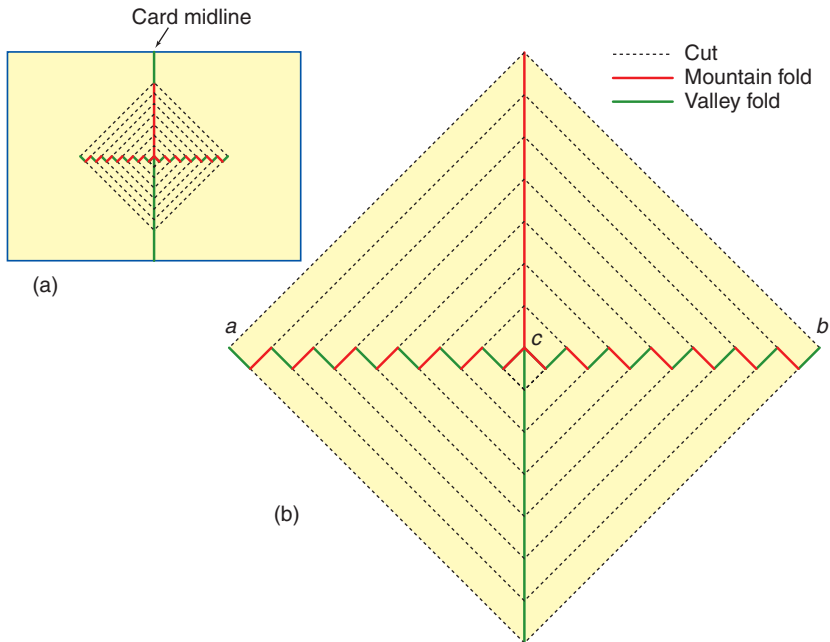


Figure 3.11. (a) Pop-up spinner card design. (b) Nested diamond details: cuts and folds.

heart of the pop-up spinner, and the key to understanding its operation, is – surprisingly! – the Unit-90° Theorem (Theorem 3.1). To see this, we need to study its construction in detail. Figure 3.11 shows a template for the construction, distinguishing cuts, mountain folds, and valley folds. As their names imply, a

mountain fold is a crease that bumps outward, while a *valley fold* is a crease that dents inward. Try cutting the pattern on card stock and crease as indicated – no description can fully convey the beauty of the physical action. See Box 3.1.

Box 3.1: Cut & Fold the Pop-up Spinner

Either draw the pattern with a ruler on card stock, or download the template from this book's web site <http://www.howtofoldit.org> and print on the heaviest weight paper your printer will accept. First crease the card down its midline, one long mountain fold. Don't worry that at this point some of the valley folds along that midline are folded backward. Now cut the diamond diagonal cut lines with scissors, stopping just at the central zig-zag, as in Figure 3.12. Fortunately, scissors suffice: No razor knife is needed as in many intricate constructions. Reverse the folds along the card midline that should be valley folds. Lay the design flat. Now, starting at point *a* in Figure 3.11 and proceeding across to *b*, crease the zig-zag segments with your fingers, alternating valley/mountain as indicated. Once every crease is folded the correct way, you can sharpen the creases with additional pressure. At this point you should be able to start twisting and compressing from the center point *c* outwards, reversing the animation snapshots in Figure 3.10. Then open: Voilá!

Figure 3.11 is too complicated for grasping the structure, so we will analyze portions of it separately. At a first glance, it is not evident how it all hangs together as one connected piece. Figure 3.13(a) shows the cuts for just two of the nested square diamonds, revealing clearly that a horizontal strip remains free of cuts. Figure 3.13(b) shows how a zig-zag of mountain/valley creases runs through the uncut strip, which in Figure 3.11 is seen to connect the left and right tips of the outermost diamond. Call these tips *a* and *b*.

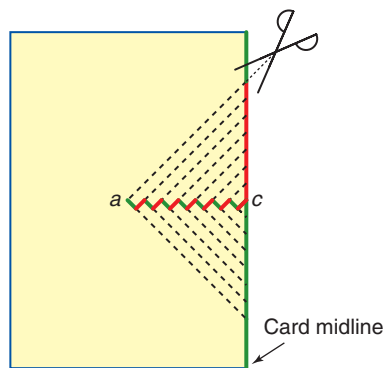


Figure 3.12. Cutting the pop-up spinner template, Figure 3.11(a).

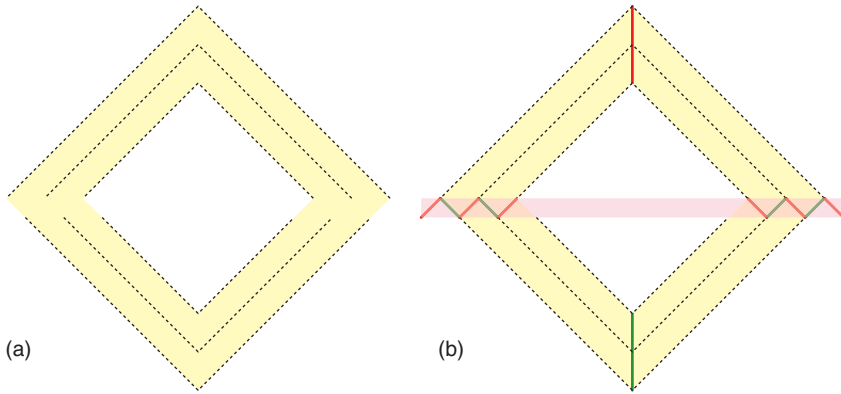


Figure 3.13. (a) Nested diamond-frames are connected. (b) Uncut strip containing zig-zag mountain/valley folds.

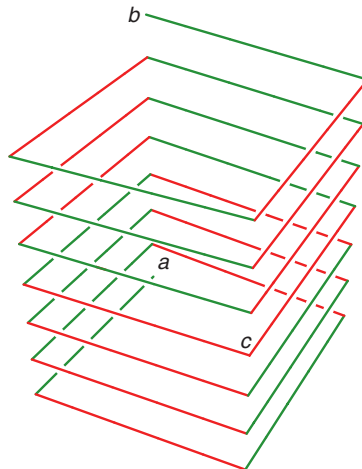


Figure 3.14. The zig-zag chain of mountain/valley creases, shown stretched out from a to b in Figure 3.11(b), curls up when the card is closed, forming a compressed spiral.

Now, here is the key observation: This mountain/valley zig-zag path is a unit 90° -chain! It is a chain of creases in the horizontal strip, with the angles fixed to 90° by the construction. When the card is closed, this chain (call it C) is curled up into a spiral like a coiled spring, as I have crudely depicted in Figure 3.14. The diamond-frames attached to C are twisted from their flat position as many times as the chain spirals around. When the card is opened, the two endpoints of C at the diamond tips a and b get stretched apart, forcing the chain to head toward its maxspan configuration. But we know from the Unit- 90° Theorem (Theorem 3.1) that the maxspan configuration is the *planar* staircase configuration. So C must

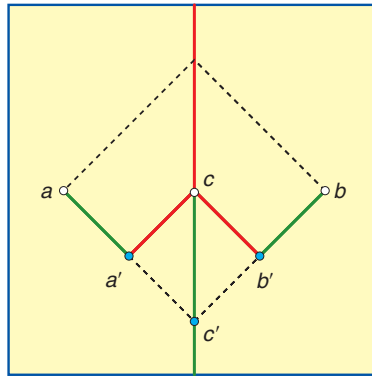


Figure 3.15. Minimal pop-up spinner, with innermost spinner tip c' identified. (Exercise 3.4)

uncurl from its initial, highly nonplanar state to the planar state, unwinding the initial spiral with all its attached diamond-frames. Those larger frames nearest a and b unwind the least; those more deeply nested, smaller frames nearest the centerpoint c have the most to unwind, and so spin faster. The inverse variation of frame size with spin speed adds to the elegance of the motion.

Exercise 3.4 (Understanding) Spinning Motion. The “minimal” functioning pop-up spinner follows the design in Figure 3.15. Point c' in this figure is the innermost tip of the spinner, where two cuts meet on the card midline. Describe the 3D motion of c' as the card opens, with respect to the points a' and b' , which remain fixed to the left and right card halves.

3.7 Above & Beyond

3.7.1 Folding@Home

Our narrow focus on fixed-angle chains has moved rather far from the protein-folding problem that inspired that focus. Although detailed geometric analysis is one line of attack on the protein-folding problem, there have been notable successes via atom-to-atom force simulations – something like a vast Sim City simulation for proteins. A typical simulation starts at an unfolded state of the protein, computes the forces between atoms from known physical principles, moves each atom slightly according to the total force vector applied over one small time step, and repeats. These simulations are computationally intensive for two reasons. First, experience has shown that the simulation time step – the time between simulation snapshots – must be on the order of 1 nanosecond (ns), one billionth (10^{-9}) of a second, whereas the real time it takes for a protein to fold is typically many thousandths (10^{-3}) of a second, that is, millions (10^6) of nanosecond time steps. So the simulation has to run for millions of time steps, often many millions. Second, what the simulation must do within each

time step is itself daunting, for accurate simulations must incorporate the forces between each of perhaps 10,000 atoms with one another – and that’s 50 million atom-to-atom interactions!

In the face of these computational challenges, researchers have resorted to massively parallel computation, either through development of special-purpose computers, such as IBM’s *Blue Gene*, or by harnessing many ordinary computers. The champion of the latter approach is the Folding@Home project, which developed software to run on idle computers. Whenever a machine with this software has been idle long enough for screen-saver to kick in, it starts working on simulating a “folding pathway” for a particular protein, sending its partial results to a server at Stanford University, which integrates all the data it receives. At this writing, 350,000 workstations around the world are participating in this project.

A notable early success of Folding@Home was an accurate and scientifically revealing simulation of the 36-amino acid protein known as the “villin headpiece,” which we used as illustration in Figure 3.2. This is a good target molecule not only because it is small (so there are fewer atom-to-atom forces to compute), but because it is one of the fastest folding proteins known, folding in as little as 700 ns (so it doesn’t need millions of simulation steps). Although the villin headpiece contains only 36 amino acids, the simulation includes 3,036 water molecules surrounding the protein; these represent the cellular fluid, which plays a crucial role in folding. Thus the simulation tracks approximately 10,000 atoms.

3.7.2 Locked Chains

Returning to the geometry of fixed-angle chains, we end this chapter with an unsolved problem suggested by protein folding but simple enough to be approachable in isolation, and perhaps solved by an enterprising reader. Let us say that a configuration C of a chain is *locked* if it is tangled in such a way that there is at least one other configuration C' of the same chain that is inaccessible from C by continuous motions without self-intersection. In other words, a configuration C is locked if it can only reach some other configuration C' by portions of the chain passing through itself. In technical mathematical language, a locked chain means that the *configuration space* of the chain is disconnected – some spots of the space are inaccessible from other spots. Because protein chain links have nearly the same length, it is especially interesting to know whether a “near-unit” 90° -chain can lock. To make this concept more precise, define the *length ratio* L of a chain to be the ratio of the length of its longest link to the length of its shortest link. So unit chains have length ratio precisely $L = 1$, and a near-unit chain has length ratio larger but close to 1. The unsolved question is: What is the smallest L that permits a 90° -chain to lock?

To date, the smallest length ratio found for a locked 90° -chain is $L > \sqrt{2} \approx 1.414$, achieved by the chain shown in Figure 3.16(a). In this 6-link example, 4 links have length 1, while the end links (v_0v_1 and v_5v_6) have length slightly more than $\sqrt{2}$. An attempt to untangle the chain as in Figure 3.17 just barely

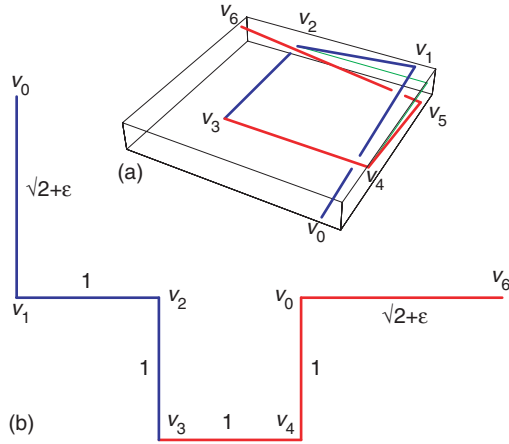


Figure 3.16. (a) Initial position of 6-link chain. (b) Unreachable planar configuration. ϵ is some small positive number.

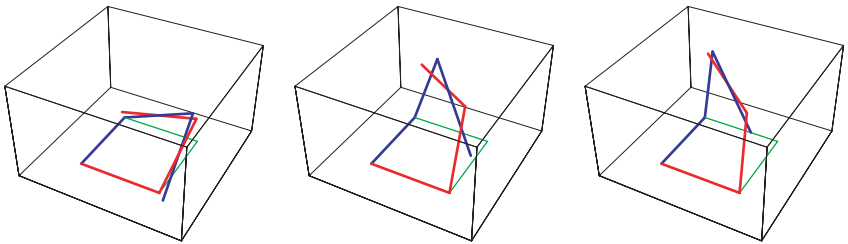


Figure 3.17. Three positions in an attempted opening motion. (See <http://howtofoldit.org> for an animation.)

fails with those lengths. Thus the configuration in Figure 3.16(a) cannot reach the planar configuration in (b). It is unknown if any 90° -chain with length ratio smaller than 1.414, even if composed of hundreds or thousands of links, can lock. In particular, this is an unsolved (“open”) problem:

OPEN PROBLEM: *Locked Unit 90° -Chain?*
Can a unit 90° -chain lock?

Exercise 3.5 (Challenge) Locked 4-Chain. Show that there is a locked 36° -chain of four links and nearly unit link lengths. By “nearly unit,” I mean lengths that could be chosen to be arbitrarily close to 1. Notice that 36° is the angle at the tips of a pentagonal star.

PART II

Origami

Origami is an art form with roots in Asia more than 1,000 years in the past, and likely coinciding with the invention of paper nearly 2,000 years ago. The Japanese word ‘origami’ literally means ‘fold, paper.’ Interest in the mathematics of origami arose only in the last century, and a focus on “computational origami” only since the 1990s.

In this part, I have selected three approachable topics. The first two concern “flat foldings,” a specialized form of origami. We first concentrate on single-vertex flat folds that, although not very exciting as origami, include some beautiful mathematical regularities. These regularities will help us in the next chapter explain the amazing “Fold and One-Cut” theorem, which is perhaps the prettiest result so far obtained in mathematical origami. And we close this Part of the book with another surprising but more specialized theorem, the “Shopping Bag” theorem.

4

Flat Vertex Folds

Although an origami folding generally produces a 3D object, such as the ubiquitous crane, intermediate stages of the folding are often *flat*, that is, parallel layers of paper squashed into a plane, as in Figure 4.1. In fact, flat origami as an end-product is its own well-developed art form.

In this chapter, we examine some of the surprising regularities present in flat origami, and then touch on the perhaps even more surprising technical unknowns lurking in a problem as commonplace as folding a map.

4.1 Mountain and Valley Creases

When you fold a sheet of paper in half, you create a straight-line *crease* that extends from one edge of the paper to an opposite edge. A crease snaps fibers in the paper, which is why the crease imprint remains after the creasing pressure is released, and why you cannot erase a crease completely by uncreasing – the fibers remain broken. Origami creases need not in general extend from edge to edge of the paper being folded. With some care, you can crease a line segment in the interior of the paper, with neither endpoint at the paper edge.

Creases come in two varieties: those created by a *mountain fold* and those by a *valley fold*, with natural meanings; see Figure 4.2. Traditionally, valley folds are indicated in origami diagrams as dashed lines $- - - - -$, and mountain folds by a dash-dot pattern, $- \cdot - \cdot - \cdot -$. Because these patterns are easily confused by the eye, we opt for the unconventional red for mountain and green for valley. (Memory aide: red sunset hitting peaks, lush green valleys.) Whether a crease represents a mountain or a valley fold depends on the point of view: From the underside, a mountain fold becomes a valley fold, and vice versa.

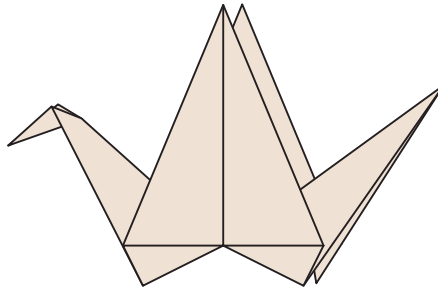


Figure 4.1. The standard origami crane, shown as a flat folding, before wings flap into 3D.

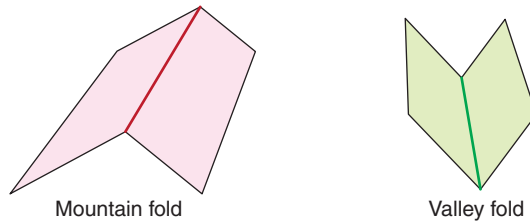


Figure 4.2. Mountain and valley folds.

4.2 Single-Vertex Flat Folds

There is already a rich mathematical structure in one of the simplest flat origami constructions: a flat folding containing a single vertex. A *vertex* in an origami construction is any point not on the boundary of the paper at which two or more creases meet. A simple example is the result of folding a sheet of paper in half twice: once top-to-bottom, and then left-to-right, which produces a vertex at which four creases meet; see Figure 4.3, in which the two sides of the paper are colored different shades.

Box 4.1: Folding Creases

Folding a crease that goes straight through a vertex is as easy as folding a sheet of paper in half. Folding a crease that stops at a vertex requires a somewhat different technique. One method is to fold the crease lightly right through the vertex, and then only firm up the crease (perhaps by pressing against a table) for the desired half. The method I use myself is to first draw the crease on the mountain side with a ruler. Then I hold the paper in the air and pinch at

several spots along the crease between my thumb and forefinger, up to but not through the vertex. Only once it has been precreased in this way do I set it on a table and sharpen the crease, either by sliding my thumbnail along it, or – better – pressing the edge of a ruler along the crease.

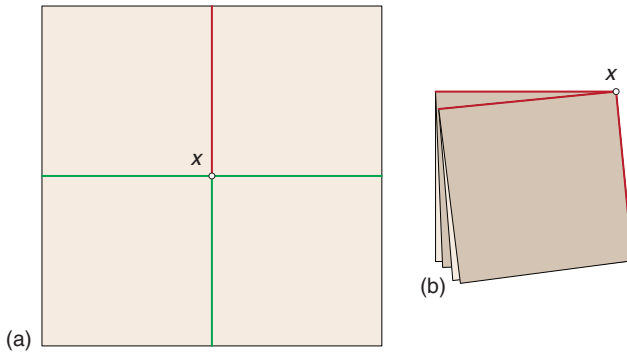


Figure 4.3. Degree-4 vertex: (a) Mountain/Valley creases on lighter side of paper; backside is darker. (b) Flat folding. The three valley creases become mountain creases on the darker side.

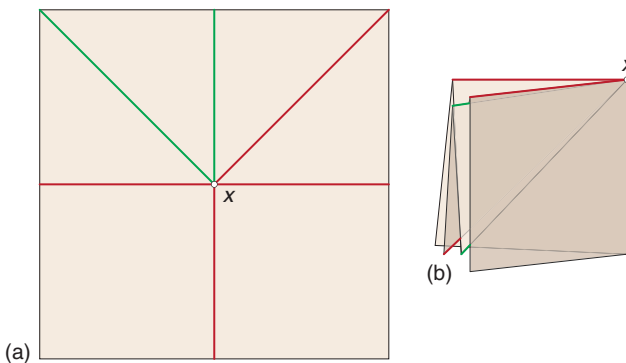


Figure 4.4. Degree-6 vertex: (a) crease pattern; (b) folding. Some sheets are shown partially transparent.

More complicated examples can be made by terminating a crease at the vertex, for example, as in Figures 4.4. and 4.5.

Exercise 4.1 (Practice) Four Mountain Creases. Create four mountain creases meeting at a central vertex, as shown in Figure 4.6, as follows. Fold a piece of paper in half, top to bottom. Now unfold completely, and fold it in half, left to right so that the two perpendicular creases are both mountain creases (or

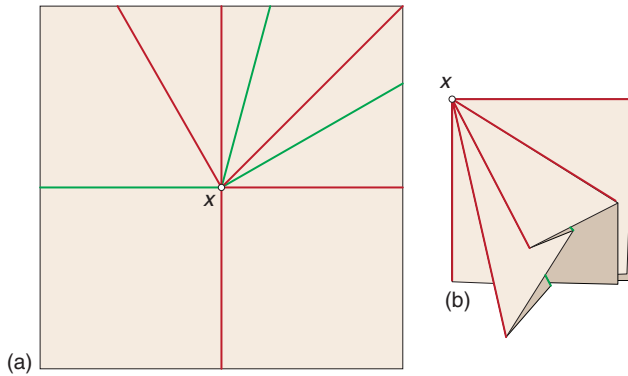


Figure 4.5. Degree-8 vertex: (a) crease pattern; (b) folding.

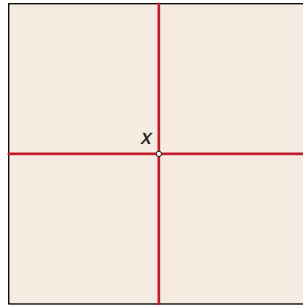


Figure 4.6. Four mountain creases meeting at vertex x . (Exercise 4.1)

valley creases from the opposite side). Open the paper again. Convince yourself by manipulation that the paper cannot fold flat with just those four creases mountain-folded and meeting at the central vertex x (as they do in Figure 4.3(b)).

Is there any pattern to the single-vertex flat foldings we've examined so far? I encourage the reader to experiment with sheets of paper and formulate conjectures. Perhaps the first regularity to become apparent is that the number of creases meeting at the vertex must be even in order for the pattern to fold flat: 4 in Figure 4.3, 6 in Figure 4.4, 8 in Figure 4.5. And if we consider the midpoint of the single crease formed by folding the sheet in half, a special type of vertex where two mountain folds meet along the same line (collinearly), then again there must be an even number: 2.

Indeed this regularity holds universally:

Theorem 4.1 (Even Degree)

A vertex in a flat folding has even degree.

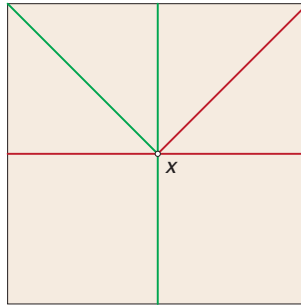


Figure 4.7. The crease pattern of Figure 4.4(a) with different mountain/valley folding labels.

In this case, *degree* has nothing to do with angular measure, but rather is the technical term for the number of creases coming into (*incident to*) the vertex. This theorem will turn out to be a consequence of a deeper regularity that we will see in the next section. Exercise 4.1 shows there must be more here, because there four creases would not fold flat. Another clue is the six-crease example in Figure 4.7, where the creases follow the same lines as in Figure 4.4 but with a different mountain/valley folding. Try as you might, you cannot fold this diagram flat. There must be both some imbalance between mountain and valley folds, and some near-balance. The regularity here is captured in a beautiful theorem named after the two people who first discovered it (independently of one another) in the 1980s, Jun Maekawa and Jacques Justin.

4.3 The Maekawa-Justin Theorem

Theorem 4.2 (Maekawa-Justin)

If M mountain creases and V valley creases meet at a vertex of a flat folding, then M and V differ by 2: either $M = V + 2$ or $V = M + 2$.

We check our examples so far (Table 4.1.), verifying that they do indeed satisfy Theorem 4.2. We now prove Theorem 4.2.

Let's start with a circular piece of paper (Figure 4.8(a)) so we are not distracted by the corners, which are irrelevant to what happens in the *neighborhood* of the single central vertex. Now we consider an arbitrary single-vertex flat folding of the paper; our goal is to prove that M and V differ by 2. Lay the folding flat, as in Figure 4.8(b). Now look at the side of the folded paper toward the vertex inside to see a closed zig-zag path of circular arcs, as depicted in (c) of the figure. Each

Table 4.1. The number of mountain and valley creases (M and V respectively) in our examples, checking the Maekawa-Justin theorem (Theorem 4.2).

Figure	M	V	$M - V$	Theorem 4.2 satisfied?
Figure 4.3	3	1	2	X
Figure 4.4	4	2	2	X
Figure 4.5	5	3	2	X
Figure 4.6	4	0	4	×
Figure 4.7	3	3	0	×

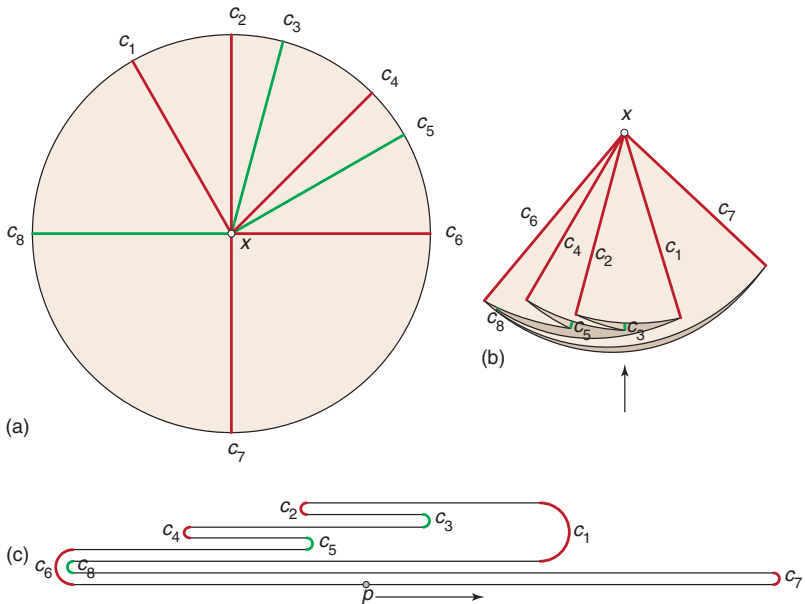


Figure 4.8. The example of Figure 4.5 revisited: (a) Crease pattern on circular paper. The eight creases are labeled c_1, \dots, c_8 . (b) Flat folding. (c) Expanded view looking from folded boundary toward vertex. Sharp turns at creases are shown as circular arcs to illustrate the nesting. Starting direction vector from p toward right.

arc is a piece of the circular boundary flattened between two creases, which, viewed edge-on, appears as a straight segment. Select any point of the path not directly at a crease, for example, point p in (c), and imagine walking toward the right. Let's view your direction of travel as a vector (see Chapter 1, Box 1.3, on vectors). Then the start direction vector points at angle 0° in the standard

coordinate system, in which angles are measured counterclockwise from the positive x -axis, which points toward the right.

Each mountain fold you encounter in your walk rotates your direction vector through $+180^\circ$ (+ meaning counterclockwise), and each valley fold rotates your vector through -180° (- for clockwise). Although it is true that rotation by $+180^\circ$ and by -180° bring the vector to the same final heading – exactly opposite to the heading before rotation – the intermediate headings are different. For mountain turns, the headings point to the exterior of the folding; for valley turns, the headings point to the inside of the construction.

Now, we know that by the time we return to the starting point p after traversing the entire diagram in (c), we approach p from the left heading right, so again the vector has direction 0° , which is the same as 360° . In other words, we must twist a total of a full 360° by the time we return to start.

So we must have:

$$M \cdot 180^\circ + V \cdot (-180^\circ) = 360^\circ$$

Dividing through by 180° leads to $M - V = 2$. Remembering that M and V are two sides of the same coin, we know that flipping the paper over in Figure 4.8(a) would interchange the roles of M and V , and we'd reach the conclusion that $V - M = 2$. Combining both possibilities into one phrase: M and V differ by 2. That is the exactly the claim of the theorem; so we have proved Theorem 4.2.

The Maekawa-Justin theorem easily implies the Even-Degree Theorem (Theorem 4.1). Suppose $M = V + 2$. Then:

$$M + V = (V + 2) + V = 2V + 2 = 2(V + 1)$$

and so the total number of creases $M + V$ into a vertex x (as in Figure 4.8a,b) is even. The same logic applies when starting with $V = M + 2$ and reaches the same conclusion: $M + V$ is even.

Most theorems have many proofs, often starting from different background assumptions. An alternate proof of Theorem 4.2 using polygons is presented in Box 4.2.

Box 4.2: Proof of Maekawa-Justin Theorem via Polygons

The following proof was found by Jan Siwanowicz when he was a high-school student. The starting point of his proof is another theorem: The sum of the internal angles at the n vertices of a polygon is $(n - 2)180^\circ$. (This in turn follows from the theorem that every polygon can be partitioned by diagonals between

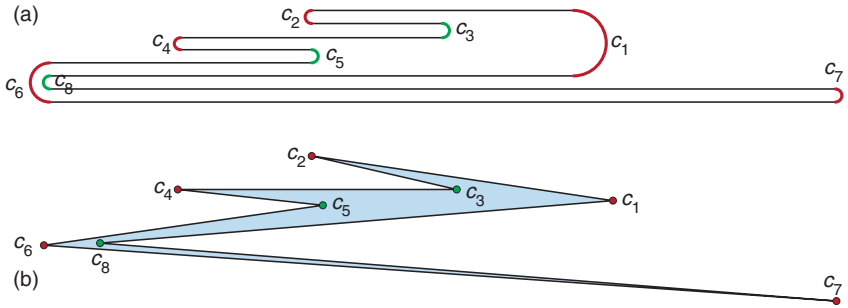


Figure 4.9. An 8-vertex polygon (b) corresponding to Figure 4.8(c), repeated as (a) here, with 5 mountain vertices $\{c_1, c_2, c_4, c_6, c_7\}$ and 3 valley vertices $\{c_3, c_5, c_8\}$.

its vertices into $n - 2$ triangles, so that the total internal angle is that of $n - 2$ triangles, each of which has angle sum 180° .) The idea is to view the zig-zag path in Figure 4.8(c) as a squashed polygon, as in Figure 4.9, which is closer to how it looks with sharp creases. If we imagine compressing this polygon completely flat, all the mountain vertices have an internal angle near 0° , and all the valley vertices have an internal angle near 360° . So the total internal angle sum after complete flattening is:

$$M \cdot 0^\circ + V \cdot 360^\circ$$

and this must equal $(n - 2)180^\circ$, where n is the total number of vertices of the polygon. In this construction, each vertex derives from a crease, so $n = M + V$. Therefore:

$$V \cdot 360^\circ = (M + V)180^\circ$$

and dividing by 180° yields $2V = M + V$ or $M - V = 0$.

Exercise 4.2 (Practice) Maekawa-Justin Theorem. Exercise 4.1 argued that four mountain creases lead to an unflattenable vertex. Add additional creases to Figure 4.6 so that it can flatten, and verify the Maekawa-Justin Theorem (Theorem 4.2) for your construction.

4.4

The Local Min Theorem

The pattern in Figure 4.10(a) shows that we still haven't plumbed the depths of single-vertex flat foldings fully. It satisfies the Maekawa-Justin Theorem (Theorem 4.2) with $M = 4$ and $V = 2$, and therefore satisfies the Even-Degree Theorem (Theorem 4.1) with degree 6. And yet, if you try to fold it flat, you will see it is impossible. Why? The essence of the impediment is that a 40° pie-slice wedge

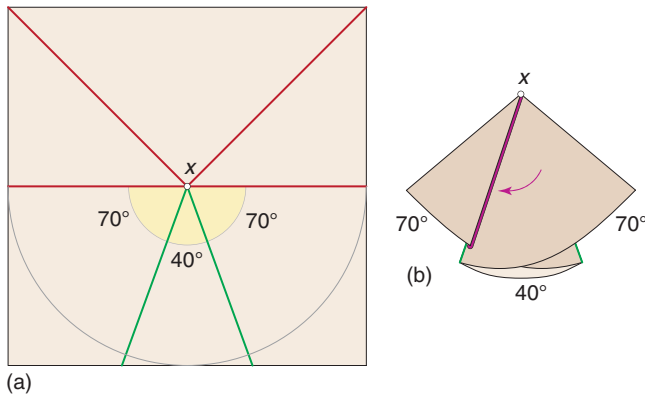


Figure 4.10. (a) A crease pattern that cannot fold flat. (b) Attempting to fold the 40° wedge results in paper penetrating itself. Here we've restricted the folding to the semicircle (a) to make the angular relationships clear.

delimited by two valley folds is surrounded by larger angles on either side – 70° . This forces paper to pass through itself, as depicted in (b) of the figure. Whenever we have such a pattern of consecutive wedge angles: {large, small, large}, the folds delimiting the central wedge cannot both be valley, nor can both be mountain: One must be mountain and the other valley. The central angle is called a *local min*, because locally – that is, in its immediate neighborhood – it is a minimum angle, smaller than its neighbors to either side. We can phrase this condition in a theorem as follows:

Theorem 4.3 (Local Min)

In any flat folding, any wedge whose angle is a local min must be delimited by one mountain and one valley fold.

Exercise 4.3 (Practice) *Three Theorems Check.* Check which of Theorems 4.1 (Even Degree), 4.2 (Maekawa-Justin), and 4.3 (Local Min) are satisfied by the crease pattern in Figure 4.11.

The three regularities we've uncovered so far are what mathematicians call *necessary conditions*: Every single-vertex flat folding necessarily satisfies them. But they may or may not be *sufficient conditions*: conditions on the crease pattern which, if satisfied, imply the diagram can be folded flat. Indeed our three conditions, in pairs or even all three together, are not sufficient conditions. The holy grail in mathematics is a set of necessary and sufficient conditions, which then completely characterize the situation. For single-vertex flat folds, these are embodied in the Kawasaki-Justin Theorem.

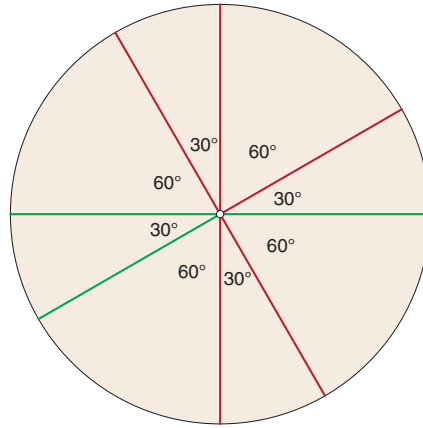


Figure 4.11. A single-vertex crease pattern for checking. (Exercise 4.3)

4.5 The Kawasaki-Justin Theorem

The Local-Min Theorem (Theorem 4.3) indicates that the measures of the wedge angles defined by the crease pattern are important. Let us call the wedge angles around the vertex in sequential order, $\theta_1, \theta_2, \dots, \theta_n$. We know from the Even-Degree Theorem (Theorem 4.1) that n is even, because an even number of creases determine an even number of wedges. We also know that:

$$\theta_1 + \theta_2 + \dots + \theta_n = 360^\circ$$

because the angles completely surround the vertex. The Kawasaki-Justin Theorem claims that a simple condition on the angles, completely ignoring the mountain-valley pattern, provides necessary and sufficient conditions for flat foldability:

Theorem 4.4 (Kawasaki-Justin)

A set of an even number of creases meeting at a vertex folds flat if, and only if, the alternating sum of the determined wedge angles is zero:

$$\theta_1 - \theta_2 + \theta_3 - \theta_4 + \dots + \theta_{n-1} - \theta_n = 0^\circ$$

The term *alternating sum* means that every other term has opposite sign: The odd terms $\theta_1, \theta_3, \theta_5, \dots$ are added and the even terms $\theta_2, \theta_4, \theta_6, \dots$ are subtracted.

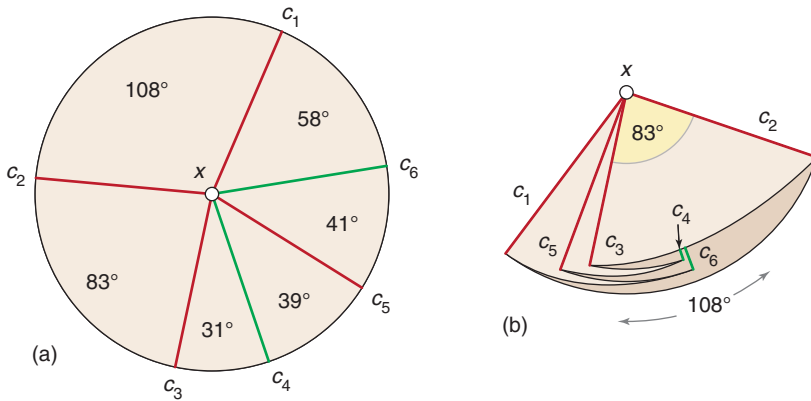


Figure 4.12. Illustration of Kawasaki-Justin Theorem 4.4: $31^\circ + 41^\circ + 108^\circ = 39^\circ + 58^\circ + 83^\circ$.

So the alternating-sum equation is equivalent to:

$$\theta_1 + \theta_3 + \theta_5 + \dots = \theta_2 + \theta_4 + \theta_6 + \dots$$

the sum of the odd-indexed angles equals the sum of the even-indexed angles. The phrase “if, and only if,” is mathematician’s shorthand for claiming necessary (“only if”) and sufficient (“if”) conditions.

Figure 4.12(a) shows a 6-crease example with six wedge angles:

$$31^\circ + 39^\circ + 41^\circ + 58^\circ + 108^\circ + 83^\circ = 360^\circ$$

Their alternating sum is indeed zero:

$$31^\circ + 41^\circ + 108^\circ = 180^\circ = 39^\circ + 58^\circ + 83^\circ$$

so

$$31^\circ - 39^\circ + 41^\circ - 58^\circ + 108^\circ - 83^\circ = 0^\circ$$

The flat folding guaranteed to exist by the theorem is shown in (b) of the figure.

Exercise 4.4 (Practice) Kawasaki Theorem Check. Check if Theorem 4.4 is satisfied by the example used in Exercise 4.3, Figure 4.11.

The claim that Theorem 4.4 provides a complete characterization of flat foldability is rather remarkable, because it says nothing explicitly about the pattern of mountain and valley folds on which we’ve been concentrating! But because its conditions are sufficient, the alternating angle sum must somehow imply both the Maekawa-Justin Theorem (Theorem 4.2) and the Local-Min Theorem (Theorem 4.3). Kawasaki’s theorem implies that there must exist a way to select creases for mountain folds and other creases for valley folds to make those theorems work out.

The proof of necessity proceeds just as with the Maekawa-Justin argument, analyzing the zig-zag circular paper boundary path, as in Figure 4.8(c) (p. 62). Again we imagine walking around this path. But now rather than concern ourselves with the gyrations of the direction vector of travel, we concentrate on how far we travel, measuring “how far” not in terms of linear distance, but in terms of angular travel as seen from the central vertex. Let’s use Figure 4.12(b) as an example. Starting at the leftmost edge of the folding and traveling rightward on the bottommost flap, we travel an arc of 108° with respect to the apex x . At the mountain fold we reverse direction and travel an arc of 83° leftward, then reverse again and travel 31° rightward, and so on. Whether we encounter a mountain or a valley fold is irrelevant if we are just concerned with total angular travel. By the time we return to the start point, the total travel must be 0° . And so the alternating sum must be zero, which means it is necessarily zero.

That the alternating sum condition is also sufficient for the pattern to be flat-foldable is not as easy to see, and we will have to leave it as a claim that the Kawasaki-Justin Theorem 4.4 completely characterizes single-vertex flat foldability. Given any crease pattern incident to a single vertex, and a protractor, you can tell in advance whether or not it may be folded flat. Moreover, in an even less obvious manner, the Local-Min Theorem (Theorem 4.3) can be used to determine a mountain/valley assignment for the creases that will fold it flat. Indeed, there are in general many such assignments – eight for the pattern in Figure 4.12(a). Thus, in some sense, single-vertex flat foldings are completely understood.

Exercise 4.5 (Understanding) *Kawasaki Revisited.* Exercise 4.1 concluded that the four creases of Figure 4.6 cannot fold flat. But Theorem 4.4 is satisfied: $90^\circ - 90^\circ + 90^\circ - 90^\circ = 0$. So it should fold flat. Where is the contradiction?

4.6

Above & Beyond

4.6.1 Flat Foldability is Hard

Flat origami of any artistic interest includes more than one vertex. For example, the elegant “oval tessellation” (Figure 4.13) designed by Robert Lang has 136 vertices. Each must, individually, satisfy all the theorems of this chapter, but it is known that, in general, this does not suffice: There are diagrams with every single-vertex crease pattern locally “legal,” but the whole pattern cannot be folded flat. A complete characterization of which patterns of creases are flat foldable has remained out of reach. Perhaps the first result of what has come to be known as *computational origami* implies that it might remain forever out of reach. Marshall Bern and Barry Hayes proved in the 1990s that deciding whether a crease pattern (even with mountain/valley labels explicitly provided) is flat foldable is *NP-hard*, a computational complexity classification that means: at least as hard as the NP-complete problems, which we saw, in Chapter 1 (p. 21), are “intractable.” Many mathematicians believe that any NP-complete problem

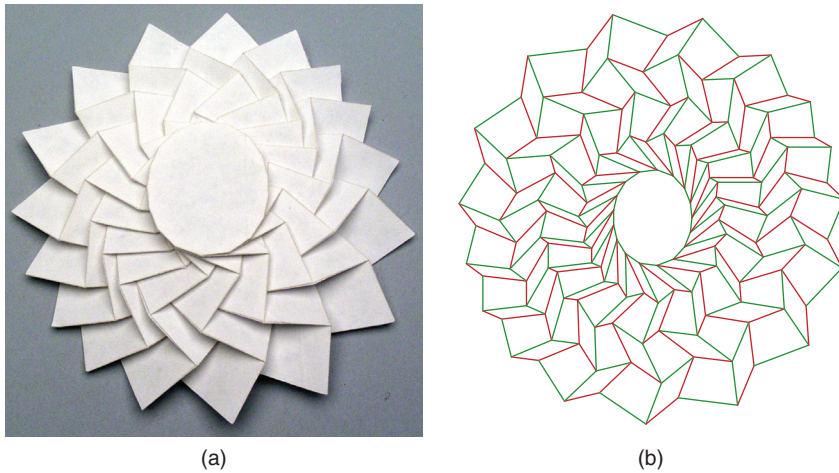


Figure 4.13. Robert Lang's *Oval Tessellation*, 1999.

is not only impractically difficult to solve computationally, but also that it will forever resist being captured in a concise set of necessary and sufficient conditions (because these would likely lead to tractable computations). Without the possibility of a complete mathematical classification, the artistic core of flat origami is not at risk of being overrun by mechanization.

4.6.2 Map Folding Complexity

I close this chapter with an unsolved problem: deciding whether or not a map crease pattern can be folded flat. Anyone who has struggled with correctly refolding a map in a car will appreciate the practical difficulty of the task, but the unsolved problem concerns its computational complexity: Essentially, is it “tractable” (technically, achievable in *polynomial time*) or is it intractable (NP-complete or worse)?

You might wonder why this question is not already settled by the Bern-Hayes result I mentioned in the previous section. The answer is that map folding is a very special case. The *map* is assumed to be rectangular, with the creases forming a regular grid of squares, with each crease segment labeled mountain or valley. The Bern-Hayes proof fails on this special case, leaving hope that this specific problem is tractable.

By this point you are likely wondering: What could be so hard about folding a map? I encourage you to try to fold the example in Figure 4.14, even with the help of the illustrated solution. The freedom to tuck layers over/under/inside of one another gives the problem a rich combinatorial structure that has resisted understanding. Even to answer the map-folding question for $2 \times n$ maps – two grid squares high by n grid squares wide – remains unsolved:

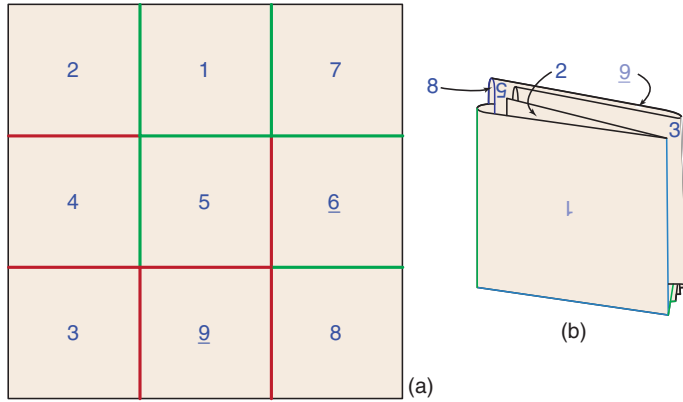


Figure 4.14. (a) A map-folding puzzle. (b) Solution, with several squares labeled (lightly shaded labels are facing away from viewer).

OPEN PROBLEM: Map Folding

Is there an efficient method (*algorithm*) for deciding whether or not a given rectangular map can fold flat, with each grid crease segment pre-marked as either a mountain or a valley fold?

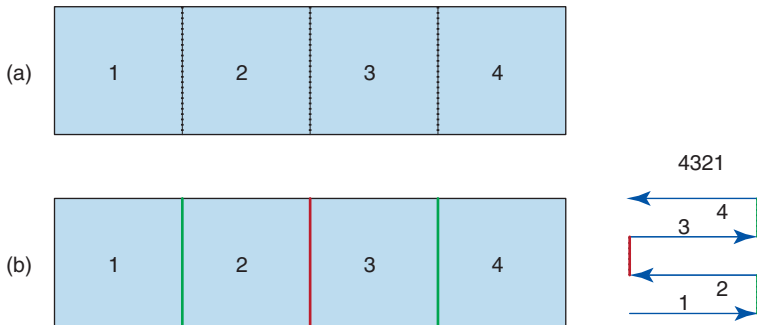


Figure 4.15. Folding four stamps. (Exercise 4.6)

Exercise 4.6 (Challenge) *Stamp Folding.* If you have a strip of four stamps, labeled on their tops with the numbers 1, 2, 3, 4 as shown in Figure 4.15(a), how many different permutations of 1234 can you achieve by folding the stamps along their perforated connections into a stack? There are $4! = 24$ different permutations of 1234. Can all of them be achieved? The convention for counting is that, after folding, orient the stack so that the 1-stamp (wherever it is) is facing upward, and then read off the stamp numbers from the stack top to bottom. For example, (b) in the figure shows a folding that achieves the permutation 4321.

5

Fold and One-Cut

The most impressive result of mathematical origami discovered to date is so surprising that it has been used as a magic trick by none other than Harry Houdini, the great magician. The result is this. Make any straight-line drawing on a sheet of paper. For example, you might draw a number of shapes: a rectangle, a triangle, a star, some block letters, and so on (but no circles or other curves). See Figure 5.1. Then it is possible to (cleverly!) fold the paper flat so that, with one straight scissors cut completely through the paper, all the shapes you drew are simultaneously cut out, leaving the original paper missing exactly those shapes.

5.1

Examples

Square. This is so hard to believe that we'll start with a simple example so that you see that it at least might be possible: a square centered on a square piece of paper; see Figure 5.2. To cut out this shape, we must fold the paper to place the four edges of the square aligned on top of one another. Call the four corners of the square A, B, C, D as in (b) of the figure. Crease the paper with a mountain fold along diagonal AC . This aligns the bottom edge AB with the left edge AD underneath, and similarly aligns the right edge BC with the top edge CD . Now the square boundary has become an L-shape. One more symmetric mountain fold through B as shown in (c) results in all four edges of the square lying directly on top of one another in the four layers of the folding. Now cutting AB and the edges underneath cuts out exactly the square and nothing else.

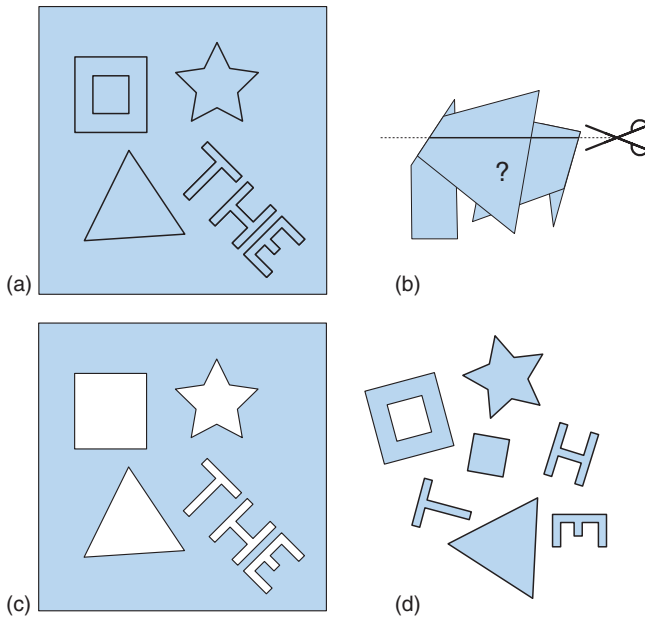


Figure 5.1. (a) A straight-line drawing. (b) Flat folding of the paper is cut straight through by scissors. (c) The shapes are cleanly cut out from the paper and fall out separately (d).

This is the main idea: fold the paper so that all the edges of the drawing align one on top of another, and so that nothing else is along this line. Then cutting through that line cuts all the edges, and nothing else. The surprise is that this can always be done for any straight-line drawing.

Regular 5-pointed star. Let's take another example, the regular 5-pointed star described by Houdini in his 1922 book *Paper Magic*. He first folds it down the central line of symmetry to align the left and right halves, and then methodically aligns more and more of the edges via folds through the star center; see Figure 5.3. A total of four folds align all ten edges on top of one another (in eight layers of paper), and one straight cut excises the star. Voilà! This trick was known to Betsy Ross, for she convinced George Washington to use the regular 5-pointed star on the American flag because it was easy to produce by fold and one-cut.

The emphasis here is on the *regularity* of the star: All its edge lengths are the same, all the spike angles are equal, and all the dent angles are equal. For it is this regularity that ensures that a fold that bisects an angle (that is, splits into two equal angles) places two equal-length edges on top of one another.

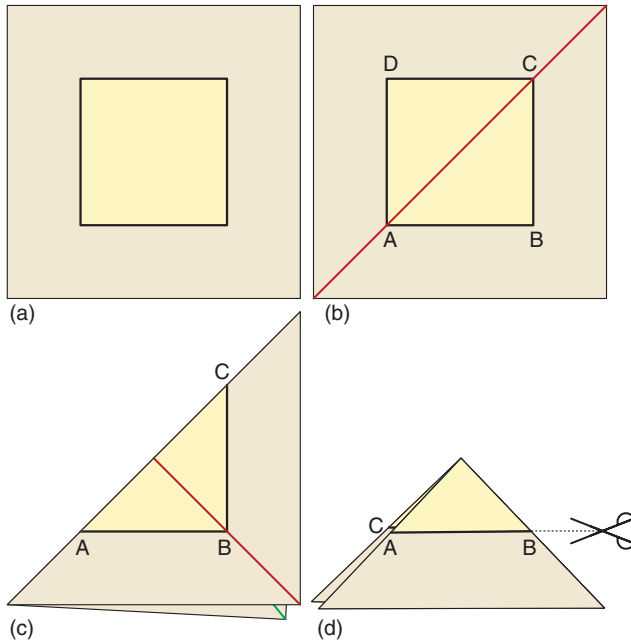


Figure 5.2. (a) A square to cut out. (b) Mountain fold along diagonal AC, and then along opposite diagonal (c) aligns all four edges on top of one another (d).

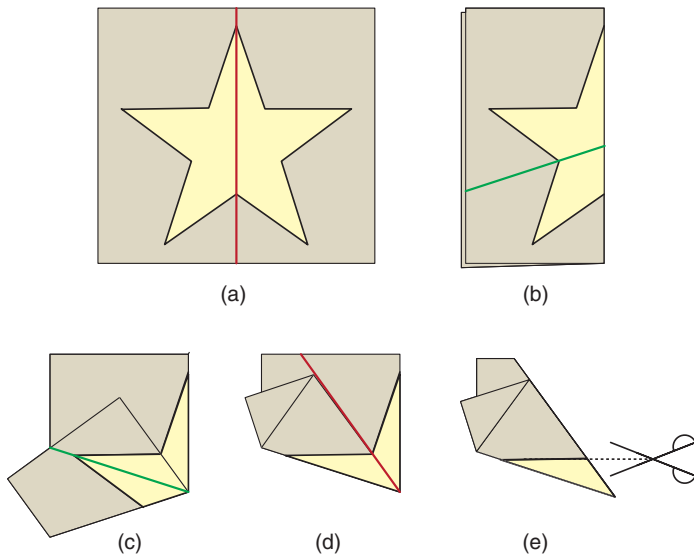


Figure 5.3. Houdini's method of folding a regular 5-pointed star for one-cut.

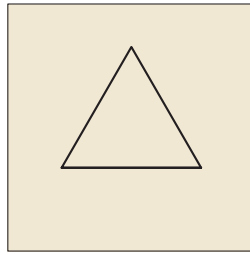


Figure 5.4. An equilateral triangle for fold and one-cut. (Exercise 5.1)

Exercise 5.1 (Understanding) Equilateral Triangle. Fold flat a piece of paper with an equilateral triangle drawn on it (Figure 5.4) so that the triangle can be cut out with one straight cut.

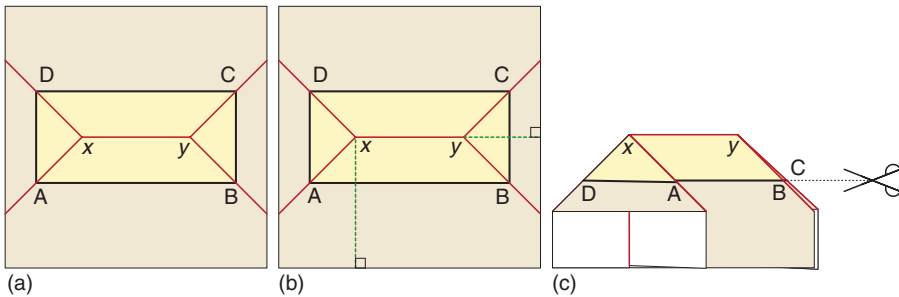


Figure 5.5. (a) Mountain-fold bisectors. (b) Valley-fold perpendiculars. (c) Flat folding ready for one-cut.

Let us now explore examples that take a few steps away from regularity: a rectangle, and then an irregular triangle.

Rectangle. A rectangle is not as “regular” as a square, because the side lengths differ. And indeed, bisecting the four corner angles gets us started in the right direction, by laying the short sides on top of the long sides, but does not complete the job, because the central portions of the long sides (AB and CD in Figure 5.5(a)) are not yet accounted for. A little experimentation should convince you that a crease down the center of the rectangle, as illustrated, suffices to draw the two long sides together on top of one another. But then still we are not finished, because the points at which the bisectors converge fail to satisfy the Even-Degree Theorem (Theorem 4.1) from Chapter 4: That theorem says there must be an even number of creases incident to a vertex of a flat folding, but there are three incident to x and y in the figure. More precisely, flattening demands that we

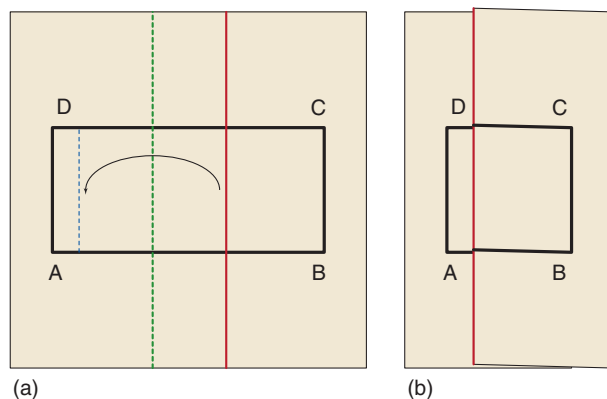


Figure 5.6. Converting a rectangle to a square via a crimp. (Exercise 5.3)

satisfy the Maekawa-Justin Theorem (Theorem 4.2) and have the number of mountain and valley folds incident to these vertices differ by two. In order to satisfy this demand and at the same time not destroy the alignment already achieved of the rectangle edges, we introduce creases known as *perpendiculars*, one valley crease each from x and from y , which cross the rectangle sides at right angles. Now we have $M = 3$ and $V = 1$ at x and y , satisfying Maekawa's Theorem. There are several choices here for the direction of the perpendicular crease at each vertex. One is shown in Figure 5.5(b) (we draw perpendiculars dashed), which leads to the flat-folding in (c) that achieves the needed fold and one-cut overlapping of all the rectangle's edges.

Exercise 5.2 (Practice) Perpendiculars. How many distinct ways are there to draw perpendiculars from x and y in Figure 5.5(b)? (Don't count symmetric options as different.) Convince yourself that all lead to a one-cut fold for the rectangle

Exercise 5.3 (Understanding) Rectangle Crimp. Suppose instead of following the plan in Figure 5.5, you attempt to fold and one-cut a rectangle as illustrated in Figure 5.6. First a "crimp fold" consisting of parallel mountain and valley folds are used to shorten the rectangle to a square, as shown in (b) of the figure. Then the plan for fold and one-cut of a square illustrated previously in Figure 5.2 is followed. Will this strategy work?

Triangle. The same strategy employed for a rectangle works for an irregular triangle. First, mountain-fold crease along angle bisectors of the three triangle corners. These three bisectors meet at a single point, a (non-obvious!) theorem known to Euclid. See Box 5.1. Call that single point x . Choose a valley-fold perpendicular from x to any one of the three sides; see Figure 5.7(a) Now fold and one-cut as in (b).

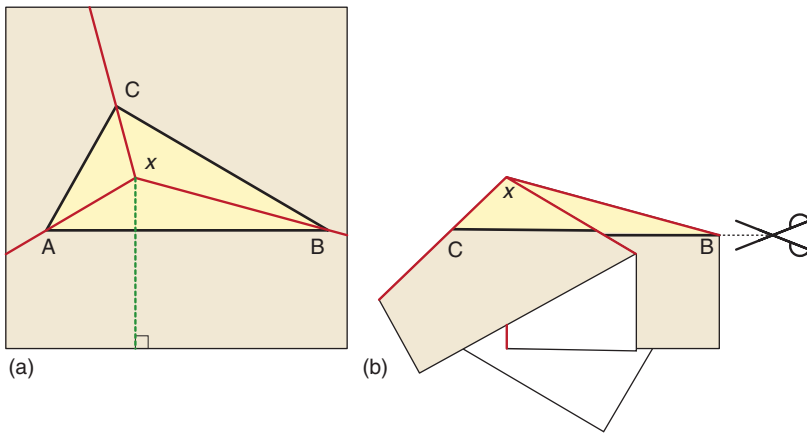


Figure 5.7. (a) Creases for triangle. Perpendicular shown dashed. (b) Flat folding ready for one-cut.

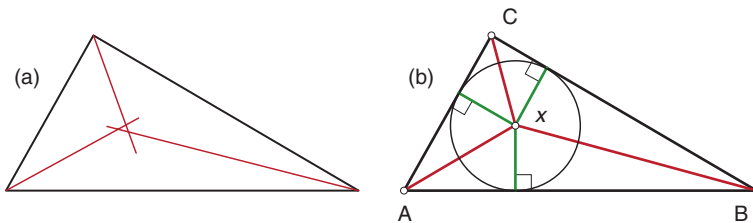


Figure 5.8. (a) Bisectors could conceivably miss converging on a single point. (b) Bisectors in actuality meet at the center of the inscribed circle.

Box 5.1: Angle Bisector Theorem

The three angle bisectors of a triangle meet in a point. This claim is by no means obvious; the bisectors could fail to meet at a single point, as in Figure 5.8(a). However, when you fold the bisectors, it is immediately convincing that they do in fact meet at a single point; see (b). It amounts to almost a physical “proof.” This bisector theorem follows from Proposition 4 in Book IV of Euclid. This proposition concerns inscribing a circle in a triangle, and in the course of the proof that there is indeed a unique circle that touches all three sides, he proves that connecting the circle center to any corner bisects the angle at that corner. Consequently, the three angle bisectors meet at a single point, the center of the inscribed circle.

So far we have concentrated on single shapes, but the claim of the theorem is that any collection of shapes can be cut out simultaneously. This is one of

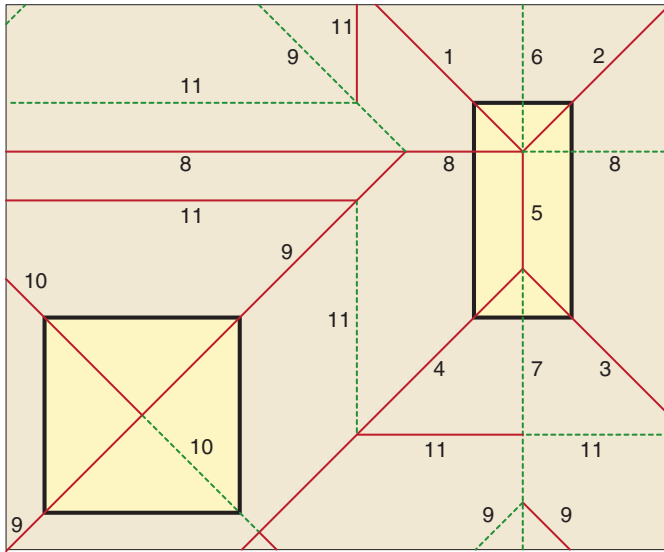


Figure 5.9. Fold and one-cut of a square and rectangle simultaneously. Creasing in the order indicated by the labels aligns exactly the edges to be cut.

the many aspects of this complex result that we have to skim over. I invite the reader to fold the example shown in Figure 5.9 to see that indeed a square and a rectangle can be both cut out at the same time by one slice of the scissors.

5.2 Fold and One-Cut Theorem

I hope you are now prepared to at least believe that the fold and one-cut theorem might be true. Here it is:

Theorem 5.1 (Fold and One-Cut)

Any straight-line drawing (one composed of straight segments) on a sheet of paper may be folded flat so that one straight scissors cut completely through the folding cuts all the segments of the drawing and nothing else.

Unfortunately, the only known proof of this theorem is quite complicated and beyond the scope of what we can present here. Its difficulty is indicated by the tortured history of the proof. The theorem was discovered by three researchers, whose junior—and driving—member Erik Demaine was 17 years old at the time. But the proof (they subsequently realized) did not cover all possible drawings, just “nearly all.” Later, Erik joined three other researchers to create a completely

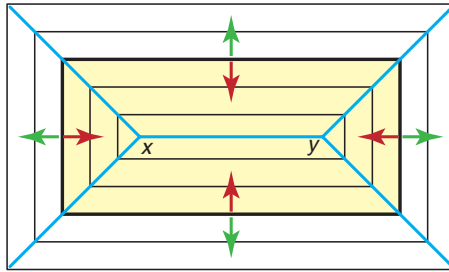


Figure 5.10. Growing and shrinking to trace out the straight skeleton.

different proof that did cover all drawings, but subsequently they uncovered a flaw in that proof. Now the proof has been repaired twice independently, and the theorem regained. (See the Further Readings chapter for details.)

We will content ourselves with hinting at the “nearly-all” proof, which is easier to grasp than the final “all” proof. The construction relies on a network of line segments called the *straight skeleton*, which coincides with the creases we have used prior to adding perpendiculars. The mountain-fold creases in Figure 5.5(a) constitute the straight skeleton of the rectangle. They can be viewed as the tracks of the four corners of the rectangle if it were continuously enlarged or continuously reduced by moving all sides parallel to themselves at the same rate toward the exterior or interior of the rectangle. See Figure 5.10. The central crease xy corresponds to the rectangle “winking out” to a segment.

This enlargement/reduction process might be clear for a rectangle or triangle, but it becomes less evident for shapes with “dents” or concavities. Figure 5.11(a) shows the magnification and reduction for the boundary in a crude representation of the letter A, resulting in the straight skeleton shown in (b). We will not pause to define the straight skeleton precisely, but you can sense that it bisects the angle at every vertex of the shape and produces a network of segments with its own set of vertices where creases meet. Just as in the rectangle and triangle cases, each of these vertices may need one or more perpendicular creases to reach a flat folding that aligns all edges. For example, the seven perpendiculars added to Figure 5.11(c) suffice for the A-shape, as you can verify by folding the template.

These are the two main ingredients to the proof: (1) Construct the straight skeleton of the drawing; (2) Add perpendiculars incident to every skeleton vertex. It turns out that not every vertex needs a perpendicular, but that’s another detail we will not pursue. Then it remains to be shown that this construction leads to a flat folding aligning exactly the segments of the drawing on top of one another.

For drawings more intricate than we have so far illustrated, the behavior of the perpendiculars can be complex, as is hinted at in the two-shape Figure 5.9 and especially for the turtle in Figure 5.12. Notice that some perpendiculars in the turtle example “wander” about the drawing, from their start at a skeleton

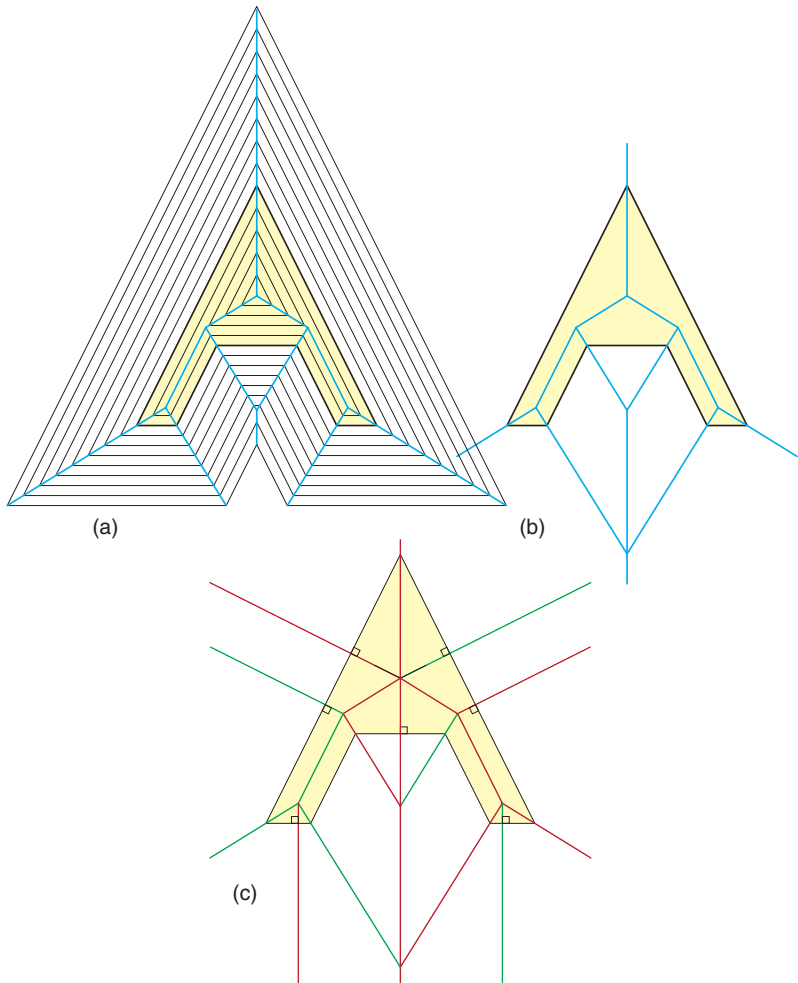


Figure 5.11. (a) Growing and shrinking the shape. (b) Straight skeleton. (c) Perpendiculars added, and mountain/valley folds indicated by color.

vertex, before escaping to the border of the paper – for example, those emanating downward from the base of the turtle’s neck. It turns out that under contrived circumstances, this wandering can go on forever, spiraling densely without leaving the boundaries of the paper. And this is why this proof works for only “nearly all” drawings: For some special drawings, it requires an infinite number of creases, and therefore an infinitely thick stack of paper to scissor through!

Exercise 5.4 (Practice) *Verify Turtle Flattening.* Check that each of the 24 vertices (junctions of creases) in Figure 5.12 satisfies the Maekawa-Justin Theorem (Theorem 4.2), as they must in order to fold flat.

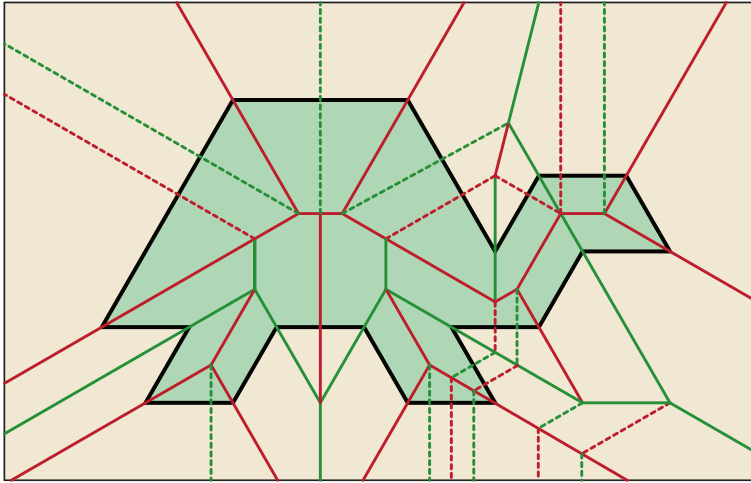


Figure 5.12. Turtle. Perpendiculars are dashed.

Finally, in Figure 5.13 we make good on the claim in the Preface that the paper dolls can be cut out with a single slice!

5.3 Above & Beyond

The same researchers who discovered the Fold and One-Cut Theorem (Theorem 5.1) asked an interesting related question: Can every polyhedron be flattened? The general idea here is perhaps best seen with the natural flattening of a box shown in Figure 5.14. A box is among the simplest of the class of 3D objects known as *polyhedra*, which will be the focus of Part III of this book. A *polyhedron* is any 3D shape whose surface is composed of flat polygons, known as its *faces*, and which is *closed* in the sense that it could enclose water without leaking at any orientation. The box has six faces, each a rectangle. Polyhedra can be quite complicated. For example, the elegant head in Figure 5.15 is a polyhedron of more than 2,000 faces. (Note its eyes are paneled with flat faces to ensure the surface is closed.) The posed question applied to this example asks: If this polyhedron were a hollow paper model, could it be creased and folded flat? During the folding and flattening, the surface should not tear or pass through itself, although it is perfectly permissible to bend the faces as much as needed during the “crushing” process. We imagine poking a pinhole in the surface to let the enclosed air escape.

This problem can be viewed as fold and one-cut in one higher dimension. When describing objects, let’s use 1D, 2D, and 3D to mean one-, two-, and three-dimensional, respectively. In the fold and one-cut problem, we have a drawing made of 1D straight segments on a 2D sheet of paper, and the goal is to flatten all segments to the same 1D line. The polyhedron-flattening problem

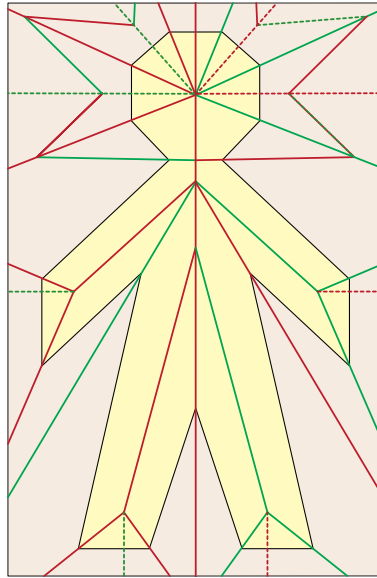


Figure 5.13. Fold and-One-Cut creases for one of the paper dolls in Figure 0.1. Fold down the vertical centerline first, then fold the octagonal head from the top down to the neck, then fold the remainder. At spots the flattened paper is 14 sheets thick, so folding all four people in Figure 0.1 at once will require cutting through 56 sheets!

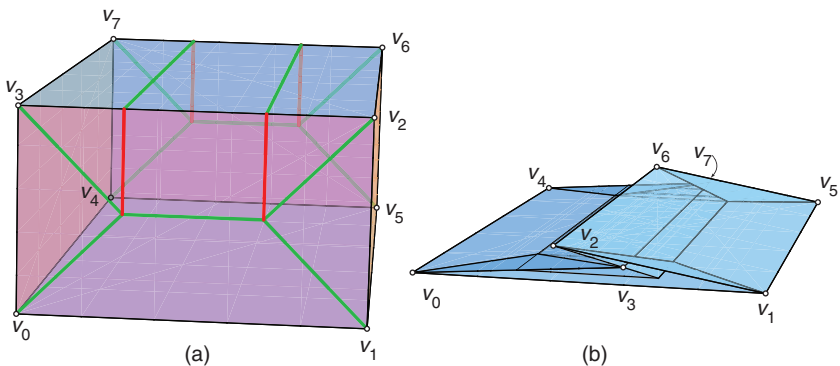


Figure 5.14. (a) A $10 \times 6 \times 20$ box. (b) Flattening according to the creases in (a).

starts with a polyhedron (the object comparable to the drawing) made of 2D flat faces in 3D space, with the goal of flattening all faces to the same 2D plane:

fold and one-cut:	1D segments	on 2D paper	fold to 1D line
polyhedron flattening:	2D faces	in 3D space	fold to 2D plane

Just recently, the flattening question was half-answered: YES, every polyhedron has a flattened equivalent. The reason this is only half an answer is that

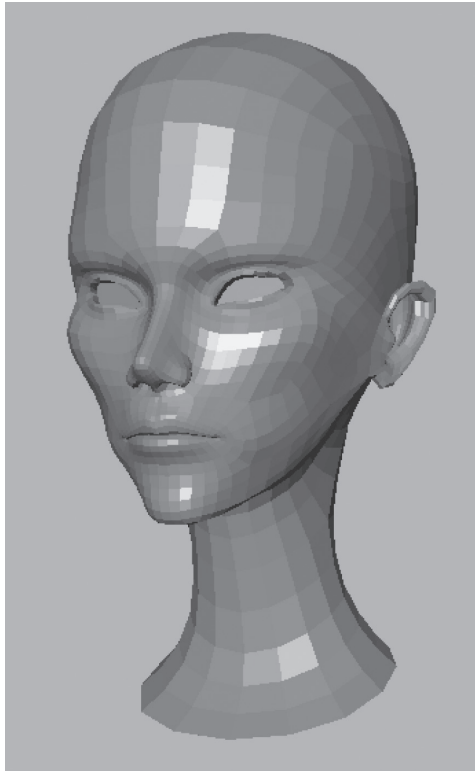


Figure 5.15. A polyhedron of 2,290 faces, most of them quadrilaterals.

it remains unknown whether there is a continuous motion from the original polyhedron to its flat version that avoids tearing and paper penetration. So the half-result can be interpreted as saying that the polyhedron surface can be cut up and reassembled to a flat version, taping back together the same two sides of every cut. But it remains unresolved whether there is a flattening that never needs to cut the surface, or – equivalently – whether the flattened version can be continuously “inflated” back to the original polyhedron. This gap remains despite the intuition that if you simply squash the polyhedron in Figure 5.15 with your boot (remembering to pre-puncture with a pin hole to let the air escape), it will indeed inexorably flatten.

OPEN PROBLEM: *Flattening Polyhedra*

Can every polyhedron be creased and then continuously flattened?

Exercise 5.5 (Challenge) *Flattening a Cube.* Find creases that permit flattening a cube.

6

The Shopping Bag Theorem

In Part I of this book, the linkages we studied all used rigid one-dimensional links. Although the flat origami foldings we've examined in the previous two chapters end up with flat "faces" between the creases, it is a rare origami folding that can reach its final folded state without bending the faces during the construction process. Origami designs that tuck flaps of paper into pockets, such as the map-folding puzzle in Figure 4.14, clearly require face bending. Even forming the standard origami crane (Figure 4.1) is impossible if the paper is only bent along the final creases. As long as bendings do not become creasings, the properties of paper ensure that the *faces* of the design – the regions *bounded* (i.e., outlined) by the creases – can be smoothed flat in the final design.

The new field of *rigid origami* is focused on designs that can be folded without bending the faces. We can think of the faces as rigid steel plates, hinged along creases to the adjacent plates. Although rigidity limits the range of designs that can be folded, those that can fold often have useful applications.

We set as the goal of this chapter proving the surprising result that the standard grocery shopping bag (Figure 6.1) cannot fold flat without bending the flat portions of the bag. Of course collapsing a grocery bag requires bending at its creases, but it also requires bending its flat faces. Another way to express this result is that, were the bag built of metal plates and hinges, then an open shopping bag would be rigid: it could not flex at all. Similarly, a closed shopping bag would be stuck in its flattened configuration. This is in fact one of the reasons the shopping bag is such a successful engineering design: Because its faces are stiff (but not rigid), it tends to stay open rather than spontaneously collapsing, because collapsing requires bending the resistant bag paper.



Figure 6.1. A shopping bag.

In the previous two chapters, we concentrated on static, 2D origami designs. In contrast, the focus in rigid origami is on 3D, and in particular on the 3D dynamics as the creases are hinged. Here we see again the dihedral motions we encountered in [Chapter 3](#) – for example, in [Figure 3.1](#) – but now the motions are more intricate because they are related in patterns more complicated than those along a single fixed-angle chain. Before turning to the Shopping Bag Theorem, we examine two rigid origami folds that demonstrate these beautiful 3D motions.

6.1 Two Rigid Origami Examples

6.1.1 The Miura Map Fold

The Japanese astrophysicist Koryo Miura designed a clever rigid origami fold in the 1970s for the purpose of unfolding a satellite solar array in space. The individual solar panels are inflexible, so rigidity must be maintained. A version of his fold was used for the Japanese radio astronomy satellite launched in 1997. Its solar panels unfold to 10 meters in diameter; see [Figure 6.2](#).

The panel unfolding is based on what is now known as the *Miura map fold*. Starting with a rectangular piece of paper, it is made with horizontal creases forming strips and zig-zag vertical creases in the mountain-valley pattern shown in [Figure 6.3\(a\)](#). Because the creases partition all but the left and right portions of the paper into parallelograms, it is more natural to start, not with a rectangle, but with a piece of paper whose boundaries have been snipped to match the parallelograms, as in (b) of the figure.

Exercise 6.1 (Practice) *Miura Map Creases*. Verify that the Maekawa-Justin Theorem ([Theorem 4.2](#)) is satisfied at each vertex of [Figure 6.3](#).

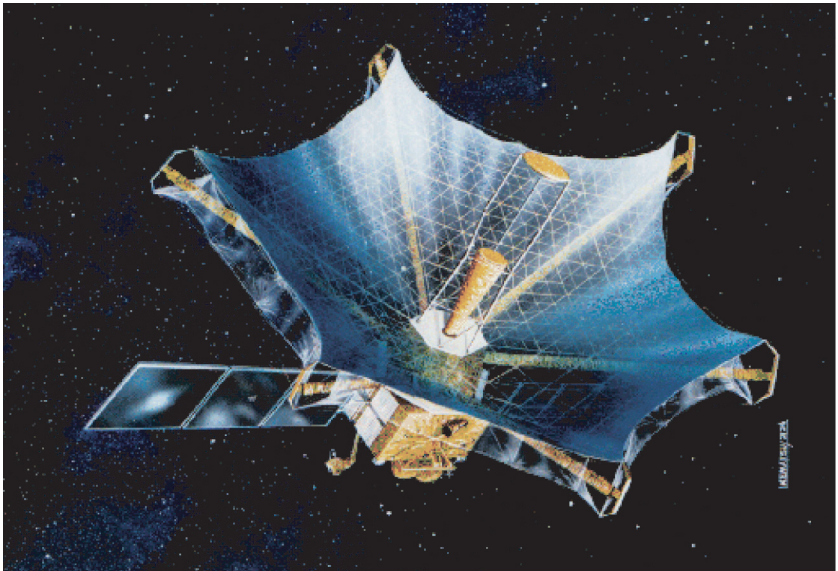


Figure 6.2. The Japanese radio telescope with its solar panels unfolded.

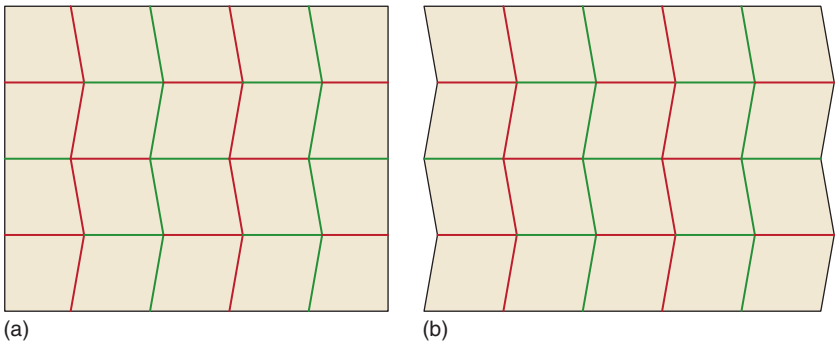


Figure 6.3. (a) Miura Map folding pattern for rectangle. (b) For nonrectangular paper tiled by parallelograms.

If the paper is creased sharply and cleanly, and squeezed between diagonally opposite corners, it collapses beautifully, as depicted in Figure 6.4. As its name implies, the Miura Map fold is especially useful for folding maps: see Figure 6.5. As these figures indicate, bending occurs at all the creases simultaneously – rather unlike the usual origami folding. The ultimate end result is that the entire sheet of paper folds flat into the space of one of its “primal” parallelograms. It is this remarkable compaction that makes the Miura Map Fold so useful for stowing and deploying solar panels. Or maps!

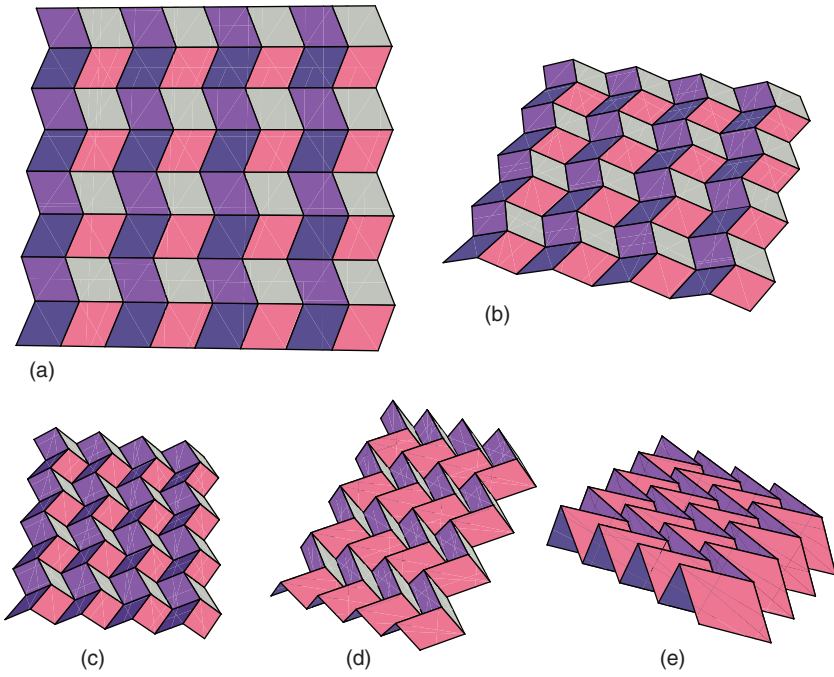


Figure 6.4. A Miura Map folding collapsing from (a) flat to partially collapsed (e).



Figure 6.5. Various Miura map foldings.

However, the emphasis in this chapter is another essential property for the application: The parallelogram panels do not bend during the 3D collapsing motion (or the reverse expanding motion). Thus they can be rigid solar cells joined by hinges along the folding segments. Let's look at another beautiful rigid origami motion known as the "square twist."

Exercise 6.2 (Practice) *Eight Rigid Right Triangles.* Crease a square in half twice, then twice more along its diagonals, to achieve the mountain/valley crease pattern shown in Figure 6.6. Convince yourself that it folds rigidly to a star-like shape, an 'x' when viewed from above the central vertex.

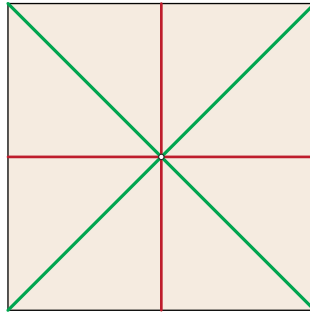


Figure 6.6. A crease pattern to fold as rigid origami. (Exercise 6.2)

6.1.2 The Square Twist

The "square twist" is a way of rigidly folding the crease pattern shown in Figure 6.7, leading to the pleasing motion shown in Figure 6.8. When flattened, the original square sheet is compacted to a square half the size. This fold is often used repeatedly to make intricate (and usually colorful!) "origami tessellations." We will see that the pattern of creases at the four vertices of the twisting square is the same pattern of creases found on the side of a shopping bag.

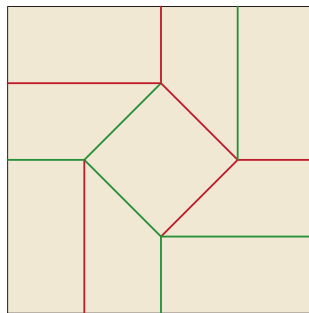


Figure 6.7. Crease pattern for the square twist.

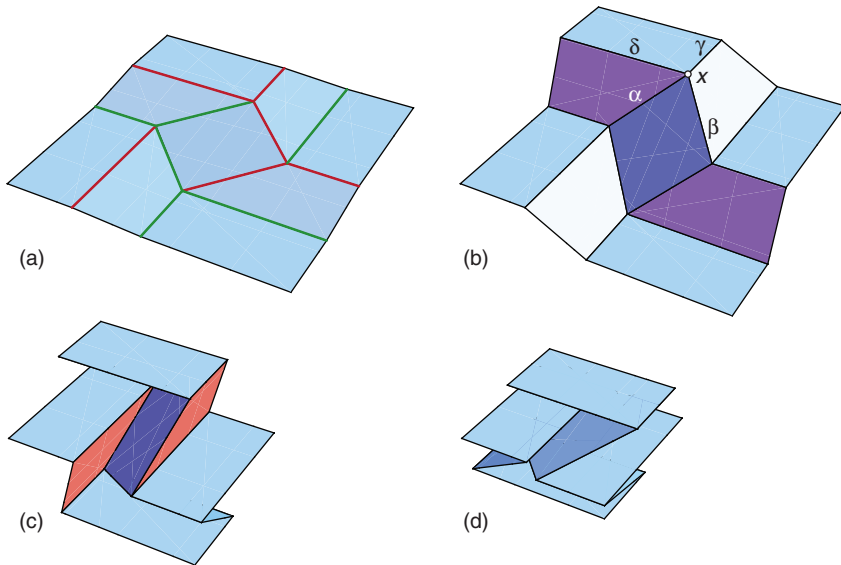


Figure 6.8. The square twist. (a) is slightly creased from Figure 6.7. By (d), the folding is nearly flattened to one-half its original size. Labels in (b) correspond to Figure 6.10(b) further in the chapter.

6.2 Dihedral Angle Constraints

If you try the Miura Map fold or the Square Twist, you can sense that the motions depicted in Figures 6.4 and 6.8 are forced: Once you start to fold one crease, all the others follow in lock step, and simultaneously. Indeed this is true for degree-4 vertices, which are the only type of vertices that appear in either construction. (See p. 61 for a reminder of the meaning of “degree” in this context.) The forcing relationship is mathematically rather complicated. The only proofs I know use 3D geometric techniques beyond what we can present here. However, although the precise justification may be technically difficult, the relationship itself is quite believable, especially if you play with the constructions. So the way we will proceed is to gather together in this section a number of lemmas – none of which we will prove, but each of which is, I hope, believable. Accepting these lemmas will permit us to sketch a proof of the Shopping Bag Theorem.

We need four lemmas, each of which concerns 3D dihedral angles. When concentrating on flat foldings in Chapter 4, we naturally focused on the angles between creases as measured within the paper surface – planar angles. But with rigid origami, which is fundamentally about 3D motions, it is the angles in 3D that matter. In Chapter 3 (p. 39) we introduced the notion of a *dihedral angle*: the angle in 3D between two planes. That is the key notion we need now, because that is precisely what is forced in the rigid motions we have just seen.

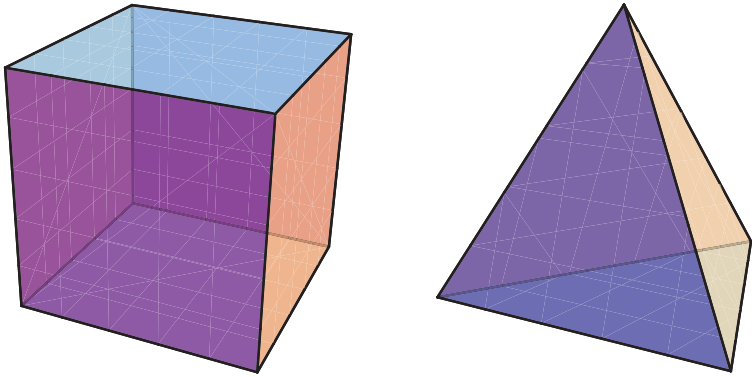


Figure 6.9. A cube and a tetrahedron. (Exercise 6.3)

Exercise 6.3 (Practice) Dihedral Angles. Refer to Figure 6.9. Calculate: (a) the (internal) dihedral angle between a pair of faces of a cube that share an edge, and (b) the (internal) dihedral angle between a pair of faces of a regular tetrahedron. [Trigonometry is needed for the tetrahedron.]

Box 6.1: Greek Letters for Angles

In honor of the ancient Greek mathematician Euclid, the “Father of Geometry,” it is common to use Greek letters to represent angles. The use is not only honorific. It is often much less mentally taxing to mark an angle α than to cite it as, say, “ BAC , which is typical in U.S. high-school textbooks. The first four letters of the Greek alphabet, corresponding to the English a, b, c, d , are used the most frequently. See the table below.

Name	Symbol
alpha	α
beta	β
gamma	γ
delta	δ
theta	θ

When just one angle needs to be identified, the common convention is to use θ . Or when many need identification (as in the Kawasaki-Justin Theorem 4.4), subscripting θ is typical: $\theta_1, \theta_2, \dots$. We will follow these conventions – consider it training for reading professional mathematics!

6.2.1 Degree-4 Lemma

The first lemma we need is the mathematically deepest:

Lemma 6.1 (Degree 4)

If four creases meet at a vertex in the middle of the paper (surrounded by 360° of paper), then the dihedral angles of opposing creases are equal.

The lemma is easiest to see when the creases form perpendicular compass directions, as illustrated in Figure 6.10(a). The crease out the East side of x must be the same as the crease to the West side in the sense that the 3D dihedral angle along segment ax must equal that along the segment xc : so $\alpha = \gamma$. The two are tied together through the collinearity of $\{a, x, c\}$. (Recall from p. 32 that “collinear” means “lying on the same line.”) The same holds for the North-South creases bx and xd : $\beta = \delta$. The lemma is far less obvious when the creases are not perpendicular, but it remains true, for example, for each vertex of the Square Twist construction shown in Figure 6.10(b). Again there is forced equality between the dihedral angles of opposing creases: $\alpha = \gamma$ and $\beta = \delta$. This can best be seen in the top left corner of Figure 6.7(b) – or better yet, by folding the square twist yourself!

6.2.2 Plus-Sign Lemma

The second lemma is easy to see, but we first need to establish some conventions on the exact meaning of the dihedral angle at a crease. A dihedral angle of 180°

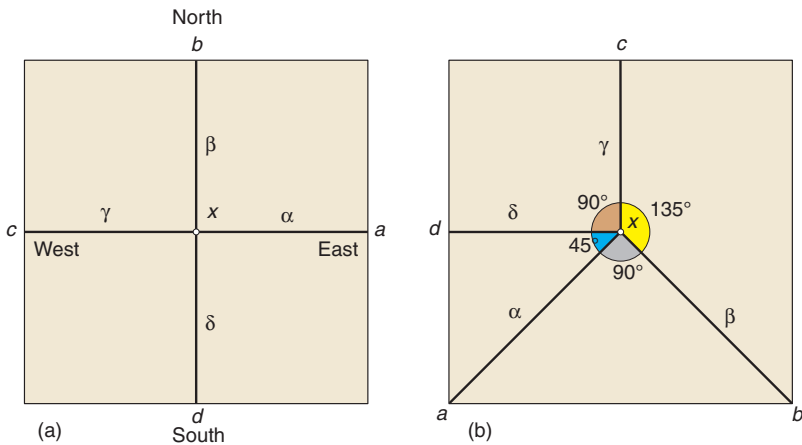


Figure 6.10. (a) Creases meeting orthogonally. (b) Creases meeting as in the Square Twist fold, Figure 6.8(b). In both cases, it must be that $\alpha = \gamma$ and $\beta = \delta$.

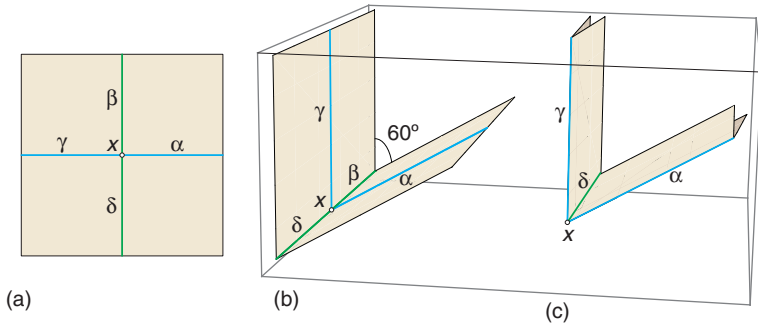


Figure 6.11. (a) The + sign creases. When β and δ are folded at some nonflat dihedral angle (here 60°), then either $\alpha = \gamma = 180^\circ$ as in (b), or $\alpha = \gamma = 0^\circ$ as in (c).

means entirely uncreased, whereas a dihedral angle of $0^\circ = 360^\circ$ is a sharp fold bringing the paper on either side of the crease into contact (whether a mountain or valley fold will not matter). These extreme dihedral angles correspond to the flat foldings studied in Chapter 4, so we will call them *flat angles*. Nonflat dihedral angles – angles intermediate between 0° and 180° – have a fundamental 3D structure.

Lemma 6.2 (Plus Sign)

If four creases meet perpendicularly (as in Figures 6.10(a) and 6.11(a)) and an opposing pair is folded to a nonflat dihedral angle, then the other pair must be folded to a flat angle.

I encourage the reader to fold a + sign into a sheet of paper and experiment as in Figure 6.11 until the claim of the lemma feels right.

6.2.3 Degree-3 Lemma

The third lemma we need says, essentially, degree-3 vertices are impossible.

Lemma 6.3 (Degree 3)

If exactly three creases are incident to a vertex x , no two of which lie on the same line, then all their dihedral angles are 180° (i.e., they are not creases), so x is not a true vertex – the neighborhood of x is flat.

(Recall from p. 61 that a “neighborhood” of a point is a small region surrounding the point.)

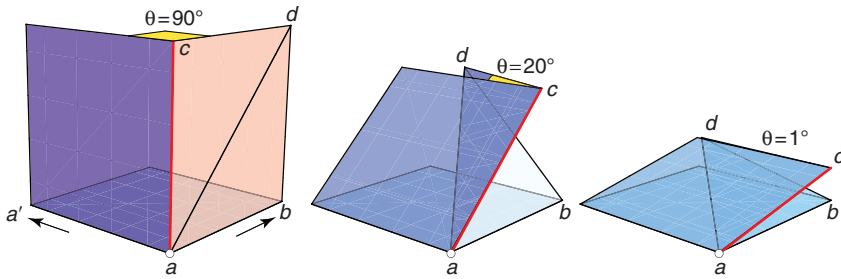


Figure 6.12. Collapsing the corner of a shopping bag.

The Even-Degree Theorem (Theorem 4.1) from Chapter 4 showed that a vertex must have even degree, but that was in order to fold flat. Lemma 6.3 is saying that, even in 3D, under the rigid origami model, you cannot fold the paper to a surface where the creases create a vertex of degree 3. Look again at the configuration in Figure 6.10(b). If the segment dx is increased ($\delta = 180^\circ$), then the degree-4 vertex x reduces to degree-3 and the Degree-4 Lemma (Lemma 6.1) then says that bx must be increased as well ($\beta = 180^\circ$). Now x is reduced to degree 2. And a degree-2 vertex is only possible if the two incident creases align along the same line, in which case x is not a true vertex.

6.2.4 Bag Corners

The last lemma we need concerns the corner of a shopping bag, where a vertex is surrounded by 270° of paper – three 90° angles – rather than the usual 360° of paper. We will specialize the lemma to just what we need:

Lemma 6.4 (Bag Corner)

If the dihedral angle θ in Figure 6.12, along the vertical crease incident to the corner a , is 0° , then the whole corner is flattened.

There is more that could be claimed to justify the 3D motion illustrated in the figure, but we only need that closing that one dihedral angle forces the entire construction to fold flat. Again this lemma becomes plausible through physical manipulation of a such a corner.

6.3 The Shopping Bag Theorem

Let a grocery shopping bag (Figure 6.1) have width W , depth D , and height H . At my grocery store, $W = 12$ inches, $D = 7$, and $H = 17$. Because $H \geq \frac{1}{2}D$, two 45° valley creases from the bag corners meet at a point d on the side of the bag, as

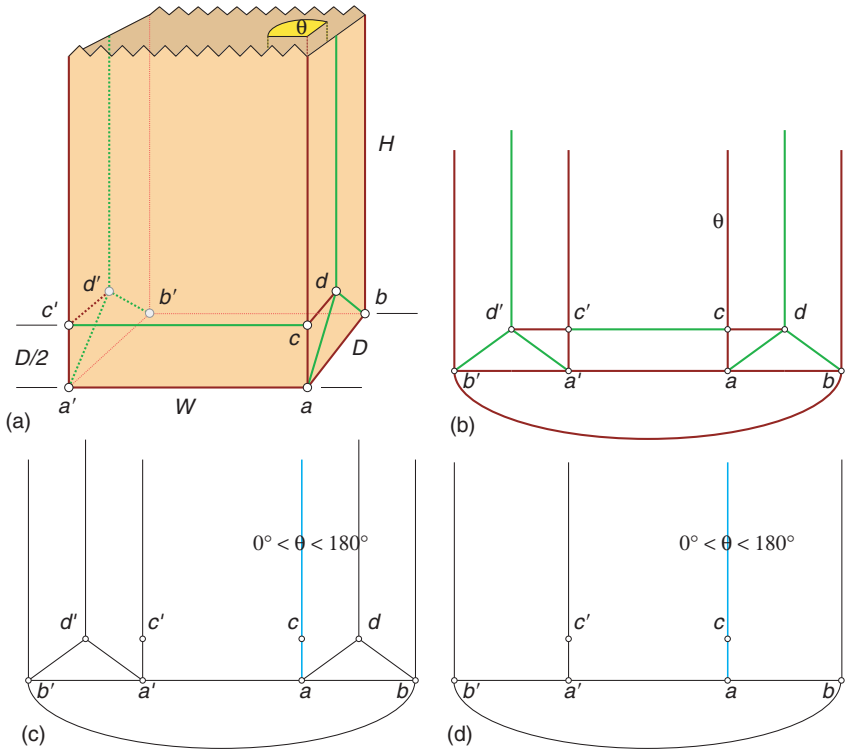


Figure 6.13. (a) Model of a “tall” open grocery shopping bag. θ is the dihedral angle between the front and right faces. (b) Network of creases between vertices. The arc from b to b' represents the crease along the back of the bottom. (c) When $0^\circ < \theta < 180^\circ$, creases cd , cc' , $c'd'$ disappear. (d) And now the creases incident to d and d' disappear, leaving an open shopping bag.

illustrated in Figure 6.13(a). We say that any bag with $H \geq \frac{1}{2}D$ is a *tall shopping bag* (to distinguish it from a short bag latter).

Crease cc' in the figure is the valley fold that permits collapse of the bag, at which time cc' is tucked behind the aa' mountain fold. Call any state of the bag – open, closed, or intermediate – a *configuration* of the bag.

Theorem 6.1 (Shopping Bag)

If the faces of a tall shopping bag are rigid, then the bag may be either fully opened (as illustrated) or fully collapsed flat. It has no other configurations. In particular, the bag is rigid in either the opened or flattened configuration.

We now offer a proof sketch of this theorem by repeated use of Lemmas 6.1–6.4. It is a sketch rather than a formal proof, not only because we did

not prove any of the lemmas we are using, but also because even accepting those lemmas, we still rely on intuition at several junctures. Warning: The proof requires some patience and concentration as it marches methodically through cases and subcases!

We label the vertices as in Figure 6.13(a). Part (b) of the figure displays the network of crease connections between the vertices in a way that will make the reasoning easier to follow. We partition the analysis into three cases that exhaust all possibilities, with the cases depending on the dihedral angle (call it θ) along the vertical crease incident to vertex c from above: θ is nonflat, $\theta = 0^\circ$ (closed flat), and $\theta = 180^\circ$ (opened flat).

Case 1. θ is nonflat, so it lies between 0° and 180° . This case covers the standard open bag, when $\theta = 90^\circ$. Figure 6.13(a,b) shows that c is a degree-4 vertex at which the creases meet perpendicularly. So the Plus-Sign Lemma (Lemma 6.2) applies, and says that the cross creases must have dihedral angle either 0° or 180° . We consider each option in turn.

(a) The dihedral angle along cc' and cd is 180° , that is, there is no crease. Applying the Degree-4 Lemma (Lemma 6.1) to c' says that the outgoing crease $c'd'$ must also have dihedral angle 180° —again, no crease. So the network of creases has been reduced to that shown in Figure 6.13(c).

Now apply the Degree-3 Lemma (Lemma 6.3) to vertex d (and symmetrically, d'): These degree-3 vertices must be flat – in effect, not there either. Now the network is reduced to Figure 6.13(d), which shows that both the front of the bag, and its sides, are uncreased. And so we have the fully open configuration of the bag.

(b) The dihedral angle along cc' and cd is 0° , that is, it is folded closed flat. If the dihedral angle along cd is 0° , the Degree-4 Lemma (Lemma 6.1) says the outgoing crease db is also 0° . This forces the right side of the bag to fold flat, and the left side follows symmetrically. Thus, both the front and sides are folded flat, and we are in the flat configuration of the bag.

Case 2. $\theta = 0^\circ$. The Degree-4 Lemma (Lemma 6.1) implies that the crease ca incident to the bag corner is also at 0° . The Bag-Corner Lemma (Lemma 6.4) now says that the entire corner is flat, which then forces the whole bag to fold flat.

Case 3. $\theta = 180^\circ$. This last case is impossible. If we don't crease down the verticals incident to a and to a' , then the bag splays out (roughly like the network drawings in Figure 6.13[b,c,d]), and there is just not enough paper to wrap around the back and close the bag.

Thus all lines of reasoning lead to just the two configurations of the bag: either fully opened, or folded flat in the standard manner. Because there are just two isolated configurations, the bag cannot rigidly move between them. Next time you open or close a shopping bag, see if you can sense the bending of the faces, which this theorem says are necessary for any movement to be possible!

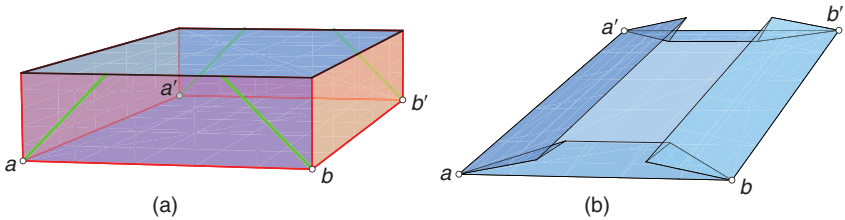


Figure 6.14. (a) A box with H short enough so that the valley creases from the corners do not meet. (b) Flat folding of the box.

Gift-Box Folding. We called the standard shopping bag a tall shopping bag because its height is at least half its depth, $H \geq \frac{1}{2}D$. If instead $H < \frac{1}{2}D$, then the vertices d and d' on the side of the bag disappear, because the 45° creases from the bag corners do not have enough height H to meet. The result is something closer to a gift box (without a top), and indeed this can now fold to a flat configuration in the rigid origami model, with all intermediate positions achievable. See Figure 6.14. (Note that this is different from the flattening of a box considered in Chapter 5 (Figure 5.14), which permitted bending of the faces.) Given that our proof sketch of the Shopping Bag Theorem 6.1 traced logic along the network of creases, it is perhaps not surprising that if that network is fundamentally altered, the proof no longer holds.

6.4 Above & Beyond

You can think about the rigid origami examples in this chapter as objects that transform between two different shapes by movements articulated at hinges: from a compact storage form to the fully unfolded solar panels; from an open gift box to a flattened gift box, and so on. In this sense, rigid origami constructions are akin to the “transformer toys” so popular now. The mathematical equivalent of a transformer toy is a 3D *hinged dissection*. There is a *dissection* between a pair of objects if there is a partition of one object into pieces that may be reassembled to form the other object. A hinged dissection is a dissection where the pieces are hinged at vertices so that the whole hangs together as a single piece (it is *connected*), and the transformation can be effected by rotations about the hinges. 2D dissections have been a staple of puzzle enthusiasts for over a century, and there is an extensive literature on the topic. But here we concentrate on 3D, where less is known. Any rigid origami construction with two preferred configurations can be viewed as a 3D hinged dissection where the pieces of the dissection happen to be thin: the rigid faces of the construction.

Exercise 6.4 (Practice) Four Hinged Squares. Figure 6.15 shows a 4×1 rectangle made from four unit squares hinged together. Can it be hinged in the plane to produce a 2×2 square so that squares 1 and 4 are diagonally opposite, that is,

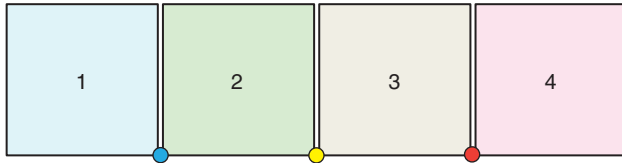


Figure 6.15. Four squares hinged to a rectangle. (Exercise 6.4)

they share only a corner, not a side? During the hinging motions, one square is not permitted to overlap another.

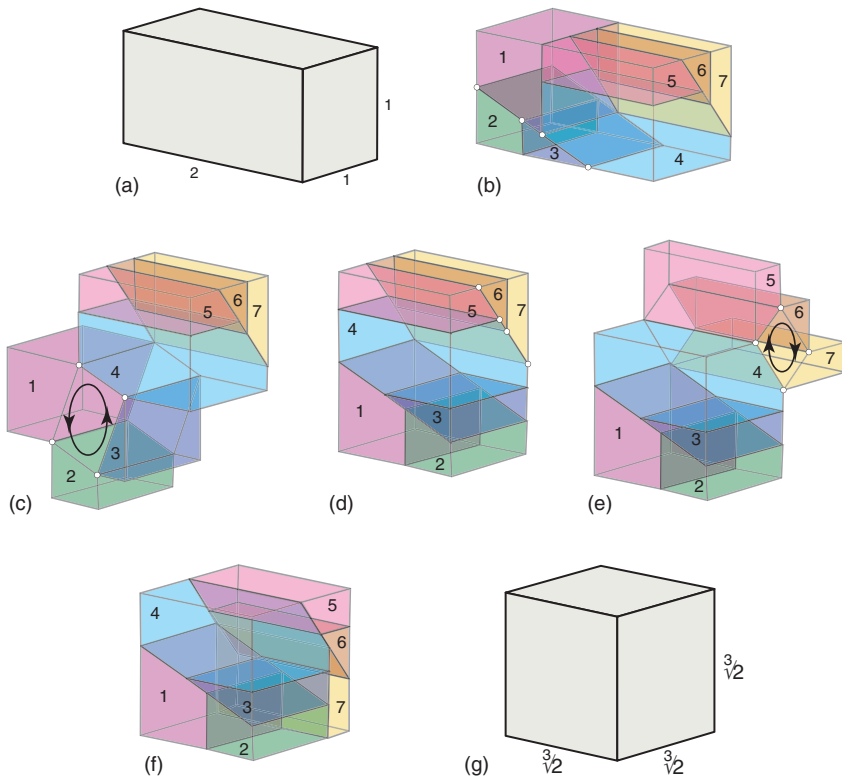


Figure 6.16. Hanegraaf's hinged dissection of a $2 \times 1 \times 1$ block (a) to a cube (g).

Although many fascinating hinged dissections of 3D objects have been found by now, to me the pinnacle of the art is the amazing hinged dissection between a $2 \times 1 \times 1$ rectangular block and a cube shown in Figure 6.16. This dissection was found in the 1960s by Anton Hanegraaf, a Dutch civil engineer with a passion (and talent!) for dissections, answering an open problem circulating at the time. Dissections between this pair of objects were known, but not hinged

dissections. His elegant 7-piece hinged dissection resolved the issue. The seven pieces into which the block in Figure 6.16(a) is partitioned are shown in (b). Four hinges along a line connect pieces {1, 2, 3, 4} together. These rotate counterclockwise (c), carrying pieces {5, 6, 7} along, to produce an intermediate block (d). Four different hinges, again along a line, connect pieces {4, 5, 6, 7}, which rotate clockwise (e) to produce the cube (e, f). Because the cube must have the same volume as the $2 \times 1 \times 1$ block, the cube's side length is $\sqrt[3]{2}$, the cube root of 2, about 1.26.

Exercise 6.5 (Practice) *Hanegraaf Block*. What are the dimensions of the intermediate block in Figure 6.16(d)?

A famous question posed by the mathematician David Hilbert in 1900 asked if any pair of objects (polyhedra) with the same volume have a dissection into polyhedral pieces of one to the other. The answer no soon followed: For example, there is no dissection of a regular tetrahedron to a cube.

Another question, at least implicit a century ago, asks: If two 3D objects do happen to have a dissection, then do they also have a hinged, dissection? Although this is not definitively settled as of this writing, there has been a recent breakthrough on the same question in 2D, which was also unresolved. A group of researchers (including three college students) proved that every pair of 2D shapes of the same area have a hinged dissection, building on the long-known fact that, unlike in 3D, there is, always a dissection in 2D. Whether their proof technique can extend to 3D remains unclear. In any case, this venerable topic is very much alive today.

Polyhedra

The third and final part of this book explores folding and unfolding the surface of a polyhedron, a 3D solid shape whose surface is made of flat *faces*. We break the tradition of the previous two parts of always including at least one beautiful theorem in each chapter, for a central question in unfolding polyhedra has so resisted solution that there are as yet no general theorems. We explain this central open problem, “Dürer’s Problem” for convex polyhedra, in the next chapter, and follow that with a variation for “orthogonal polyhedra” on which there are results to report. We close with the inverse of unfolding, folding a piece of paper to a polyhedron, which has at its core a beautiful and powerful theorem of the Russian geometer Alexandr Alexandrov. Investigation of folding polyhedra has led to many surprises and leads to several unsolved but accessible problems for the reader to ponder.

7

Dürer's Problem: Edge Unfolding

7.1

Albrecht Dürer's Nets

In 1525, the German painter and thinker Albrecht Dürer published his masterwork on geometry, whose title translates as “On Teaching Measurement with a Compass and Straightedge.” The fourth part of this work concentrates on polyhedra: the Platonic solids, the Archimedean solids, and several polyhedra “discovered” by Dürer himself. Figure 7.1 shows his famous engraving, “Melencolia I,” in which he used a polyhedron of his own invention a decade earlier. His book presented each polyhedron by drawing a *net* for it: an unfolding of the surface to a planar layout. The net makes the geometry of the faces and the number of each type of face immediately clear to the eye in a way that a 3D drawing, which necessarily hides part of the polyhedron, does not. Moreover, a net almost demands to be cut out and folded to form the 3D polyhedron. Figures 7.2 and 7.3 show two examples of Dürer's nets. The first is a net of the *snub cube*, which consists of six squares and 32 equilateral triangles. The second is a net of the truncated icosahedron, consisting of 12 regular pentagons and 20 regular hexagons. We know the spherical version of this polyhedron as a soccer ball.

Dürer's nets, an apparently original representational invention, have since become a standard presentation method for describing polyhedra. For example, Figure 7.4 shows a modern display of nets for the so-called Archimedean (or *semiregular*) solids. But in the nearly 500 years since Dürer's work, no one has proved that a net exists for every convex polyhedron, even though there are often several or even many different nets for any given polyhedron. It is this long-unsolved problem that we will examine in this chapter.

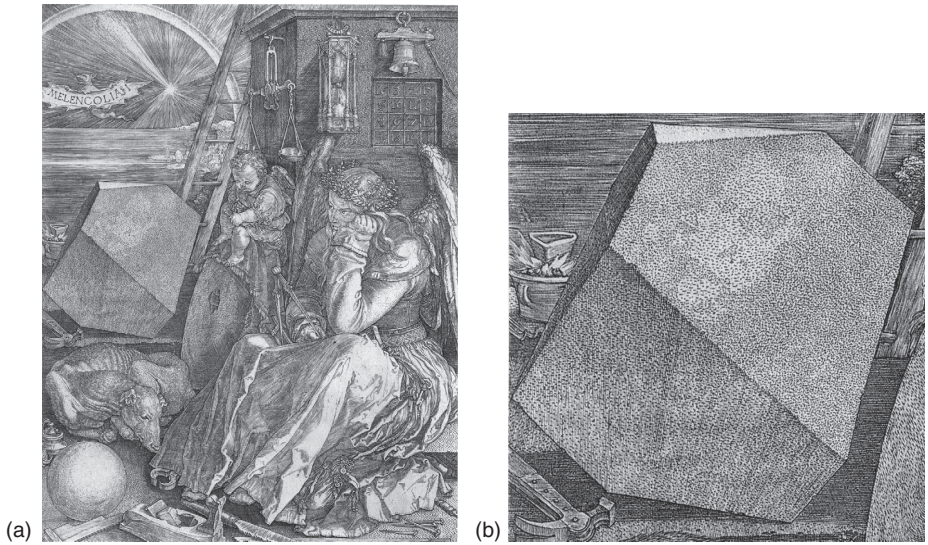
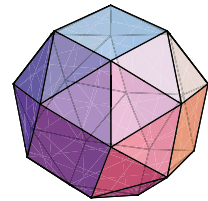
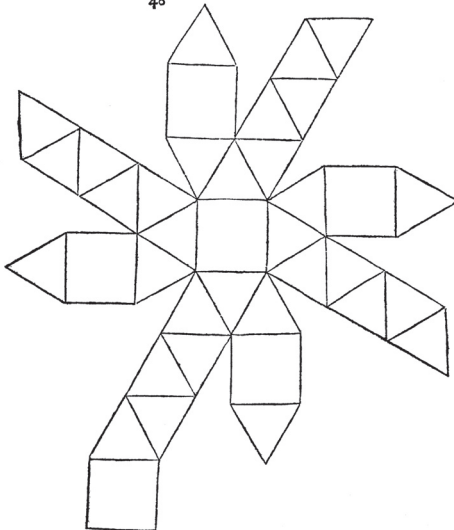


Figure 7.1. Dürer's 1514 engraving, "Melencolia I," with polyhedron enlarged.

Das Sechß corpus/so das außßan wirt/hat es sechs geferte/ond zwey vnd dreyßig Dyanßliche fider/so man das zúßamen leget/gerwin es vier vnd zweyßzig eck/ vnd sechsßig schärpffer seyen.

40



Das Sibent vnd nach folget corpus/so es offen leyt/hat es sechs acht eckter/ond acht sechs eckter/ond zweß vier eckter fider/ vñ so man das zúßamen leget/so gerwin es acht vnd vierßzig eck/ond zwey vnd sibentßig schärpffer seyen.

Figure 7.2. Dürer's net for a snub cube.

L In anders das mach auß zweynsig sechseckter flachen seibern/ gleichförmig vnd wincklich/ so man darzu thut zwey fünffteckter flacher seibern/ so die gleichförmig gegen den sechsecken seiden sind/ vnd in jnen seibe auch gleichwincklich vnd ewenlich an einander gesetzt wesen/ wie ich das offen im plano hernach hab außgerissen/ Es man dann das alles zusamen schleußt/ so würt ein corpus daraus/ das gewinnet zwey vnd sechsig eck/ vnd neünzig schärpff seuen/ die Corpus rüret in einer helen kugeln mit allen seinen ecken an.

43

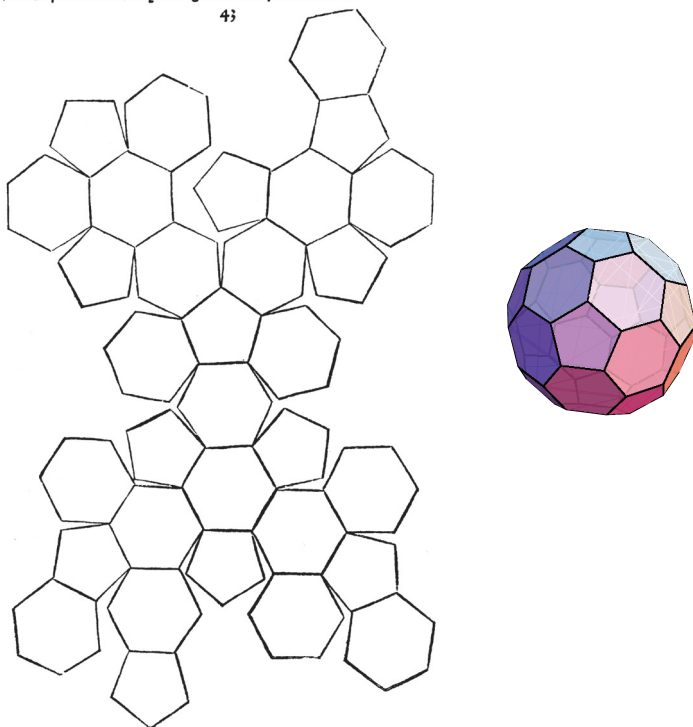


Figure 7.3. Dürer's net for a truncated icosahedron.

7.2

Convex Polyhedra

To understand Dürer's problem, we need to understand what a polyhedron is, what specifically a convex polyhedron is, and what a net is.

Polyhedron. A *polyhedron* is the surface of a 3D solid object, composed of flat, convex polygonal *faces*. We will be concerned in this chapter only with polyhedra without holes, so we exclude polyhedral tori and other such objects that could be hung on a thread through a hole. Each face of a polyhedron is bounded by *edges* that are straight segments. Edges meet at *vertices*, the sharp corners on the surface. We briefly introduced the concept of a polyhedron in Chapter 5, but for our current purposes, we need this more precise characterization.

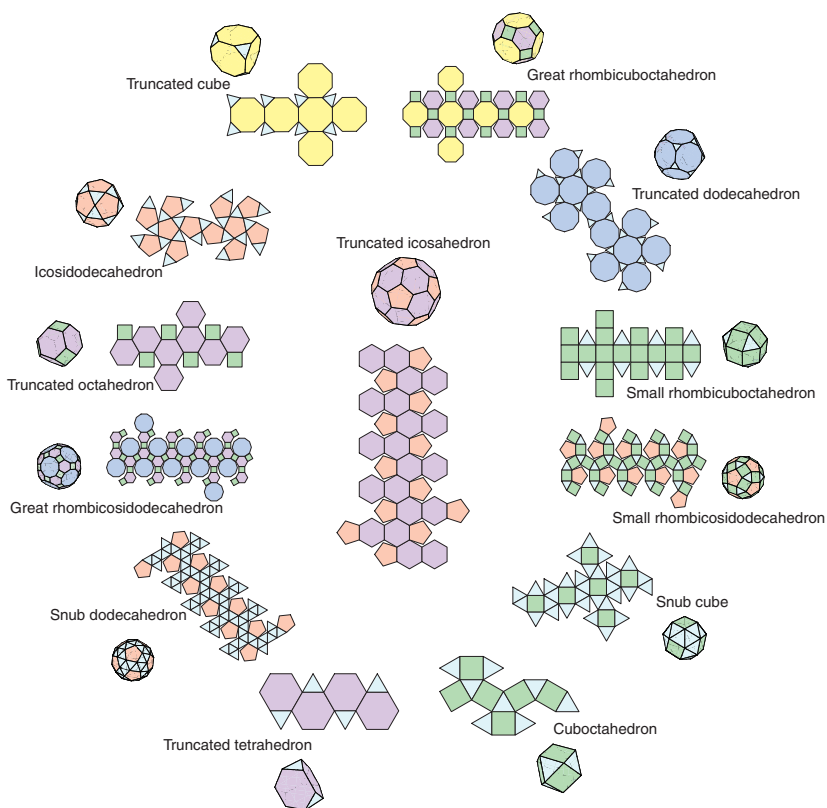


Figure 7.4. Nets for the 13 Archimedean solids. The nets here for the snub cube and for the truncated icosahedron are both different from those used by Dürer (Figures 7.2 and 7.3).

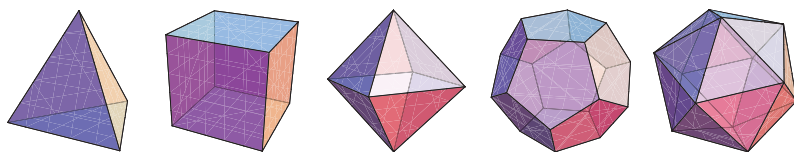


Figure 7.5. The five Platonic solids: tetrahedron, cube, octahedron, dodecahedron, icosahedron.

The most famous polyhedra are the five “regular” *Platonic solids* shown in Figure 7.5, known for at least 2,500 years. Despite the name “solid,” our definition of a polyhedron is the thin surface enclosing the volume, rather than the solid itself.

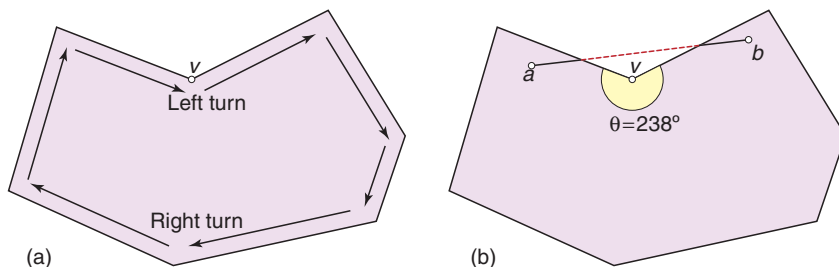


Figure 7.6. This polygon is not convex because (a) a clockwise traversal makes a left turn, or (b) not all the points on ab are in the polygon. Here v is a reflex vertex.

Box 7.1: Convexity

A shape S is *convex* if, for any two points a and b in S , all the points of the line segment ab connecting those points are also in S . As Figure 7.6(b) illustrates, if a polygon has a vertex v at which the internal angle θ is greater than 180° (a *reflex vertex*), then one can find two points a and b on either side of v such that the segment ab is partially exterior to the polygon. Conversely, if ab is wholly inside for every possible a and b , then there can be no dents. The same holds for 3D polyhedra as well.

Convexity plays an important role in mathematics. My college's library has twenty-two books with "convexity" in their title, including one named simply "Convexity"! We describe convexity from the point of view of angles on the boundary of a shape. But a little thought shows that the angles of a convex polygon being less than 180° is a consequence of the formal definition above.

Convex Polyhedron. The Platonic solids are all *convex polyhedra*, which means, essentially, they have no dents. In 2D, a *convex polygon* is a closed figure composed of straight edges joined at vertices, with the property that if you walked around the boundary clockwise, you would make only right turns at each vertex. A left turn would constitute a dent. See Figure 7.6(a). Another way to think about this is that the internal angle at every vertex of a convex polygon is less than 180° . A vertex at which the internal angle exceeds 180° is called a *reflex vertex* (which is more memorable than calling it a "concave vertex"). Box 7.1 gives a more formal definition of convexity.

In 3D, just as a convex polygon has no reflex vertices, a convex polyhedron has no reflex edges. Recall that the dihedral angle (which we encountered on p. 89) is the 3D angle at the join of two planes. For a convex polyhedron, the dihedral angle formed by the two faces meeting at each edge is less than 180° . For the cube, this dihedral angle is 90° at each convex edge (as we saw earlier

in Exercise 6.3[a]). The polyhedron we showed in Figure 5.15 is a nonconvex polyhedron with many reflex edges – for example, those ringing the elegant neck.

Net. Now that we know what a convex polyhedron is, we finally define what constitutes a net. A *net* is an unfolding of the surface of a polyhedron produced by cutting the polyhedron along some of its edges and flattening the surface to a single, nonoverlapping piece in the plane.

The key aspects of this definition are:

1. The net is planar.
2. It is a single piece.
3. It is the result of cutting polyhedron edges. This is called an *edge unfolding*.
4. It is non-self-overlapping in the plane, so that noncut points do not unfold on top of one another.

These four properties mean that we could cut out a net drawn on paper with scissors, crease along edges, and fold back to the 3D polyhedron. Our definition is intended to capture the standards Dürer set down by his example. His net for the truncated icosahedron (Figure 7.3) is just barely nonoverlapping, with some pentagons almost touching some hexagons, as is the net for the same polyhedron in Figure 7.4. It makes sense to permit boundaries to touch, because that would still allow the net to be cut out with scissors. What is forbidden is more substantial overlap, where cutting out the net would slice through the interior of some faces.

Exercise 7.1 (Understanding) *Melencolia Polyhedron.* Is the drawing of the visible faces of the polyhedron in Dürer's "Melencolia I," shown earlier in Figure 7.1, consistent with the interpretation that it is a cube with two diagonally opposite vertices cut off (*truncated*)?

7.3

The Open Problem

Now we can pose the unsolved question, which we take the liberty of calling "Dürer's Problem" even though there is no evidence that he recognized it as claim that needed proof:

OPEN PROBLEM: *Dürer's Problem*
Does every convex polyhedron have a net?

Despite almost 500 years of many people drawing nets for convex polyhedra, no one has come up against an example that has no net. On the other hand, there is no proof that every convex polyhedron does have a net, despite years of

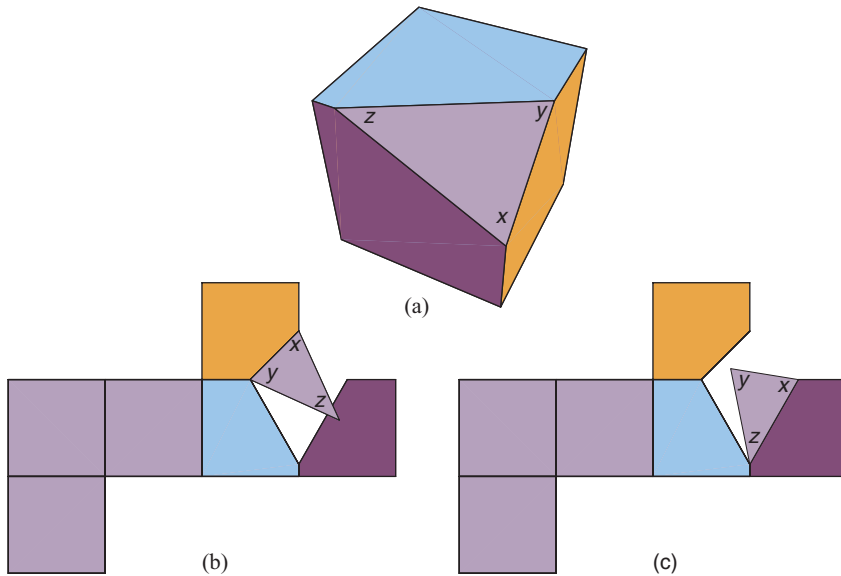


Figure 7.7. (a) Cube with corner truncated. (b) Overlapping unfolding. (c) A net: nonoverlapping.

effort since the problem was first formalized as a precise mathematical question in 1975.

You might think that it is obvious that a convex polyhedron has a net, because when the surface is cut open and flattened, it “spreads out,” and so there should be plenty of room in the plane to avoid overlap. This is certainly true for the regular and semiregular polyhedra that have been the focus of attention for hundreds of years. It might not even be clear how overlap can ever occur. Figure 7.7(b) shows an unfolding of a cube with one corner truncated that leads to overlap. But (c) of the figure shows that it is easy to avoid the overlap by cutting slightly differently to reposition the problematical face.

Still, it could be that a more complicated and irregular convex polyhedron, such as the egg-shaped object in Figure 7.8, may not have a net. There are an infinite number of convex polyhedra, so resolving Dürer’s problem requires a proof.

Julie DiBiase proved (when a college student) that every tetrahedron – that is, every convex polyhedron with exactly four vertices – has a net. But even this is not as obvious as it might seem. Consider the thin, nearly flat tetrahedron shown in Figure 7.9(a). One choice of cutting leads to overlap (b), although again it is easy to find other cutting choices, such as (c), that do produce nets for this tetrahedron.

Exercise 7.2 (Understanding) *Cube Net.* Can a cube be cut along edges and opened to a net in such a way that the cuts form one, single connected path,

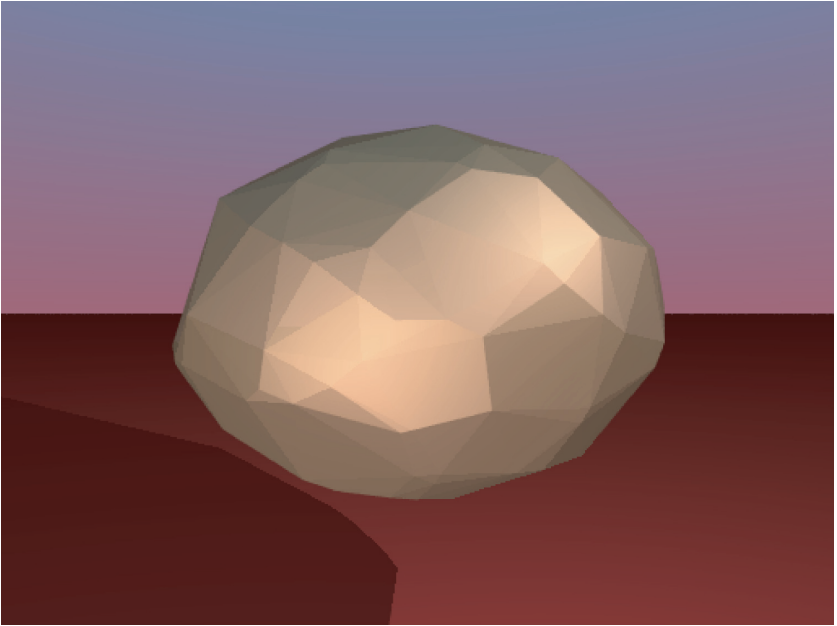


Figure 7.8. A convex polyhedron whose 100 vertices are randomly sprinkled on the surface of an ellipsoid.

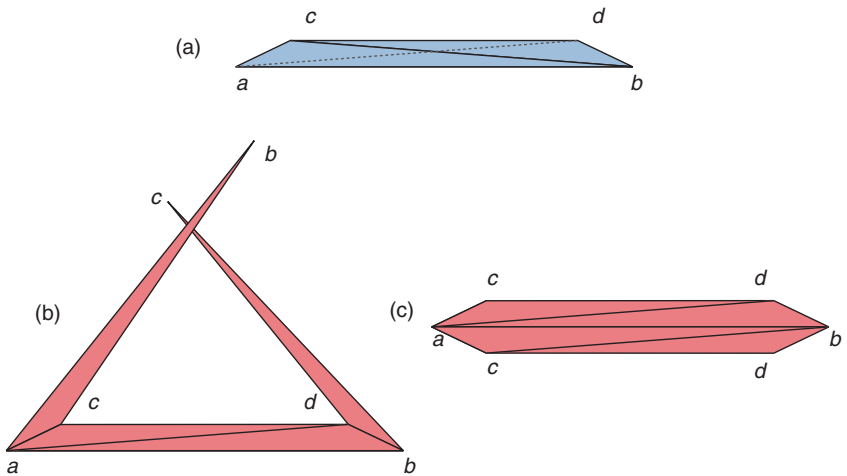


Figure 7.9. The four faces are colored blue on the outside and red on the inside. (a) A nearly flat tetrahedron. Edge ad is in back. (b) Overlapping unfolding from cutting the edges ab , bc , and cd . (c) A net obtained by cutting the edges ac , cd , and db .

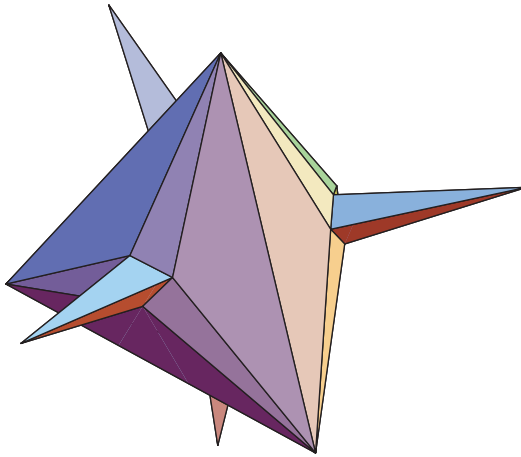


Figure 7.10. A nonconvex polyhedron of 36 triangle faces that has no net.

with no “branching”? In other words, could the cuts be achieved by one trail of a knife over the surface, without ever lifting the knife off the surface?

7.3.1 Nonconvex Polyhedra

I have emphasized that Dürer’s open problem is only unsolved for convex polyhedra. In the late 1990s, several groups of researchers independently discovered nonconvex polyhedra that have no net. One of these discoverers, Alexei Tarasov, was a high school student at the time (now he’s a professor). Figure 7.10 shows an especially elegant example, a “spiked tetrahedron,” found by Erik Demaine when he was 18 years old, in collaboration with several others (now Erik is also a professor). I will not give a proof here that this polyhedron has no net. But, as you can imagine, the proof must show that every possible edge unfolding leads to overlap, and there are many possible edge unfoldings.

Because any proof that a polyhedron has no net must grapple with “every possible edge unfolding,” we next take a look at some necessary properties edge unfoldings must possess.

7.4 Spanning Cut Tree

The *skeleton* of a polyhedron is the network of edges and vertices on the surface – effectively a wireframe view of the polyhedron without its faces. It is a portion of the skeleton that is cut to make a net. We call an edge of the polyhedron that is cut as part of an unfolding a *cut edge*. Figure 7.11 shows a set of seven cut edges that unfold a cube to the “Latin-cross” net. Five of the 12 edges of the cube skeleton are not cut in this unfolding. What we seek are a set of conditions on the collection of cut edges that must hold for the unfolding to be a net. We first

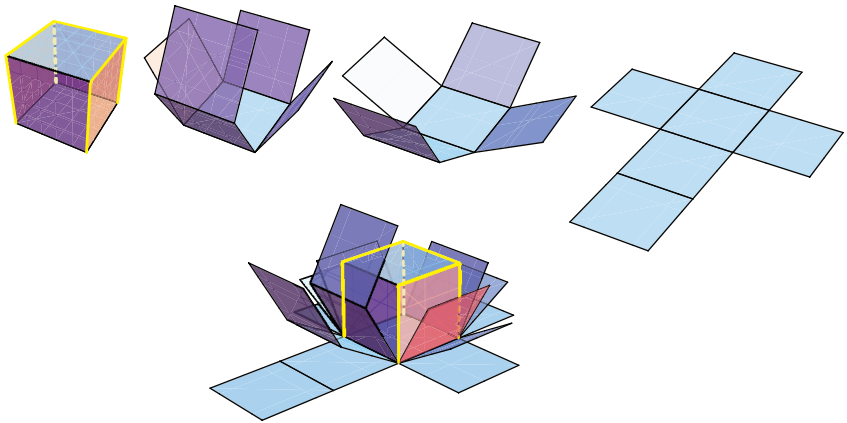


Figure 7.11. The Latin-cross unfolding of the cube. The cut edges are marked in yellow.

state the necessary conditions using technical jargon, after which the terms will be explained:

Lemma 7.1

The cut edges of an unfolding that produces a net for a polyhedron form a spanning tree of the skeleton of the polyhedron.

We need to explain the phrase “a spanning tree of the skeleton.”

A *tree* is a collection of edges that never forms a cycle or loop. It is called a tree because the edges “branch out.” The branches of a tree cannot dovetail and connect, for that would form a cycle of edges. A cycle must be avoided because it will surround some collection of faces and separate the surface into two or more pieces, as in Figure 7.12. Finally, a *spanning tree* is a tree that touches every vertex. So the eight vertices of a cube are each touched by the spanning cut tree in Figure 7.11. A more complex example is shown in Figure 7.13(b), which displays the cut edges that Dürer used to obtain his net of the snub cube. A close look shows that these edges indeed form a spanning tree of the polyhedron’s 24 vertices.

Exercise 7.3 (Understanding) *Octahedron Spanning Trees.* What is the fewest number of edges in a spanning tree of an octahedron (Figure 7.14)?

Now we can prove Lemma 7.1. If the cut edges do not form a collection of trees (a *forest*), then they contain at least one cycle, by the definition of “tree.” We’ve already seen that this leads to violation of the one-piece condition of the definition of a net (p. 106). The cut edges are connected by the boundary of the

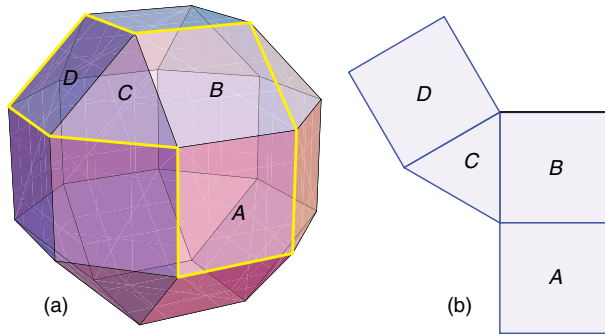


Figure 7.12. A cycle of cut edges separates out a piece of the surface. (The polyhedron here is the small rhombicuboctahedron.)

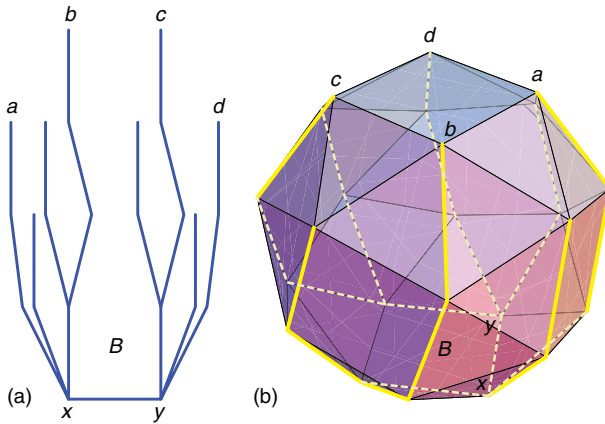


Figure 7.13. (a) A tree. (b) Dürer's spanning cut tree on the snub cube. The tree in (a) is structurally the same as the tree in (b). The base face B (on which the polyhedron rests) and several vertices are labeled in both.

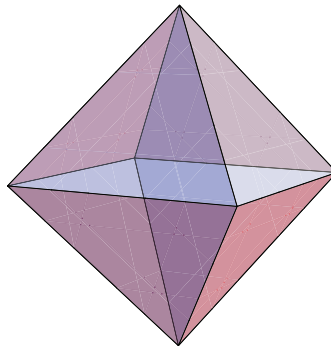


Figure 7.14. Octahedron to cut open. (Exercise 7.3)

net, so they must form a single tree, rather than a collection of disconnected trees.

Suppose the tree is not spanning. That means that there is some vertex v of the polyhedron not touched by any cut edge. But this means that v retains its 3D structure, and so cannot be flattened. This violates the planarity condition of the definition of a net. For the net to be planar, the cuts must span the vertices.

Thus we have established Lemma 7.1: The cut edges must form a spanning tree to have any hope of producing a net. So the spanning tree condition is necessary. Does this help us to solve Dürer's problem? The lemma effectively covers three of the four conditions on a net, but does not address the crucial fourth condition: The net should be nonoverlapping. And indeed the spanning tree condition is far from sufficient to guarantee a net. In fact, the cuts used to produce our overlapping examples (Figures 7.7 and 7.9) do form spanning trees.

To this day, no one has discovered necessary and sufficient conditions for a collection of cut edges to unfold to a net. Consequently, any potential counterexample to the conjecture that every convex polyhedron has a net must thwart every possible spanning tree. And there are a lot of spanning trees! For the egg-shaped polyhedron in Figure 7.8, there are more than 500,000 different spanning trees.

Exercise 7.4 (Challenge) *All Cube Nets.* Find all the different nets for a cube.

7.5

Some Polyhedra with Nets

The investigation of Dürer's Problem has as yet yielded no grand theorem, let alone a definitive answer. However, a few special classes of convex polyhedra are known to have nets. To describe these classes, we will use the concept of the convex hull. In 2D, suppose we mark a set of points in the plane by pounding in nails at each point, leaving a portion of the nail above the plane, and then stretch a rubber band around these nails. The *convex hull* of the points is the polygon determined by the shape of the stretched rubber band. Perhaps you can see that it is always a convex polygon – no concavities. In 3D, we have to imagine a set of points fixed in space. The convex hull of the points is the the convex polyhedron determined by wrapping a set of points in 3D as tightly as possible with plastic wrap. I “constructed” the polyhedron in Figure 7.8 by computing (via a complex computer program) the convex hull of random points on an ellipsoid.

Now with this notion, we can define the classes of polyhedra for which nets are known. Let B be a convex polygon in a plane. Make a second copy of B directly vertically above B , lying in a parallel plane; call that copy A . The convex hull of the vertices of A and B is a *right prism*, “right” because all the vertical side faces (the *lateral faces*) are rectangles with right angles. If we permit A to shift around, remaining in a plane parallel to the base face B , the result is still a prism, but now an *oblique prism*, whose lateral faces are parallelograms. See Figure 7.15(a,b). Both of these classes are subclasses of a more general shape, a

prismoid, whose side faces are trapezoids. Now A and B are no longer congruent copies of one another, but each edge of A is parallel to one of B . I proved that every prismoid has a net, as illustrated in Figure 7.15(c): All lateral edges are cut, and all but one edge of A is cut, and the shape unfolds like the petals of a flower exposed to the sun. The only delicate part of the proof is deciding which edge of A not to cut. Not all choices always lead to non-overlap (although they do in this example).

A related class of shapes includes the pyramids. A *pyramid* is the convex hull of a base convex polygon B and a single point a (the apex) anywhere above B . The natural *petal unfolding* determined by cutting every edge incident to a leads to a net; see Figure 7.16(a). A *dome* is a generalization of a pyramid with the

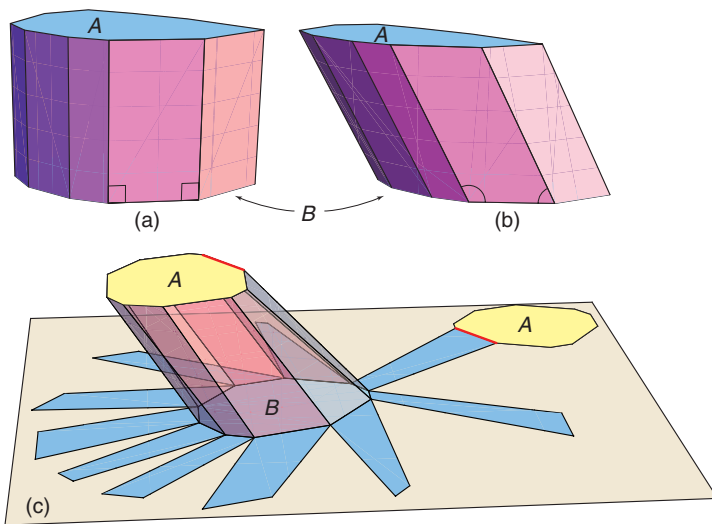


Figure 7.15. (a) A right prism. (b) An oblique prism. (c) Unfolding of a prismoid to a net.

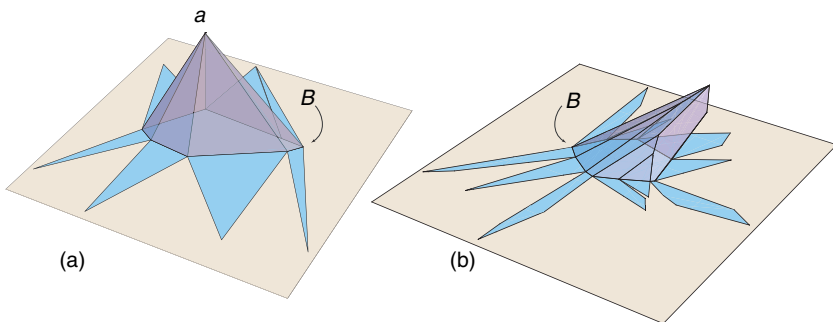


Figure 7.16. (a) A net for a pyramid. (b) A net for a dome.

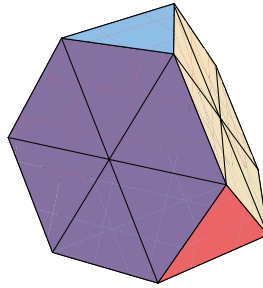


Figure 7.17. A deltahedron formed by partitioning the faces of a truncated tetrahedron (one of the Archimedean solids in Figure 7.4) into equilateral triangles.

property that every face shares an edge with the base convex polygon B , just as in a pyramid, but now there is no apex vertex that all faces touch; see Figure 7.16(b). Again the petal unfolding leads to a net, a theorem for which there are now three different proofs.

One more class, “higher-order deltahedra,” was proved to always have a net by Daniel Bezdek, as part of his award-winning ninth-grade Canadian Science Fair project. These are polyhedra whose surface is composed entirely of equilateral triangles, including faces (e.g., hexagons) partitioned into several triangles. See Figure 7.17 for an example.

These few classes – prismoids (including oblique and right prisms), domes (including pyramids), and deltahedra – are the only infinite classes of convex polyhedra for which it has been proved that they each have a net.

A related class of polyhedra for which there is as yet no proof that nets always exist is the prismatoids. A *prismatoid* is the convex hull of two different convex polygons A and B lying in parallel planes. It is very similar to a prism or a prismoid (and its name is similarly confusing!), but in general, the lateral faces are not quadrilaterals, but rather triangles, as shown in Figure 7.18. Although this is

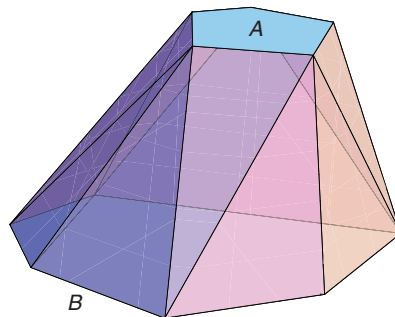


Figure 7.18. A prismatoid, the convex hull of two convex polygons lying in parallel planes.

only a tiny piece of Dürer's open problem, it is at this writing unsolved. I invite the enterprising reader to tackle this "open subproblem."

OPEN SUBPROBLEM: *Edge-Unfolding Prisms*
Does every prismatoid have a net?

Exercise 7.5 (Practice) *Unfolding Right Prisms.* The right prism in Figure 7.15(a) is a prismoid and all prismoids have petal unfoldings to nets. Can you think of a different way to unfold a right prism? Your method should work for any right prism.

7.6

Above & Beyond

If we don't restrict ourselves to only cutting along edges of the polyhedron (Condition [3] in our definition of a net), we can find some beautiful theorems. I will describe one of these to conclude this chapter.

A *general unfolding* lets us make arbitrary cuts to produce the unfolding. The cuts can run anywhere through the polyhedron faces. Our goal is the same: find a collection of cuts that unfold the polyhedron to a single, planar, nonoverlapping piece. Let us call this a *general net*. There are now several different proofs that every convex polyhedron has a general net. All of them depend on the concept of a shortest path on the surface between two points a and b . As its name suggests, a *shortest path* is the minimum-length route to travel from a to b on the surface—the optimal path for an ant to walk between the two points.

Two Properties of Shortest Paths. Shortest paths have many mathematical properties, two of which are important for this result. First, a shortest path never passes right through a polyhedron vertex: It is always shorter to go around a vertex than through it, just as it is always shorter to avoid the peak of a hill while hiking. Second, when a shortest path crosses a polyhedron edge, it does so in such a way that the planar unfolding of the two faces sharing the edge straightens the path: A shortest path unfolds straight across every edge. See Figure 7.19.

Exercise 7.6 (Practice) Shortest Paths on Cube. Let x be the point in the middle of the top face of a cube, and a , b , and c points on a vertical edge as illustrated in Figure 7.20: a and c are the edge endpoints, and b is the midpoint of the edge. (a) To which of the three points a , b , c (if any) is the shortest path on the surface from x unique, so that there is exactly one shortest path from x to that point? (b) Find all the shortest paths from x to a , b , c , including those paths tied for shortest.

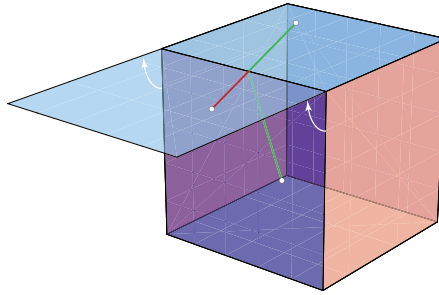


Figure 7.19. A shortest path (green) crossing an edge of a polygon unfolds to a straight segment (red).

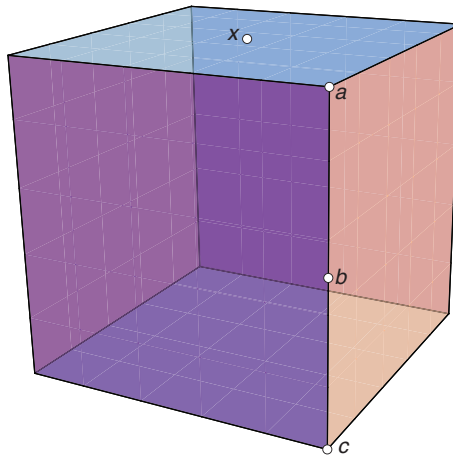


Figure 7.20. Find the shortest surface paths from x to a , b , and c . (Exercise 7.6)

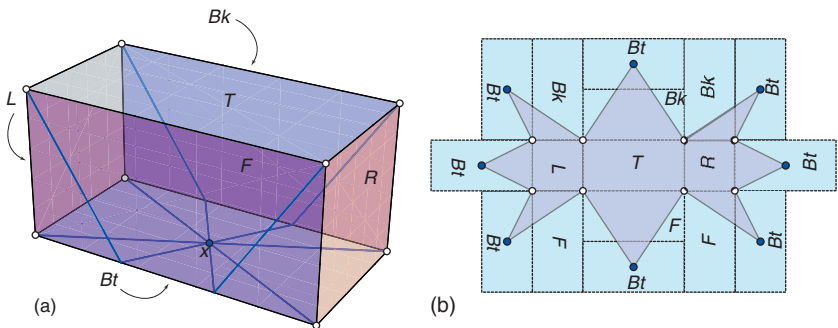


Figure 7.21. (a) $2 \times 1 \times 1$ box. Box faces are labeled: Bt, F, T, R, L, Bk for Bottom, Front, Top, Left, Right, and Back respectively. (b) Star unfolding with respect to x .

One unfolding that always produces a general net is the *star unfolding*. This concept was introduced by the great Russian mathematician Alexandr Alexandrov in 1948, but only proved to be nonoverlapping in the 1990s.

It works like this. Pick any point x on the surface of the polyhedron not at a vertex, where the shortest path from x to each vertex is unique. Now draw the shortest path from x to each vertex of the polyhedron in turn. Figure 7.21(a) illustrates these paths for a rectangular box, with x in the middle of the bottom face. The collection of these paths satisfies Lemma 7.1: They form a tree that spans the vertices. The star unfolding is produced by cutting all the shortest paths and unfolding, as shown in (b) of the figure. Each shortest path unfolds to a straight line segment, as required by the second property of shortest paths.

Although nonoverlap looks almost obvious in this symmetric example, it is less obvious, although now proved, for more generic convex polyhedra. Figure 7.22 shows a more typical star unfolding. If you cut the unfolding in (a) out of paper, and tape together the cuts between adjacent copies of x (open circles), it will form the particular convex polyhedron of 11 vertices shown in (b) of the figure (in a highly nonobvious way!). Moreover, it uniquely folds to this polyhedron, by a theorem of Alexandrov that we will explore in Chapter 9.

Exercise 7.7 (Practice) Refolding a Star Unfolding. This is an exercise for readers with both patience and some manual dexterity! Cut out the unfolding in Figure 7.22(a) (you can download it from this book's website), and tape together adjacent cuts from x to each numbered vertex, and verify that you do obtain the polyhedron with vertices labeled as in (b).

8

Unfolding Orthogonal Polyhedra

We just learned that every convex polyhedron has a general net, and we saw an example in Figure 7.10 of a nonconvex polyhedron that does not have a net (cutting along edges only). This naturally raises the question: Does every nonconvex polyhedron have a general net? This is yet another unsolved problem. As with Dürer's problem, no counterexample is known, but there is not much evidence that the answer to the general unfolding question is YES. If every polyhedron does have a general net, it would certainly make a stunning theorem!

Because this problem seems difficult, researchers have focused on a special class of nonconvex polyhedra known as orthogonal polyhedra, where "orthogonal" means "at right angles."

8.1 Orthogonal Polyhedra

You can think of orthogonal polyhedra as those you could build out of Lego blocks. An *orthogonal polyhedron* is a polyhedron where each edge is parallel to one of the axes of a standard right-angled xyz -coordinate system. If all edges are parallel to an axis, then all faces are parallel to a *coordinate plane*: either xy (horizontal) or xz (vertical, front or back) or yz (vertical, side, left or right). Any pair of faces of an orthogonal polyhedron that share an edge either lie in the same plane (they are *coplanar*) or meet at right angles to each other, that is, orthogonally.

The general unfolding question was answered in 2007 for orthogonal polyhedra: YES, all orthogonal polyhedra have a general net. The proof is rather complicated, so I will only describe it in general terms in the "Above & Beyond"

section. In the rest of the chapter, we look at a more approachable special case, and again challenge the enterprising reader with an unsolved problem.

Box 8.1: Arrays

An *array* is a rectangular arrangement of numbers, like that shown in Figure 8.1(a). The term “array” is used in computer languages, whereas “matrix” is used in mathematics to represent an array of numbers that often is viewed as a unit and manipulated in various ways. In this chapter, we use arrays solely as a convenient representation of a rectangular collection of terrain heights. Arrays have *rows* and *columns* with the natural interpretations. One number in the array occupies an array *cell*.

8.2 Orthogonal Terrains

The special class of polyhedra we examine is known as *orthogonal terrains*. Suppose we have a rectangular grid divided into squares, and on each square we place a block of some height. Each square is 1 unit on a side, a *unit square*. This results in a polyhedral representation of a Digital Elevation Model (DEM). The heights can be arranged in a rectangular array of numbers. (See Box 8.1 on arrays.) If an array cell has height h , then a $1 \times 1 \times h$ block or tower is placed on the corresponding unit square in the rectangular base. An example is shown in Figure 8.1.

DEMs are typically obtained from satellite radar data. The U.S. Geological Survey offers the National Elevation Data set covering the whole United States in a grid whose square cells cover 30×30 meters. So the entire DEM is an array of roughly $150,000 \times 100,000$ cells, one height per cell – 60 GB to download! It is a curious thought that a polyhedral version of this model has a net, for that is the claim:

Theorem 8.1

Every orthogonal terrain polyhedron has a net.

So here is a class of nonconvex polyhedra that do have an edge unfolding that avoids overlap. We should make clear, however, that to find this unfolding we need the freedom to cut along any seam between adjacent towers, even if the faces meeting there are coplanar, and so the dihedral angle (see p. 40) is 180° .

Our explanation of the steps of the proof of this theorem will rely on a series of illustrations. We will imagine that we are standing in the $-y$ halfspace in the

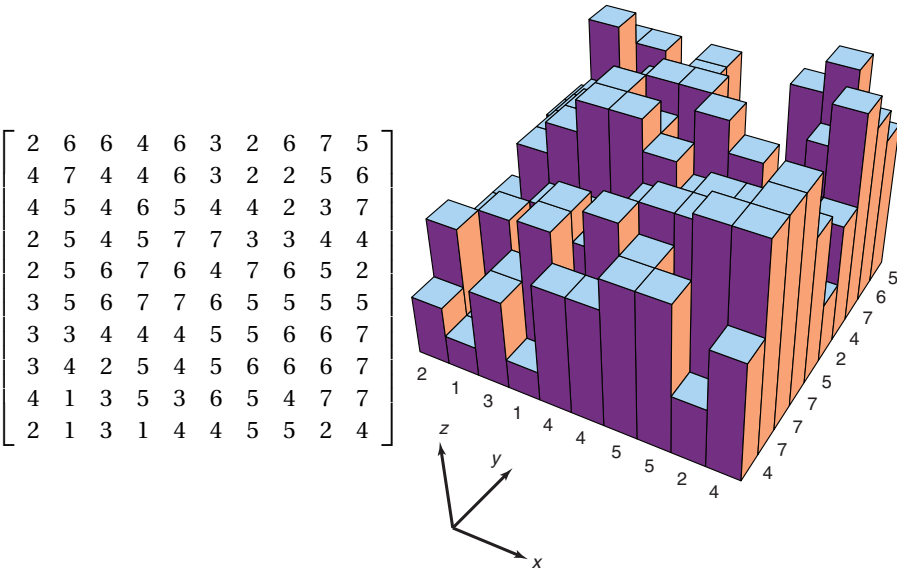
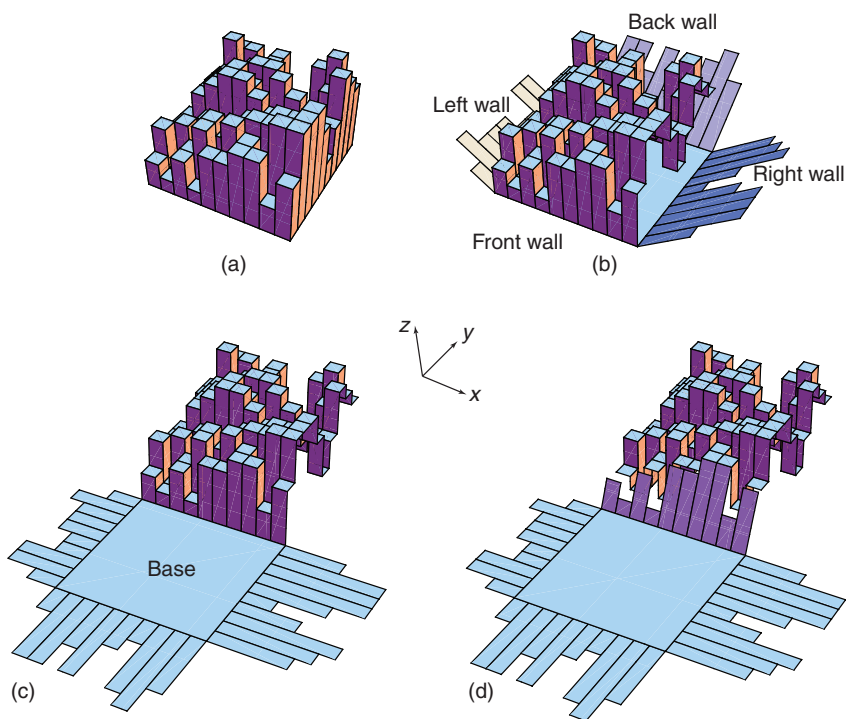
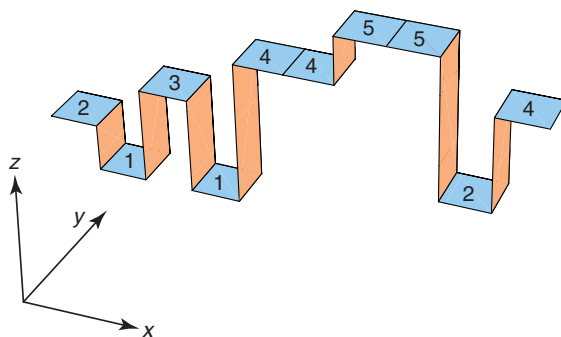


Figure 8.1. An array of heights, and the corresponding orthogonal terrain, with the frontmost and rightmost tower heights indicated.

coordinate system displayed in Figures 8.1 and 8.2. The “front” of the polyhedron is the collection of faces nearest us, forming the “front wall.” The “back” is the wall farthest away, and the “left” and “right” walls are to the left and right sides as we face the front.

1. To find the edge unfolding, we start by cutting along the top edges of the right, left, and back walls as in Figure 8.2(b), and unfold these three walls to the base xy -plane as units, leaving them attached to the rectangular base.
2. Next, we swing the base and its three dropped wall attachments 180° about the line at the bottom of the front wall, as in Figure 8.2(c).
3. Now we cut all but the top edge of the tallest front tower. If there is more than one tallest tower, we choose any one to remain attached. (In our example, two towers are tied at height $h = 5$.) This lets the front wall swing out about this uncut edge, as in Figure 8.2(d), ultimately placing the base and all four attached walls in the same horizontal plane at $z = h$. You can see that so far none of these unfoldings cause overlap.
4. The fourth step is the heart of the process. At each unit on the y -axis, there is a left-to-right sequence of faces running parallel to the x -axis on the top of the polyhedron, connecting all those towers at the same y coordinate; see Figure 8.3. We call this an x -strip. An x -strip consists of horizontal 1×1 top faces, one per tower at that y coordinate, connected by vertical left and right faces up and down each tower side, forming a snake-like, sinuous strip.

Figure 8.2. Unfolding the walls and base of P .Figure 8.3. An x -strip, for the frontmost towers in Figure 8.1.

We cut along both sides of each x -strip, except for one edge on either side, whose specification we defer to the fifth step below. These cuts permit us to unfold each x -strip as a unit, into a (long) rectangle stretching in the x -direction. Figure 8.4(a \rightarrow b \rightarrow c) illustrates the process. For example, the first x -strip, running along the edge of the front face, eventually stretches out

to a 24×1 rectangle (see ahead to Figure 8.5). We unfold all x -strips this way simultaneously.

Exercise 8.1 (Practice) x -strip length. Verify the claim that the strip shown in Figure 8.3 unfolds to a 24×1 rectangle.

5. The trickiest part of the unfolding is choosing “bridge” rectangles that connect the x -strips so that we can avoid overlap. For each pair of adjacent x -strips, we look at the height differences between the towers in the front and back of the two strips. Then we choose the *bridge* between them to be the connecting vertical rectangle with the greatest height difference. If several rectangles tie for tallest, any one may serve as the bridge. The bridge rectangle may be a front tower face or a back tower face.

If we look at the first two strips of our example in Figure 8.1, we see that the frontmost strip has a tower of height 2 with a tower of height 7 directly behind it in the second strip. The front rectangle of height 5 connecting the two strips (marked in Figure 8.4) is selected as the bridge between these two strips.

Now, when the second strip is laid out in the plane containing the base, it is separated from the first strip in the y -direction by the height of the bridge and aligned horizontally so that the bridge connects the strips. As mentioned earlier, all edges along the front edge of the second strip are cut except for the edge of the bridge rectangle. Figure 8.4 shows how the stretching of the x -strips horizontally and the bridge separation are achieved simultaneously. An overhead view of the final net is shown in Figure 8.5(a).

One more detail: If two adjacent strips are identical, they are unfolded next to one another without a separating gap. We can imagine there is an “empty bridge” selected anywhere between them.

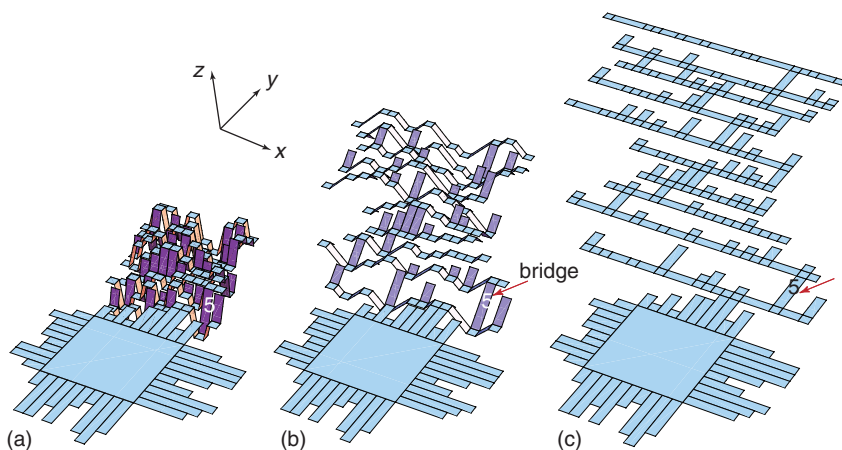


Figure 8.4. Unfolding the top faces of P into x -strips connected by y -bridges.

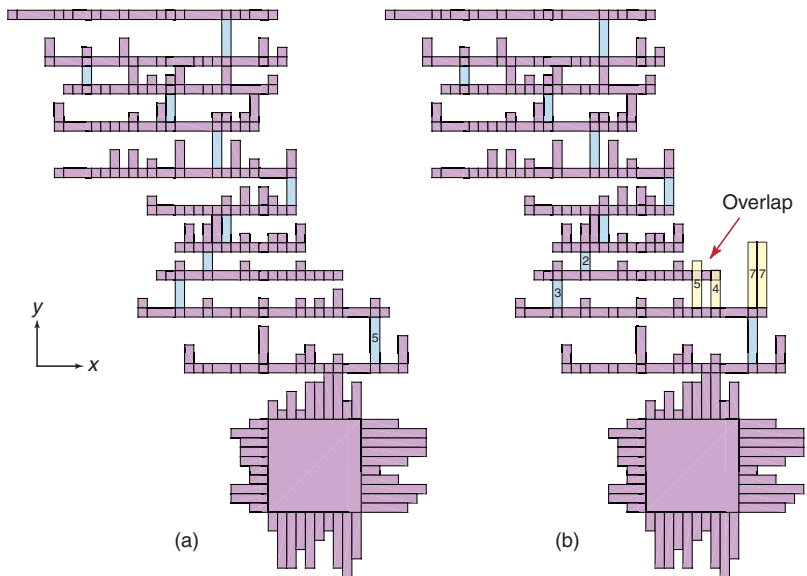


Figure 8.5. (a) Overhead view of final net for the polyhedron in Figure 8.1(b). Bridge rectangles are shaded blue. (b) Overlapping unfolding corresponding to Figure 8.6.

Choosing the tallest rectangles as bridge rectangles avoids overlap because all other vertical faces between the strips are shorter or at least no longer than the bridge, and so when they are attached to one strip they do not overlap with the layout of the next strip. In the first two strips of Figure 8.5(a), the bridge of height 5 pushes the layout of the second strip in the y -direction far enough to avoid overlap with all the shorter rectangles attached to the first strip. As this bridge displacement is repeated for each successive strip, overlap is avoided.

There is some boundary touching when several rectangles are tied with the bridge for tallest. But this is still a net: You could cut the shape out of paper and fold exactly to the polyhedron by reversing the unfolding procedure.

This completes the proof of Theorem 8.1.

Exercise 8.2 (Practice) 2×2 Terrain. Draw out the unfolding of this 2×2 terrain, following the steps in the proof of Theorem 8.1.

$$\begin{bmatrix} 3 & 4 \\ 1 & 2 \end{bmatrix}$$

Exercise 8.3 (Understanding) $n \times n \times 1$ Box. Suppose every cell of a $n \times n$ terrain has height 1:

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ \dots & & & & \dots \\ 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix}$$

Following the unfolding procedure as described, draw out the unfolding for arbitrary n .

8.3 Grid Unfoldings

Orthogonal terrains are a rare class of nonconvex polyhedra for which there is a procedure for edge unfolding to a net. Even minor generalizations of the class of polyhedra bring us to the unknown. For example, suppose we simply remove a few towers so that the base of the polyhedron is no longer rectangular. Figure 8.6 shows the polyhedron of Figure 8.1 with four towers removed, so that the base has a dent in its right side. Following the proof steps of Theorem 8.1 leads to the overlapping unfolding shown in Figure 8.5(b). What went wrong?

The second and third x -strips are separated by a bridge of height 3, but the backs of the four towers in the second row adjacent to the gaps have heights 5, 4, 7, and 7. These rectangle heights are too tall to be accommodated in the gap created by the bridge, and overlap results.

No one has found a procedure that guarantees an edge unfolding to a net for the class of orthogonal polyhedra consisting of vertical towers on an arbitrary base, a class called *Manhattan towers* in the literature. However, there is a procedure that produces a general net (p. 115) by using cuts through faces rather than solely along edges.

A general net permits arbitrary cuts as long as the final unfolding is a net. For orthogonal polyhedra, however, it is natural that the cuts be parallel to the axes rather than completely arbitrary; it is also natural to try to minimize the number of cuts inside faces. This leads to an intermediate class between edge unfoldings

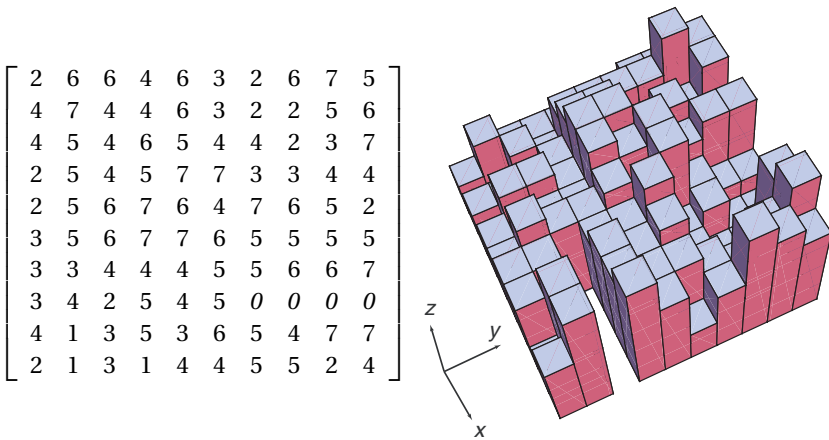


Figure 8.6. An array of heights, with four cells of zero height. The corresponding orthogonal polyhedron has a gap in the third row.

and general unfoldings, known as *grid unfoldings*. Imagine each face partitioned into a finer grid, with grid lines parallel to the axes, with all cuts along the grid lines. For Manhattan towers, there is a procedure that produces a general net if every face is first refined by a 5×5 grid, and all cuts are along these grid lines. This procedure will produce a general unfolding of the polyhedron in Figure 8.6, although it would be complicated to work out.

We can think of the net for orthogonal terrains guaranteed by Theorem 8.1 as being cut along a 1×1 grid, that is, the faces are not divided up at all. So here is another challenge to the reader. Find a procedure for unfolding Manhattan tower polyhedra to a general net, with cuts following a grid coarser than 5×5 . Because no counterexample is known, perhaps a 1×1 grid is possible, a true edge unfolding.

OPEN SUBPROBLEM: *Unfolding Manhattan Towers*
Does every Manhattan towers polyhedron have an edge unfolding to a net?

8.4 Above & Beyond

We opened this chapter by mentioning that, although unfolding nonconvex polyhedra to a general net is unsolved, it is now established that all orthogonal polyhedra have a general net. The procedure that establishes this produces

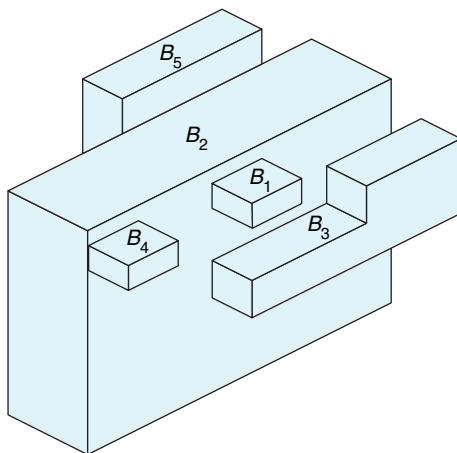


Figure 8.7. An orthogonal polyhedron, analyzed in Figure 8.8.

nets similar to grid unfoldings, with all cuts parallel to the axes. Unlike the situation with a grid unfolding, however, there is no standard-size grid that we know will always suffice. Instead each face might need be partitioned into a very fine grid, with the consequence that the unfolded net will have extremely thin sections. Let's look at how this works on the small example in Figures 8.7 and 8.8.

The orthogonal polyhedron consists of five “blocks” labeled B_1, B_2, B_3, B_4 , and B_5 in the figure. The unfolding starts at the point of B_1 marked s Figure 8.8 (s for “start”) and ends immediately to the left of that, at the point t (for “termination”). The unfolding “peels” the surface in a thin strip that winds around each block in a helical fashion, jumping from one block to another at various stages. The peels cover the top, bottom, left, and right faces; the front and back face are handled differently. Let us track the peel in the figure until the pattern becomes clear.

At s the yellow strip heads off to the right, spiraling around block B_1 in a clockwise direction until it jumps to B_2 , the large, central block. Note that the yellow spiral leaves a gap for the purple spiral to (much later) return, interleaving as it spirals counterclockwise around B_1 and ends at t . The yellow spiral only tracks along the top of B_2 for a short distance before it descends to block B_3 . Once it reaches the front face of B_3 , yellow turns to green and the strip climbs back up to the top of B_2 , and makes its way leftward over to B_4 .

After spiraling around B_4 , the strip (now yellow again) makes one complete cycle around B_2 and then climbs up to the backmost block B_5 . It spirals around B_5 , turns around on its back face, and then threads back around B_5 (now green), over to B_4 again (purple), threading the gap left in the earlier encounter with B_4 , rises to B_2 (now tan), revisits the front block B_3 , then back up to B_2 (now purple), and finally around B_1 again to terminate at t , just left of s .

The net itself is almost impossible to illustrate. Just a small piece of its beginning is shown in (b) of the figure. All the jags and turns of the spiraling strip are carefully chosen so that, when laid out in the plane, the strip unfolds to a rising and falling staircase-like shape. The strip always stretches out left to right; it never curls around leftward after taking a rightward step. The staircase nature of the strip unfolding guarantees that it never overlaps itself. After the top, bottom, left, and right faces of the polyhedron are stripped away, the front and back faces are hung in pieces like laundry from a clothesline, with the marked arrows pointing down. The left-to-right nature of the staircase ensures that there is always room for this hanging. The result is an extremely long and thin peeling of the entire surface into a single piece, a general net.

The example in Figure 8.7 hardly needs such a complicated unfolding! But the point is that the procedure works for any orthogonal polyhedron, no matter how complex.

Finding a simpler net – for example, one that cuts along a coarser, fixed-size grid – remains unsolved.

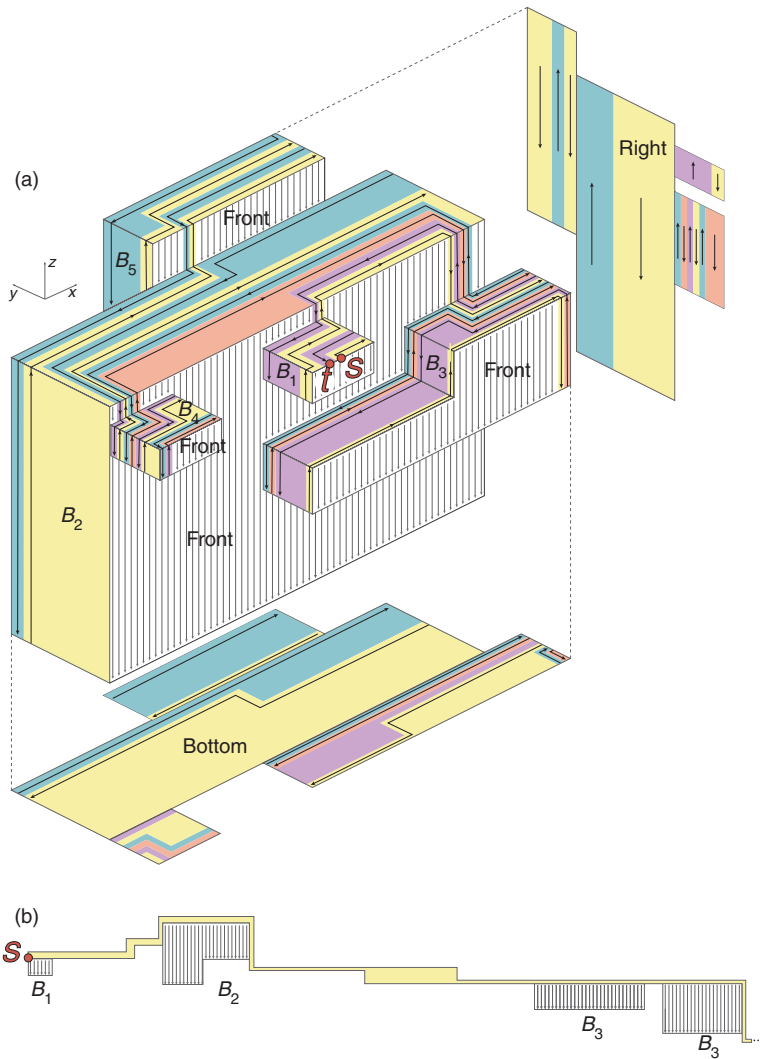


Figure 8.8. (a) Five-block example from Figure 8.7. The hidden bottom and right faces are projected out to mirrors so the paths may be tracked. The strip changes color as it alters direction of spiraling. (b) Start of the unfolding (not to same scale), with front face pieces (hashed with arrows) labeled.

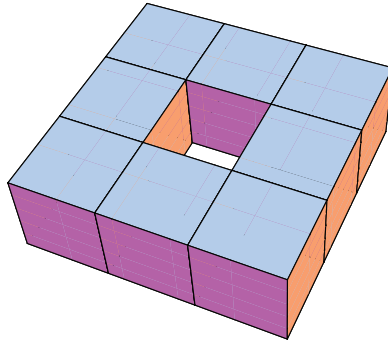


Figure 8.9. An orthogonal torus to edge-unfold. (Exercise 8.4)

Exercise 8.4 (Challenge) *Orthogonal Torus.* Find an edge unfolding of the orthogonal torus shown in Figure 8.9, composed of eight cubes glued together.

Finally, we conclude with the counterpart of Dürer's unsolved problem with which we opened the chapter, another unsolved problem of more recent vintage:

OPEN PROBLEM: *General Unfolding*

Does every nonconvex polyhedron have a general net?

9

Folding Polygons to Convex Polyhedra

The previous two chapters focused on unfolding the surface of a polyhedron to a net. It may seem that the reverse process – folding – could harbor no more secrets than unfolding. But there is a surprisingly rich structure here, and, as usual, many unresolved questions.

The situation is this. Someone presents you with a polygon cut out of paper, which may or may not be a net for a polyhedron, either an edge-unfolding net or a general net. Your task is to fold it to a polyhedron if possible. Because so little is known about nonconvex polyhedra, we concentrate almost exclusively (until the Above & Beyond section) on convex polyhedra (see p. 105 for the definition), where some deep theorems have been obtained and interesting questions raised. We will see that the richness of folding stems from that surprising fact that a net can often refold to many different different polyhedra, not just the one from which it originally derived.

For those familiar with origami polyhedra folding, that is not what is meant here. Let us stipulate that to *fold* a polygon to a polyhedron means to make creases that allow the polygon be folded to form the surface of a 3D polyhedron, without any wrapping overlap on the surface and without leaving any gaps. Origami foldings of, for example, a cube have extensive overlap. We want a folding to be the exact inverse of the unfoldings considered in previous chapters. A *polygon* is any planar (nonoverlapping) shape whose sides are all straight segments. Thus a polygon looks like a net. But it is only a net if it is the unfolding of a polyhedron.

The two examples in Figure 9.1 show that an equilateral triangle is a net for a regular tetrahedron, and that the Latin cross is a net for the cube, as we

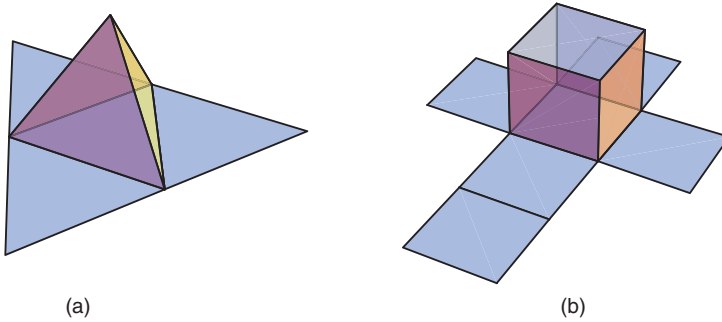


Figure 9.1. (a) Folding an equilateral triangle to a regular tetrahedron. (b) Folding the Latin cross to a cube.

saw earlier in Figure 7.11. Both are nets because in the reverse cut-open-and-unfold view, the cuts in both these examples are along polyhedron edges. The reader may be surprised to learn that the equilateral triangle can fold to an infinite variety of different convex polyhedra, and that the Latin cross can fold to precisely 23 different convex polyhedra! My goal in this chapter is to explain these two claims.

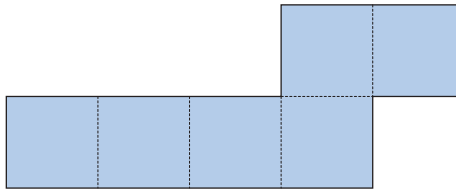


Figure 9.2. Can this polygon fold to a cube? (Exercise 9.1)

Exercise 9.1 (Practice) *Folding to Cube?* Can the polygon shown in Figure 9.2 fold to a cube by creasing along the dashed lines?

Exercise 9.2 (Understanding) *Bow Tie.* Can the “bow-tie” polygon in Figure 9.3 fold to a convex polyhedron? If so, to which polyhedron?

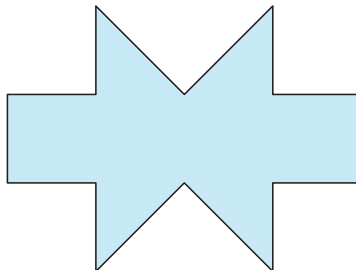


Figure 9.3. A polygon to fold. (Exercise 9.2)

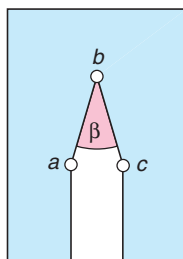


Figure 9.4. An unfoldable polygon P_u .

9.1 Questions

Consider again the situation where you are given a polygon and asked to fold it to a convex polyhedron if possible. A natural first question is: Is it always possible? In other words: Does every polygon fold to some convex polyhedron? The answer is no, and the proof is simple enough to present in its entirety.

The question can be answered by a single counterexample of an “unfoldable” polygon. We now show that the one in Figure 9.4 – call it P_u – indeed cannot be folded.

The key observation is that the total angle surrounding any true vertex v (corner) of a convex polyhedron is less than 360° . To use the technical language introduced earlier, the sum of the face angles *incident to* v is less than 360° . This result is one reason why a sixth Platonic solid does not exist: Gluing together three equilateral triangles to a vertex produces the tetrahedron, four yields the octahedron, five the icosahedron, but six times 60° equals 360° makes a flat region, not a true vertex. Another way to phrase this constraint – more useful for our purposes – is that the total angle surrounding any point p on a convex polyhedron (vertex or not) is $\leq 360^\circ$. When the angle is exactly 360° , p is not a vertex. This constraint does not hold, incidentally, for nonconvex polyhedra, for which there is no limit to the total face angle that might surround a surface point.

The consequence of the angle constraint is that when we *glue* the perimeter of any polygon to itself to form the folding (actually taping is more physically realistic, but we’ll use the word “glue”), we can never glue more than 360° around any one point, for otherwise, the polyhedron would not be convex.

Recall our earlier classification of the vertices of a polygon as either *convex*, having internal angle $< 180^\circ$, or *reflex*, having internal angle $> 180^\circ$ (p. 105). The polygon P_u in Figure 9.4 has three consecutive reflex vertices (a, b, c), with a very small complementary exterior angle β at b . All other vertices are convex, with interior angles 90° , which is larger than β .

Now we imagine how we might glue up the perimeter in the vicinity of the problematic vertex b . We have only two options. Either we “zip” together edges ba and bc , or some other point or points of the perimeter glue into the exterior β -gap at b . The first possibility forces a to glue to c , exceeding 360° there, violating the angle constraint. So this option is ruled out. The second possibility cannot occur with P_u , because no perimeter point has a sufficiently small internal angle to fit inside β at b . This constraint rules out the second possibility, and shows that P_u cannot fold to any convex polyhedron.

So now we know that sometimes polygons can, and sometimes they cannot, fold to a convex polyhedron. The natural next question is: Which polygons can fold to a polyhedron? Already here we reach the frontier of knowledge: There is as yet no satisfactory answer to this question. In particular, there is no characterization of which polygons fold and which do not, except in certain special cases, which we will explore below. Nevertheless, there is now an algorithm, implemented in publicly available software, that will take any specific polygon as input and tell you whether it can fold, and if so, give some information about the convex polyhedra to which it can fold. Before we can explain this somewhat mysterious statement, we turn to the powerful theorem that sits at the heart of this research.

9.2 Alexandrov's Theorem

Alexandrov's theorem is a deep and beautiful result, whose statement, fortunately, is elementary and so easily grasped. However, his 1948 proof is quite intricate and challenging even for professional mathematicians, and remains a subject of continuing study today. In fact, exactly 60 years later, a new proof was discovered, still intricate and challenging but with one advantage over the original, which we will explain in Section 9.5.

I will simplify and specialize his theorem to our needs. First, let us define an *Alexandrov gluing* of a polygon to be just what we need for a folding to a convex polyhedron. There are three conditions that must be satisfied for a gluing to be Alexandrov:

- (a) The gluing must entirely consume the perimeter of the polygon with sections that match up: Every point p of the perimeter must be matched with one or more points of the perimeter. Here we allow isolated points to match with themselves, as we did in Figure 9.4 when considering “zipping” in the neighborhood of b .
- (b) The gluing creates no more than 360° angle at any point. This constraint is our surface-angle condition for convex polyhedra.
- (c) The gluing should result in a *topological sphere*, that is, a surface that could be deformed to a sphere. In other words, not a donut-like torus (Figure 8.9),

or a shape with several holes, but rather what amounts to a lumpy, closed bag.

This third condition is difficult to state precisely without introducing technical language from topology. Regardless, I hope it is clear that if a gluing has any hope of producing a convex polyhedron, it must be an Alexandrov gluing, for the three conditions just specify what is obviously necessary – no gaps, the 360° condition, and producing a roughly spherical shape.

Exercise 9.3 (Challenge) *Alexandrov Gluing.* Can you find a way to glue the boundary of the polygon in Figure 9.2 to itself so that it satisfies the three conditions above, and therefore is an Alexandrov gluing?

Here is Alexandrov’s 1948 theorem:

Theorem 9.1 (Alexandrov)

Any Alexandrov gluing corresponds to a unique convex polyhedron (where a doubly covered polygon is considered a polyhedron).

Let us ignore the parenthetical caveat for a moment to emphasize what this theorem is saying: The obvious necessary conditions for a polygon to fold to a polyhedron are also sufficient! Not only that, the resulting polyhedron is unique. This means that any time you can find an Alexandrov gluing, you can be sure it will create a convex polyhedron. We will see that one catch is that Alexandrov’s proof was an “existence proof”: So you have created a particular convex polyhedron but you don’t know what it looks like!

We have already seen two Alexandrov gluings in Figure 9.1, but their foldings were obvious, both owing to their regularity and because the crease lines are self-evident. But consider the unusual folding of an equilateral triangle in Figure 9.5(a), which as we will see, produces the tetrahedron shown in (b) of the figure.

Let us first check that it is an Alexandrov gluing. Condition (a) is satisfied because no perimeter sections are left unmatched. Condition (b) is met at the gluing together of $\{x, A, B, C\}$, whose angles sum to 360° exactly, and at the four “pinch” or *fold points* $\{a, b, c, d\}$ glued to themselves, with each angle 180° , forming the flat side of the triangle. All other points glued together are $180^\circ + 180^\circ = 360^\circ$. That condition (c) is satisfied is perhaps best verified by taping a folded paper triangle according to the gluing instructions and seeing that the result is a sort of bag. I encourage the reader to try this. The triangle need not be perfectly equilateral, and any point x about a third of the way from corner A will suffice to produce this shape. Alexandrov’s theorem says that the

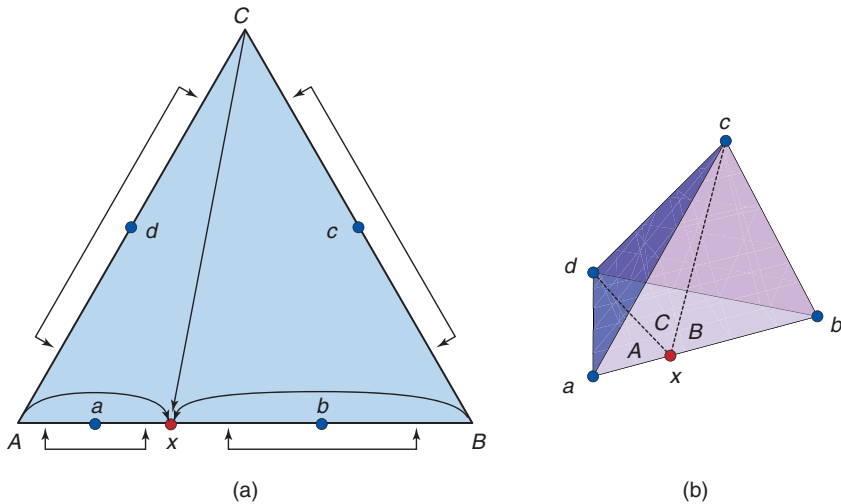


Figure 9.5. Alexandrov gluing of an equilateral triangle. (a) The three corners $\{A, B, C\}$ all glue to point x . The four fold points $\{a, b, c, d\}$ become the four vertices of the resulting tetrahedron in (b), two of whose faces are $\triangle abd$ and $\triangle bcd$. The folding of (a) to produce (b) is toward the viewer. Note that the angles at the three vertices $\{A, B, C\}$ “disappear,” forming 180° at x on tetrahedron edge ab .

resulting shape is a particular convex polyhedron. In fact it is the irregular tetrahedron, shown in (b) of the figure, which you might be able to coax out of your taped-triangle bag with some nudging.

Another Alexandrov folding of an equilateral triangle is obtained by simply creasing it down one altitude and folding it in half to a doubly covered, flat $30^\circ\text{--}60^\circ\text{--}90^\circ$ triangle. And this is the reason for the exception clause in Alexandrov’s theorem: Gluings might produce zero-volume flat “polyhedra.”

9.3 Folding Convex Polygons

We remarked earlier that the equilateral triangle folds to an infinite variety of distinct convex polyhedra. In fact, this is true for every convex polygon (Recall (p. 105) that a convex polygon is one without dents: every vertex is convex.) We now show that this surprising claim follows from Alexandrov’s Theorem (Theorem 9.1).

Take any convex polygon and mark a point x anywhere on its boundary. Walk around half the perimeter of the polygon and mark the resulting point y . Let us use the square shown in Figure 9.6(a) as an example. Next glue the perimeter half from x to y to the half from y to x . This is possible because these perimeter halves have the same length. Now we argue that this is an Alexandrov gluing.

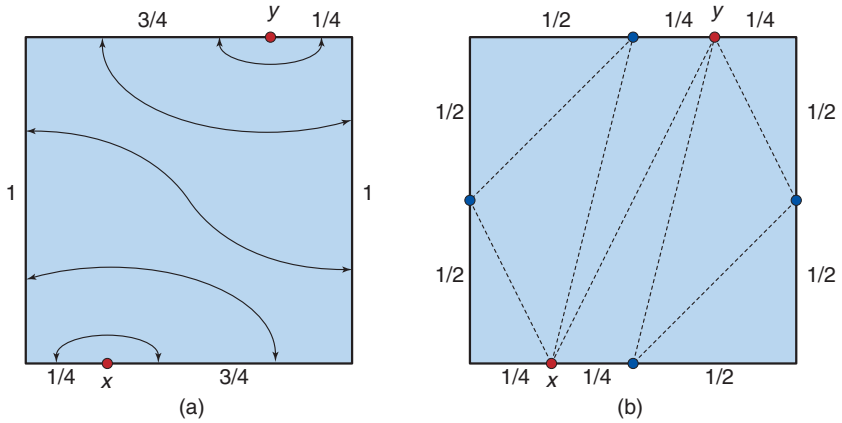


Figure 9.6. (a) A perimeter-halving folding of a unit square. x is $\frac{1}{4}$ from the lower left corner, and y is $\frac{1}{4}$ from the upper right corner. The length 2 perimeter half is glued symmetrically as indicated. The folding produces (not obviously!) an octahedron. (b) Crease pattern of edges, and vertices of octahedron. As in Figure 9.5, the corners of the square “disappear” in the folding.

Certainly it consumes all the perimeter (a), because one-half of the perimeters gets glued to the other half. The key is requirement (b): No more than 360° is glued at any one point. At the fold points x and y , the amount of angle is $\leq 180^\circ$: Exactly 180° if the fold point is not at a vertex, and less if the fold point is a vertex. Any other two points glued together either sum to exactly 360° if neither point is a vertex of P , or to less than 360° if one or the other is a vertex of P . It is here that convexity of P is used: Any convex vertex has an interior angle less than 180° . That the gluing is a topological sphere (c) can be seen if one views the perimeter-halving gluing as zipping up a pocketbook.

So, Alexandrov’s theorem says that every perimeter halving folds to a convex polyhedron. The folding of Figure 9.6(a) leads to an octahedron following the crease pattern in (b). Now if we imagine sliding x around the boundary, and y correspondingly, we obtain a “continuum” of foldings, an infinite variety. Although it is not obvious that among this continuum there are also an infinite number of different (*noncongruent*) polyhedra, that is nevertheless the case.

Figure 9.7 shows the continuum achieved by perimeter-halving foldings of a square, which I worked out together with three college students. Starting from the doubly covered $1 \times \frac{1}{2}$ rectangle achieved by creasing down a midline (3-o’clock position in the diagram) the continuum continues clockwise to the doubly covered right triangle achieved by creasing down a diagonal (9-o’clock position). This corresponds to sliding x from the midpoint of an edge of the square to an adjacent corner. Continuing sliding x repeats the shapes in mirror image (clockwise from 9- to 3-o’clock). The octahedron to which Figure 9.6(b)

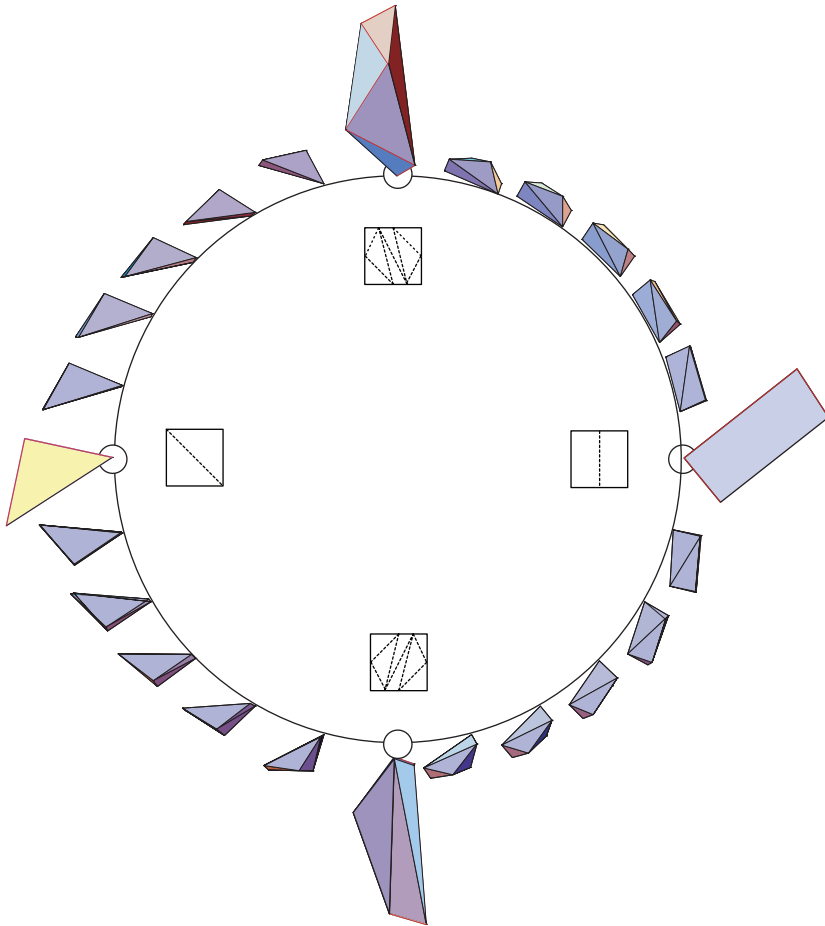


Figure 9.7. Continuum of perimeter-halvings of square. Four crease patterns are shown. The octahedron at the 6- and 12-o'clock positions corresponds to Figure 9.6(b).

folds shows up twice, one the mirror image of the other, at the 6- and 12-o'clock positions. Incidentally, this figure represents only a portion of the polyhedra foldable from a square. The full variety is even richer!

As a practical experiment, you can cut out of paper any convex polygon, start creasing it at an arbitrary x , and “zip” up the boundary from there with tape, and eventually arriving at y ; no measurement of the perimeter need be made. The result will be a handbag- or pita-like shape, which, by Alexandrov’s theorem, may be coaxed (with patience!) to reveal the creases that fold it into its unique polyhedral form.

So far researchers have explored the “space” of all foldings of regular polygons, but there remains as yet little general understanding of the phenomenon.

9.4 The Foldings of the Latin Cross

We have just canvassed the foldings of convex polygons. How about nonconvex polygons? Here we enter largely unknown territory. My coauthors (including five college students) and I decided to explore the foldings of the Latin cross, as a test case. What we found, to our surprise, is that the Latin cross folds not only to the cube (Figure 9.1(b)), but to 22 other distinct convex polyhedra: two flat quadrilaterals, seven tetrahedra, three pentahedra, four hexahedra, and six octahedra. See Figure 9.8. Here no continuum exists – the reflex vertices block the sliding possible with convex polygons that leads to the continuum of Alexandrov gluings.

How these foldings are achieved is by no means obvious. Figure 9.9 illustrates just one of the 23 foldings in detail, a delicate folding to a tetrahedron. The other foldings are equally intricate.

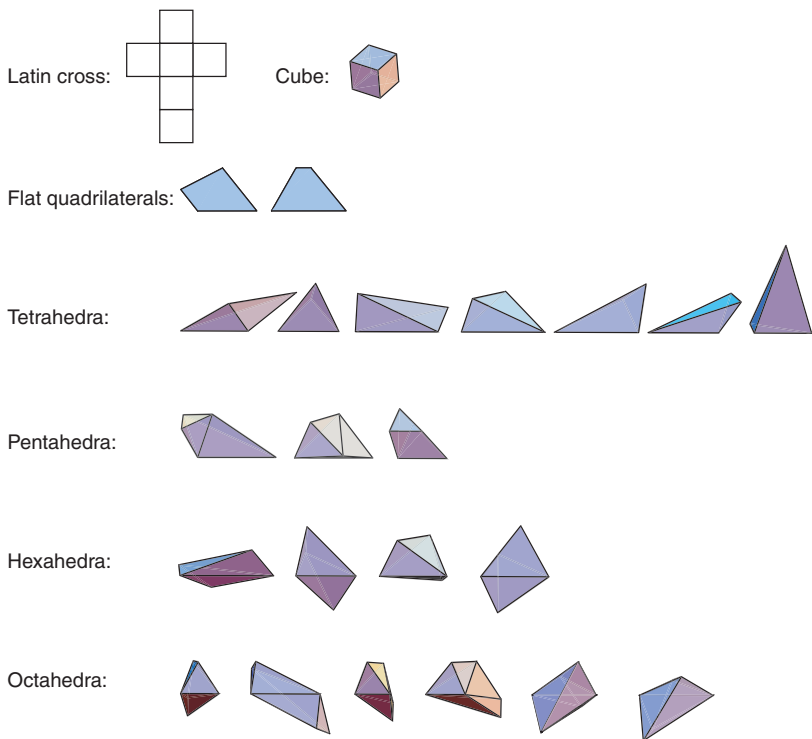


Figure 9.8. The 23 polyhedra foldable from the Latin cross.

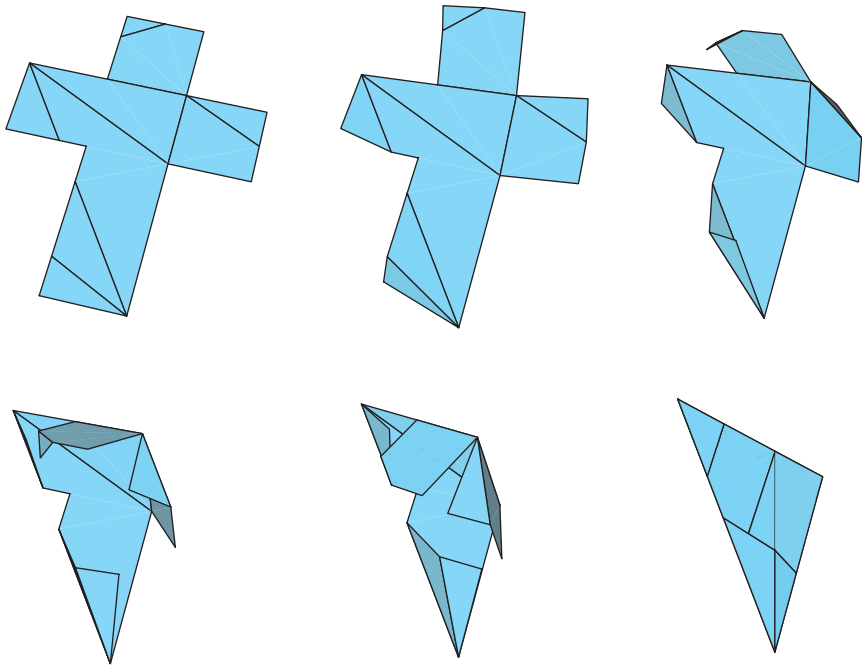


Figure 9.9. Folding the first of the tetrahedra in Figure 9.8.

Aside from this one detailed example, we are left largely without a general theory encompassing the foldings of nonconvex polygons:

OPEN PROBLEM: *Folding to Convex Polyhedra*

Which polygons fold to convex polyhedra? Is there a characterization of the shapes of polygons that do fold to convex polyhedra, distinguishing them from the shapes that do not?

Box 9.1: Algorithm

An *algorithm* is a definite procedure for accomplishing some computational task, with clearly specified steps. The steps of the algorithm must be precise enough so that they could be implemented in some appropriate programming language. Thus every algorithm is potentially encapsulated in a computer program, although many algorithms remain theoretical descriptions without programmers actually implementing them in software. An example is the procedure for unfolding an orthogonal polyhedron to a net from the previous chapter: It is an algorithm (described in a published paper), but it has not been implemented, largely because it would be rather difficult to do so.

It turns out, for subtle reasons, the reverse relationship does not hold: Not every computer program implements an algorithm! This is because a program might never stop running, but for it to constitute an algorithm, it should eventually produce an answer.

9.5**Above & Beyond**

As mentioned previously, there is an algorithm (and software) to take any given polygon and list all of its Alexandrov gluings. (See Box 9.1 on “algorithm.”) But to which (unique) polyhedra these gluings correspond is unknown. The reason is that Alexandrov’s 1948 proof of his theorem is what mathematicians call an “existence” proof: He proved that a unique polyhedron “exists” corresponding to any Alexandrov gluing, but his proof gives no hint what the polyhedron looks like. You might wonder, then, how all the polyhedra you have seen in this chapter were derived from their Alexandrov gluings? The answer is that my students and I constructed the polyhedra displayed in Figures 9.7 and 9.8 by laborious ad hoc techniques that cannot extend much beyond octahedra.

This situation dramatically changed recently when a group of researchers in Germany discovered a new proof of Alexandrov’s theorem, which is a “constructive” proof, the opposite of an existence proof: The logic of the proof leads to a method to construct the polyhedron whose existence is guaranteed by the theorem. The construction is, alas, rather complicated, requiring the solution of a particular “nonlinear differential equation” for which only approximate numerical approaches are known. Nevertheless, they have implemented software that (approximately) solves the differential equation and constructs the unique 3D shape guaranteed by Alexandrov’s theorem from a gluing. It remains to be seen if this advance will lead to new insights into the world of foldings.

I close the book with an amazing theorem: Every polygon folds to some (usually nonconvex) polyhedron! This is true even if the foldings are restricted to perimeter halvings. For a time I thought that the example in Figure 9.10 was a

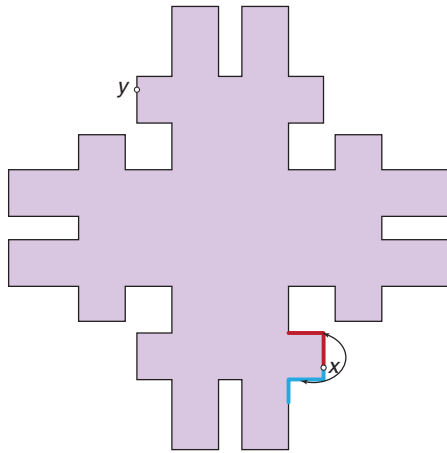


Figure 9.10. This polygon folds to a nonconvex polyhedron. Points x and y divide the perimeter into two equal halves. The folding “zips” the perimeter closed from x to y . A short initial segment in the neighborhood of x is indicated: The red portion of the boundary glues to the symmetric blue portion. This perimeter gluing continues around to y .

counterexample to this hypothesis, but instead this does fold to a nonconvex polyhedron. However, I have no idea what the polyhedron looks like!

The theorem was proved (in 1996) by two students of Alexandrov, appropriately enough: Yuri Burago and Viktor Zalgaller. Their theorem says that any gluings of polygons (including a perimeter-halving folding of a single polygon) can be realized in 3D as a polyhedral surface. The surface is in general nonconvex, and has a large number of flat triangle faces. Their proof is constructive but quite complicated. At one point they incorporate an intricate idea of John Nash, famous for both his mathematical results (he shared the 1994 Nobel Prize in Economics) and for being the subject of the 2001 movie *A Beautiful Mind*. The complexity of the techniques used in their proof leaves me without a clear idea of what the polyhedron that results from the folding in Figure 9.10 looks like. I don’t know, even roughly, how many faces it has! But I know it exists.

Here we have very much reached the frontier of current knowledge. Perhaps by the time you read this there will be a simpler proof of the Burago-Zalgaller theorem, as this is an active area of current research. Check the web site <http://www.howtofoldit.org> for updates!

10 Further Reading

Here I provide two types of resources: first, general “further reading” on the topics covered in each chapter, usually more technical and detailed presentations of similar material (there is rarely any less technical presentation available). Second, some of the original sources for the material are cited, even if they are at the graduate-student level. I try to indicate the level of expertise needed to benefit from each resource.

The most frequent citation you’ll see below is to my own monograph, *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*, coauthored with Erik Demaine, out of which this book grew. I’ll refer to this throughout as *Geometric Folding Algorithms*. This monograph is targeted at graduate students and professional researchers, but strives to be largely accessible with considerably less preparation. It has 421 scholarly citations, and rather than repeat many of these here, I concentrate on what is directly relevant to the material presented in this book.

Erik D. Demaine and Joseph O’Rourke. *Geometric Folding Algorithms: Linkages, Origami, Polyhedra*. Cambridge University Press, July 2007.
<http://www.gfalop.org>.

Chapter 1

Much of the material on reachability, including the Two-Kinks Theorem, is described in more detail in my college textbook *Computational Geometry in C*, [Chapter 8](#). The Above & Beyond material on intractability is covered in [Chapter 5](#) of *Geometric Folding Algorithms*. NP-completeness is covered in any algorithms textbook. I recommend one here but there are plenty of others.

Joseph O'Rourke. *Computational Geometry in C*. Cambridge University Press, 2nd edition, 1998.

Thomas H. Cormen, Charles E. Leiserson, Ron L. Rivest, and Cliff Stein. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 3rd edition, 2009.

Chapter 2

Kempe's remarkable idea for a linkage that "signs your name" (a formulation that was only articulated a century later) is described in his still readable 1887 monograph, *How to Draw a Straight Line*. The recounting of the history is drawn from the delightful book, *How Round Is Your Circle*, which includes a chapter on straight-line linkages, written in an accessible style. The fascinating story of *The Turk* is drawn from the book of the same title by Tom Standage. For more technical details on linkages – including Kempe's Universality Theorem – see Chapter 3 of *Geometric Folding Algorithms*.

Alfred Bray Kempe. *How to Draw a Straight Line: a Lecture on Linkages*. Macmillan & Co., 1877.

John Bryant and Chris Sangwin. *How Round Is Your Circle?: Where Engineering and Mathematics Meet*. Princeton Univ. Press, 2008.

Tom Standage. *The Turk: The Life and Times of the Famous Eighteenth-Century Chess-Playing Machine*. Walker & Co., 2002.

Chapter 3

The Unit-90° Theorem (Theorem 3.1) is due to Nadia Benbernou, who discovered it when she was a college student. A generalization of the Piercing Theorem (Theorem 3.2) that establishes necessary and sufficient conditions for any fixed-angle chain to be in maxspan configuration can be found in a technically advanced paper presented at a conference on "reconfigurable mechanisms and robots." The Folding@Home web site is at <http://folding.stanford.edu/>. If the one open problem I cited (p. 54) is not enough to keep you occupied, many others on locked chains are detailed in Chapter 6 of *Geometric Folding Algorithms*.

Nadia Benbernou. Fixed-angle polygonal chains: Locked chains and the maximum span. Undergraduate thesis, Smith College, 2006.

Nadia Benbernou and Joseph O'Rourke. On the maximum span of fixed-angle chains. In *Proceedings of the 18th Canadian Conference Computational Geometry*, pages 93–96, 2006.

Ciprian S. Borcea and Ileana Streinu. Extremal configurations of manipulators with revolute joints. In *Reconfigurable Mechanisms and Robots*, pages 279–284. American Society of Mechanical Engineers, 2009.

Chapter 4

The material in this chapter is covered in great technical detail in Chapters 12–14 of *Geometric Folding Algorithms*, including a proof of the sufficiency of Kawasaki's Theorem (Theorem 4.4) that I did not provide in this book. Perhaps the most accessible presentation of this material is that in Thomas Hull's *Project*

Origami, Activity 16, on which I relied heavily. This entire book is a rich source of the mathematics of origami, aimed at college instructors, but much of it accessible to students. The roots of computational origami are described in a paper by the Demaines in *Origami*³, the proceedings of the 3rd conference on aspects of origami. The Bern-Hayes result (in the Above & Beyond section) that shows that flat foldability is intractable was presented at a discrete algorithms conference. Bern and Hayes are also coauthors of the Fold and One-Cut Theorem in the next chapter.

Thomas Hull. *Project Origami: Activities for Exploring Mathematics*. A K Peters, 2006.
Erik D. Demaine and Martin L. Demaine. Recent results in computational origami.

In *Origami*³: *Proceedings of the 3rd International Meeting of Origami Science, Mathematics, and Education*, pages 3–16. A K Peters, 2002.

Marshall Bern and Barry Hayes. The complexity of flat origami. In *Proceedings of the 7th ACM-SIAM Symposium Discrete Algorithms*, pages 175–183, 1996.

Chapter 5

The original “nearly all” proof of the Fold and One-Cut Theorem (Theorem 5.1) based on skeletons was found by Erik Demaine, Marty Demaine (son and father respectively), and Anna Lubiw, the academic adviser of both Demaines, and presented at the annual discrete algorithms conference. The original disk-packing proof was found by Erik collaborating with Marshall Bern, David Eppstein, and Barry Hayes, and presented two years later at *Origami*³. The flaw in this latter proof was discovered by Bern and Hayes. They repaired it in a paper presented at a conference in Sao Paulo, Brazil, and Erik and I repaired it in Chapter 17 of our book, *Geometric Folding Algorithms*. The latter remains the most detailed exposition of the full, correct proof. The Brazil paper is also the one that “half-answers” the flattening question (p. 83).

Erik D. Demaine, Martin L. Demaine, and Anna Lubiw. Folding and one straight cut suffice. In *Proceedings of the 10th Annual ACM-SIAM Symposium Discrete Algorithms*, pages 891–892, January 1999.

Marshall Bern, Erik D. Demaine, David Eppstein, and Barry Hayes. A disk-packing algorithm for an origami magic trick. In *Origami*³: *Proceedings of the 3rd International Meeting of Origami Science, Mathematics, and Education*, pages 17–28. A K Peters, 2002.

Marshall Bern and Barry Hayes. Origami embedding of piecewise-linear two-manifolds. In *Proceedings of the 8th International Latin American Symposium Theoretical Informatics*, pages 617–629. Springer LNCS 4957, 2009.

Chapter 6

Koryo Muira called what is now known to us as the Muira Map Fold the “Developable Double Corrugation Surface” in *Origami*³. The best source on rigid origami is again Hull’s *Project Origami*. My presentation relies on his descriptions of the Muira Map Fold, the Square Twist, and dihedral angle relationships (Activities 18–22). Figures 6.4 and 6.8 are based on his Mathematica animations.

The Shopping Bag Theorem (Theorem 6.1) is from *Origami*⁴. Hinged dissections are described in Frederickson's delightful book, *Hinged Dissections: Swinging and Twisting*. My Figure 6.16 is based on his Figure 20.1. The breakthrough result that every 2D dissection may be converted to a hinged dissection was presented at a computational geometry conference in 2008.

Koryo Miura. The application of origami science to map and atlas design. In *Origami*³: *Proceedings of the 3rd International Meeting of Origami Science, Mathematics, and Education*, pages 137–145. A K Peters, 2002.

Devin J. Balkcom, Erik D. Demaine, Martin L. Demaine, John A. Ochsendorf, and Zhong You. Folding paper shopping bags. In *Origami*⁴: *Proceedings of the 4th International Meeting of Origami Science, Mathematics, and Education*, pages 315–334. A K Peters, 2006.

Greg N. Frederickson. *Hinged Dissections: Swinging & Twisting*. Cambridge University Press, 2002.

Timothy G. Abbott, Zachary Abel, David Charlton, Erik D. Demaine, Martin L. Demaine, and Scott D. Kominers. Hinged dissections exist. In *Proceedings of the 24th Annual ACM Symposium on Computational Geometry*, pages 110–119, 2008. To appear in the journal *Discrete & Computational Geometry*, 2011.

Chapter 7

The material in this chapter is drawn from Chapter 22 of *Geometric Folding Algorithms*. The first explicit formulation of Dürer's open problem appeared in a 1975 paper by Geoffrey Shephard, although others, including Richard Guy, wondered about it much earlier. Figure 7.4 is based on nets computed by Eric Weisstein's code `Archimedean.m`, and `PolyhedronOperations.m` available at <http://www.mathworld.wolfram.com/packages/>. Joseph Malkevitch discusses the history of the discovery of polyhedra in the book *Shaping Space*, including the 13 versus 14 Archimedean solids controversy (mentioned in the Glossary under “semi-regular polyhedron”).

Geoffrey C. Shephard. Convex polytopes with convex nets. *Mathematical Proceedings of the Cambridge Philosophical Society*, 78:389–403, 1975.

Joseph Malkevitch. Milestones in the history of polyhedra. In Marjorie Senechal and George Fleck, editors, *Shaping Space: A Polyhedral Approach*, pages 80–92. Birkhäuser, Boston, 1988.

Chapter 8

For the material in this chapter, I can only cite more technical versions. I wrote a survey that explains grid unfoldings, and organizes the confusing welter of partial results on orthogonal polyhedra. The Manhattan towers algorithm (Figure 8.6) and the algorithm to unfold any orthogonal polyhedron (Figure 8.8) are detailed in two journal articles.

- Joseph O'Rourke. Unfolding orthogonal polyhedra. In Jacob E. Goodman, Janos Pach, and Richard Pollack, editors, *Proceedings Snowbird Conference Discrete and Computational Geometry: Twenty Years Later*, pages 307–317. American Mathematical Society, 2008.
- Mirela Damian, Robin Flatland, and Joseph O'Rourke. Epsilon-unfolding orthogonal polyhedra. *Graphs and Combinatorics*, 23[Suppl]:179–194, 2007. Akiyama-Chvátal Festschrift.
- Mirela Damian, Robin Flatland, and Joseph O'Rourke. Unfolding Manhattan towers. *Computational Geometry: Theory & Applications*, 40(2):102–114, 2008.

Chapter 9

An earlier version of this chapter appeared in a “yearbook” published by the National Council of Teachers of Mathematics. Both are drawn from the more detailed presentation in Chapter 25 of *Geometric Folding Algorithms*, which contains references to the relevant literature.

The final Burago-Zalgaller theorem appeared in a Russian journal (translated to English), and I posted a short note describing their result to the electronic “arXiv.”

Joseph O'Rourke. Folding polygons to convex polyhedra. In Timothy V. Craine and Rheta Rubenstein, editors, *Understanding Geometry for a Changing World: National Council of Teachers of Mathematics, 71st Yearbook*, pages 77–87. National Council of Teachers of Mathematics, 2009.

Yuri D. Burago and Viktor A. Zalgaller. Isometric piecewise linear immersions of two-dimensional manifolds with polyhedral metrics into \mathbb{R}^3 . *St. Petersburg Mathematical Journal*, 7(3):369–385, 1996. Translated by Sergei G. Ivanov.

Joseph O'Rourke. On folding a polygon to a polyhedron. arXiv:1007.3181v1 [cs.CG]. <http://arxiv.org/abs/1007.3181>

Glossary

algorithm. See Box 9.1.

array. See Box 8.1.

amino acid. See *protein*.

annulus. The region between two concentric (same center) circles in 2D. Thus the region takes the shape of a washer or ring. In 3D, the region between two concentric spheres is often called a *spherical shell*.

Archimedean solids. The 13 semiregular polyhedra, discussed by Archimedes in a now-lost manuscript, only known because Pappus cited his discovery of “the thirteen polyhedra” 500 years later. Rediscovered by Johannes Kepler in 1620. See also the entry on *semiregular polyhedron*, which explains that perhaps we should say there are 14 Archimedean solids.

commutative. A mathematical operation is commutative if changing its order of application does not change the result. Addition and multiplication of numbers is commutative, but subtraction is not: $7 - 5 \neq 5 - 7$. For analyzing linkages, the fact that vector addition is commutative is often helpful.

configuration. A configuration of a geometric object is a particular positioning of that object in the plane or in space. A configuration of a linkage specifies the position of each joint at particular spots, which then determines the orientation of each link, and so the angles between adjacent links. A linkage with particular lengths can be viewed as an abstract object that has many possible configurations, which together form its *configuration space*. Similarly, specifying the dihedral angles at the creases of a shopping bag (Chapter 6) determines a configuration of the bag.

convexity. See Box 7.1.

degree. Degree is both a measure of angle, one-360th of a full circle, and a count of the number of lines or edges coming into (*incident to*) some point. The latter use derives from a subfield of mathematics called *graph theory*, which analyzes spanning trees and skeletons of polyhedra, among other graphs. The number of creases in an origami construction that are incident to a vertex is called the

degree of that vertex. Here a “crease” is not considered to constitute a crease unless the dihedral angle differs from 180° .

degenerate. A degenerate situation or configuration is a very special one, often somehow missing all the features of a nondegenerate situation. For example, if you collapse a triangle so that one angle becomes 180° and the other two 0° , the resulting shape (which would appear visually as a line segment) can be considered a zero-area degenerate triangle. The flat zero-volume polyhedra that might result from Alexandrov’s Theorem (Theorem 9.1) are degenerate polyhedra. Generically the intersection of two circles in the plane is either empty or two points, but as we saw in Figure 1.10, there are degenerate situations where two circles intersect in a single point of tangency, or coincide and so intersect throughout their circumferences.

degrees of freedom. The number of distinct parameters needed to specify the configuration of a linkage. Often abbreviated *dof*. A 2-link arm in the plane, with the shoulder pinned to the plane, has two degrees of freedom: one angle to specify the orientation of the first link and one to specify the orientation of the second link. A 2-link arm in 3D has four degrees of freedom, as it requires two angles (say, longitude and latitude) to specify the position of each arm.

dihedral angle. The angle formed at the ‘V’ between two planes, measured in 3D. For example, Figure Ans.21 shows that the dihedral angle between the faces of a regular tetrahedron is about 70.5° .

dissection. A shape B is a dissection of shape A if A can be partitioned into a finite number of pieces and rearranged to form B .

edge unfolding. A cutting of the surface of a polyhedron along its edges that permits the surface to be flattened to a plane. If the unfolded surface avoids self-overlap in the plane, it is called a *net*. A *general unfolding* removes the restriction that the cuts follow edges of the polyhedron.

induction. See Box 1.2.



linkage. A collection of rigid links or rods, modeled as line segments, hinged together at joints. A linkage that forms a single path is called a *robot arm*. Typically the first joint is pinned to a fixed point, the *shoulder*, about which it may freely rotate. Often interest is focused on determining where the last point of the arm, the *hand*, can reach. This last point might be the tip of welding gun for an industrial robot in an automotive factory. Linkages with connections more complex than just a single path are often called *mechanisms*.

necessary. A condition that must hold for a claim to be true. For example, it is necessary that two polyhedra have the same volume for there to be a dissection of one to the other. However, this volume condition is not sufficient: There are polyhedra of the same volume that are not dissections of one another. A cube and regular tetrahedron of the same volume are examples, although the proof of this fact (by Max Dehn in 1901) is not simple. The goal in much of mathematics is to find conditions that are both necessary and sufficient. This was achieved for

2D dissections: Any two polygons with the same area have a dissection. In this case, “same area” is both (obviously) necessary and (not obviously!) sufficient.

net. A planar polygon that results from cutting the surface of a polyhedron along edges (see *edge unfolding*). If the polygon instead derives from cuts that do not all follow edges of the polyhedron, it is called a *general net*.

NP-complete, NP-hard, PSPACE-complete. These are all technical terms classifying the difficulty of problems in terms of their “computational complexity.” The distinction between them is beyond what we can describe here, but they all mean “intractable” in the sense that no one has found an algorithm to solve any of them efficiently, and moreover it seems unlikely that any such algorithm exists. “Efficiently” could be interpreted as, say, finishing in less than one hundred years on the fastest conceivable computer.

origami folds. A *mountain fold* is a crease in paper that appears like a mountain ridge. In origami instructions, it is notated as a dot-dash pattern, $-\cdot-\cdot-\cdot-$, although I use a solid red line  in the color drawings. A *valley fold* is a crease that appears like a valley, traditionally notated by dashes $-----$, and with solid green lines  in my drawings. Every mountain fold is a valley fold from the other side, and vice versa.

pantograph. A linkage that is used to copy, enlarge, or shrink a drawing. The word is also used for linkages that have a similar structure but are used for a train to maintain electrical contact with overhead wires.

Platonic solid. See *regular polyhedron*. Although known perhaps 1,000 years before Plato, they bear his name because of the role they played in his dialog, *Timaues*, around 360 B.C. The five solids (Figure 7.5) were described precisely by Euclid in the *Elements* around 300 B.C.

polygon. A planar figure composed of straight segments joined at vertices, and forming a closed loop. Pairs of adjacent segments meet at their shared vertex, but otherwise pairs of nonadjacent segments do not touch one another. A polygon encloses some area inside, and the remainder of the plane is outside of or exterior to it.

polyhedron. A surface consisting of flat sections called *faces*, each of which is a portion of a plane. A polyhedron is the 3D analog of a polygon. Particular examples include the regular and semiregular polyhedra.

protein. A molecule made of amino acids arranged in a linear chain. There are 20 different amino acids found in organic life, each composed of between 13 and 27 carbon (C), nitrogen (N), hydrogen (H), and oxygen (O) atoms. They all have a common core, consisting of a nitrogen atom with two attached hydrogens (NH_2), a carbon with an attached hydrogen and *side chain* R (CRH), and another cluster centered on a carbon (COOH). The variation among the different amino acids comes in the variety of attached side chains R. The core of the core forms the *CCN backbone* of a protein described in Section 3.2.

regular polygon. A polygon with all angles equal and all edge lengths equal: an equilateral triangle, a square, a regular pentagon, etc.

regular polyhedron. A polyhedron each of whose faces is the same regular polygon, and with each vertex surrounded by the same number of faces. There are just five, the five Platonic solids.

rhombus. A four-sided planar figure, with all four sides the same length. A square is a special version of a rhombus, and a rhombus is a special version of a parallelogram. Plural: rhombi.

semiregular polyhedra. A polyhedron each of whose faces is some regular polygon, and with each vertex surrounded by the same number and type of faces, “similarly arranged.” There are two different notions of what constitutes “similarly arranged,” one of which leads to 14 polyhedra, the other to 13. (Unfortunately, the distinction cannot be easily explained, as it involves “orbits under symmetries.”) The 13 are generally known as the Archimedean solids, although it may be that Archimedes had in mind the definition that leads to 14 polyhedra. The 14th has an impressive name: the pseudorhombicuboctahedron!

sufficient. See necessary.

theorem. See Box 1.1.

tree. A graph without cycles. A cycle is a loop of edges starting and returning to a vertex. Every tree of n vertices has $n - 1$ edges. So the tree that looks like a ‘Y’ has four vertices and three edges.

triangle inequality. See Box 1.4.

truncation. The operation of clipping off a vertex v of a polyhedron by slicing it with a plane that has v to one side and the remainder of the polyhedron to the other side. For example, the truncated icosahedron (Figure 7.3) can be constructed by clipping off each of the 12 vertices of a regular icosahedron. Each truncation results in a pentagon, and the 20 triangular faces of the icosahedron each get clipped to hexagons.

vector. See Box 1.3.

vertex. A vertex is a point, but a point distinguished in some way. The joint between two links of a linkage is a vertex (Chapters 1–3). The corners of a polygon or a polyhedron (Chapter 7–9) are vertices (that’s the plural of “vertex”). The spot in the center of a piece of paper with incident creases is an origami vertex (Chapter 4). It can happen that a point is distinguished as a vertex even though there is no corner or “bump” there, because, for example, the two edges of the polygon incident to it form an angle of π .

Answers to Exercises

Chapter 1

Answer to Exercise 1.1. See Figure [Ans.1](#).

Answer to Exercise 1.2. The first triple fails, because, even though $C = 3 \leq 15 = A + B$, it is not the case that $|A - B| = 5 \leq 3 = C$. The second triple does satisfy the triangle inequality: $C = 6 \leq 15 = A + B$, $|A - B| = 5 \leq 6 = C$. The third does as well, the famous 3-4-5 right triangle. See Figure [Ans.2](#). The fourth triple does not, because $C > A + B$.

Answer to Exercise 1.3.

Link Lengths	r_+	r_-
(1,2,3,4,5)	15	0
(1,2,10,3,4)	20	0
(1,1,1,1,5)	9	1
(12,2,5,4)	23	1

Answer to Exercise 1.4. There are two ways to reach $v_2 = (3, 1)$, and one way to reach $v_2 = (0, 2)$. See Figure [Ans.3\(a\)](#).

Answer to Exercise 1.5. There are an infinite number of solutions. Two are shown in Figure [Ans.3\(b\)](#). A solution must place v_2 at the intersection of the red circle of radius $\ell_2 = 1$ centered on v_1 , and the green circle of radius $\ell_3 = 2$ centered on $v_3 = (2, 2)$.

Answer to Exercise 1.6. Possible uniquely reaching configurations occur when a joint angle is either 180° or 0° : stretched out straight or jack-knifed closed completely. In those cases, the link lengths add or subtract, and a circle with radius their sum centered on the shoulder v_0 can be traced out by the “hand”

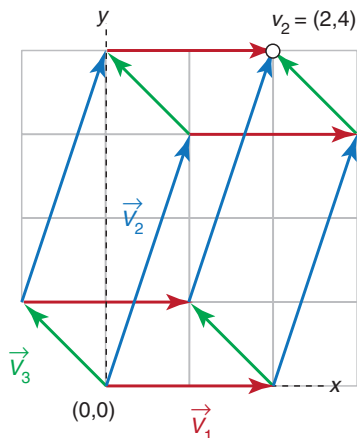


Figure Ans.1. The six permutations of adding \vec{V}_1 , \vec{V}_2 , \vec{V}_3 all reach $v_2 = (2,4)$.

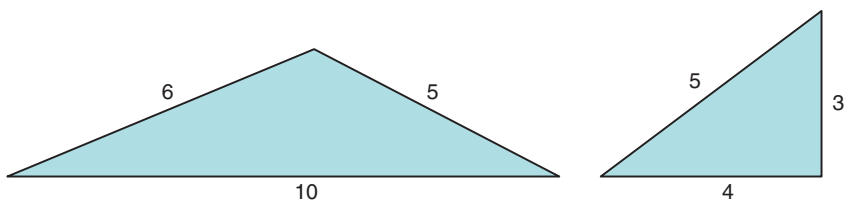


Figure Ans.2. Triangles made from lengths $(10, 5, 6)$ and $(4, 5, 3)$.

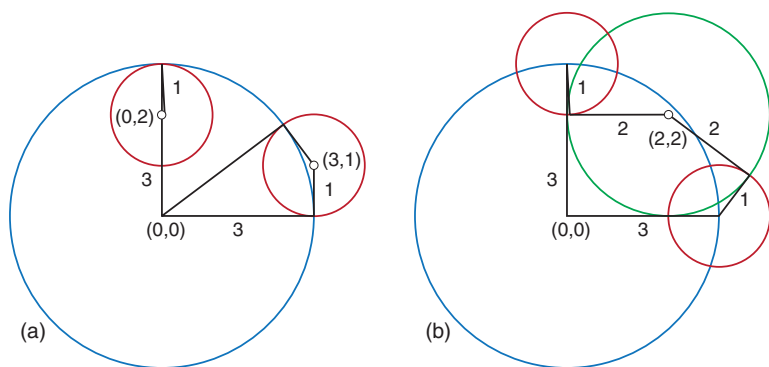


Figure Ans.3. Ways for $A_2 = (3,1)$ to reach $(3,1)$ and $(0,2)$.

v_3 . All the points on the circle of radius $2 + 1 + 2 = 5$ can only be reached by stretching all joints out straight. All the points on the circle of radius $2 - 1 + 2 = 3$ can only be reached by forming a zig-zag with the 2nd link opposing the 1st and 3rd. But the points on the circle of radius $2 + 1 - 2 = 1$ can be reached also by $2 - 1 - 2$: reversing just the third link is equivalent to reversing the 2nd and 3rd. See Figure Ans.4.

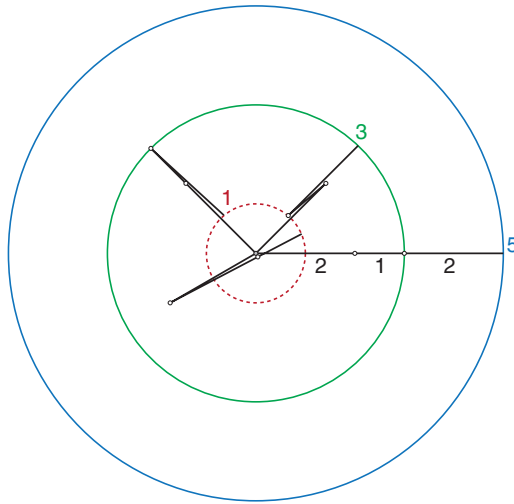


Figure Ans.4. The points on the circles of radius 5 and 3 are uniquely reachable by an $A_3 = (2, 1, 2)$ arm.

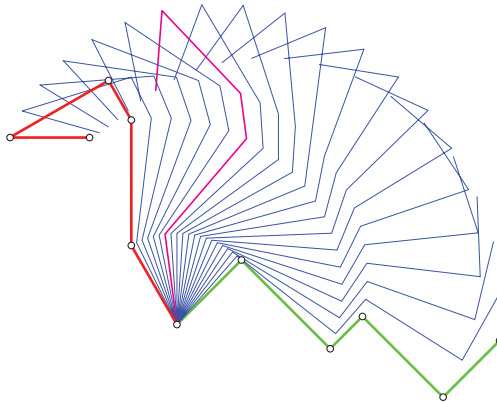


Figure Ans.5. Interpolating turn angles avoids self-crossing. One intermediate configuration is highlighted.

Answer to Exercise 1.7. Indeed there is no self-crossing of the chain when the turn angles are interpolated in this specific example. See Figure Ans.5.

It would take some effort to prove this formally for this specific example, and it is not true in general.

Answer to Exercise 1.8. The ruler can folding into length 24, by turning every joint to 0° ; see Figure Ans.6. Thus $23 - 15 + 16$ has length 24, and then $-17 + 9$ fits inside that length. Note that although $23 + 17 = 15 + 16 + 9$, and so the ruler can be arranged so that $v_0 = v_5$, that folding has overall length 31. So arranging for

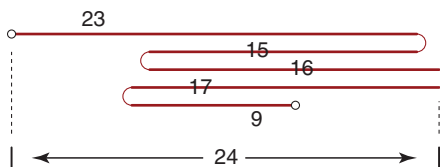


Figure Ans.6. The most compact folding of the “ruler” $A_5 = (23, 15, 16, 17, 9)$.

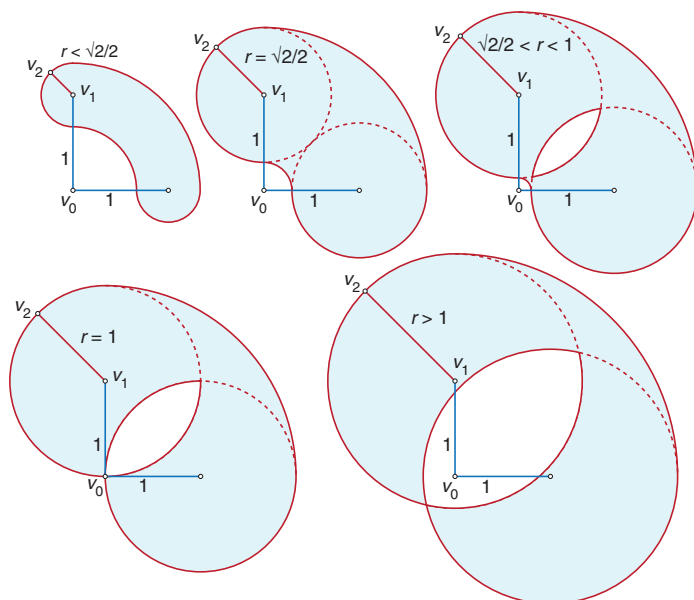


Figure Ans.7. The joint at v_0 is limited to a 90° range, but the joint at v_1 can turn a full 360° .

the hand to touch the shoulder does not necessarily result in the most compact folding.

Answer to Exercise 1.9. See Figure Ans.7. When r is much smaller than 1, the region is outlined by two quarter-circle arcs of radius $1+r$ and $1-r$, connected by two semicircle arcs of radius r . This continues up to the critical value of $r = \sqrt{2}/2$. For $r > \sqrt{2}/2$, a “lune-shaped” hole opens up, into which v_2 cannot reach. This lune grows in size as r is increased until it contacts the boundary at $r = 1$, when the inner arc degenerates to a point at v_0 (because $1-r = 0$). So the region is outlined by three semicircle arcs and includes a lune hole. For $r > 1$, the lune is larger but still connected to the outer boundary at a point along a 45° line from v_0 . The lune never touches the right outermost quarter-circle arc.

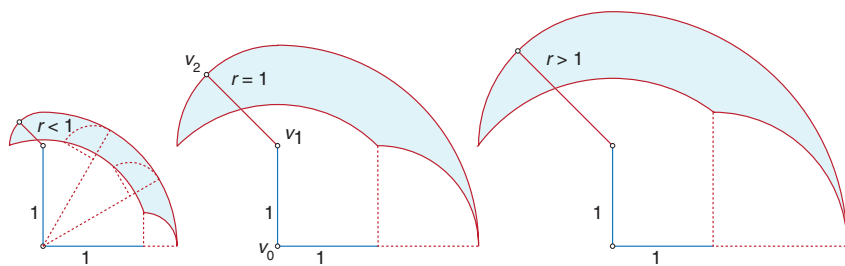


Figure Ans.8. Both the joint at v_0 and the joint at v_1 are limited to 90° ranges.

Answer to Exercise 1.10. See Figure Ans.8. The region is outlined by three circle arcs: a quarter circle arc of radius $1 + r$, attached on either end to quarter-circle arcs of radius r , which are connected by a rotated quarter-circle arc of radius $\sqrt{1 + r^2}$. The lune from the previous problem never appears, because it was formed by rotations at v_1 that are outside its 90° range here. This shape is perhaps best thought of as sweeping the quarter circle of radius r through 90° .

Chapter 2

Answer to Exercise 2.1. The extreme configuration for a counterclockwise around x occurs when the middle link ab aligns with by , as shown Figure Ans.9. Let h and δ be the height and base of the shaded right triangle in the figure. We must have $(1 - \delta)^2 + h^2 = 1$ because the radius $|xa|$ of the upper circle is 1. And we must have $(1 + \delta)^2 + (1 + h)^2 = 4$ because $|ay| = 2$. Solving these two equations leads to $\delta = \frac{1}{5}$ and $h = \frac{3}{5}$. So the angle at x with respect to the horizontal is determined by a triangle with base $\frac{4}{5}$ and height $\frac{3}{5}$. So that angle has tangent $\frac{3}{4}$, and so is $\tan^{-1} \frac{3}{4} \approx 36.9^\circ$. In this configuration, point b has moved clockwise the complement of that angle, about 53.1° . Because of the symmetry of the linkage, this means that point a moves between 36.9° counterclockwise and 53.1° clockwise, for a total of 90° .

Note that the total 90° excursion holds regardless of the length of the middle link, as long as the other two are the same length, because the extreme configurations naturally map out the rectangle visible in Figure Ans.9.

Answer to Exercise 2.2. We want $A' = 5A$. So we could choose $|xa| = 1$ and $|xc| = 5$. This means that $|ay| = 1$ and $|by| = 4$. So the two long links of length 5 include a 1×4 parallelogram $acby$.

Answer to Exercise 2.3. This is accomplished by the pantograph in Figure 2.10(b): A magnification by $\frac{3}{2}$ at point z is a reduction by $\frac{2}{3}$ at point y . So just letting z drive y in that pantograph achieves a two-thirds reduction.

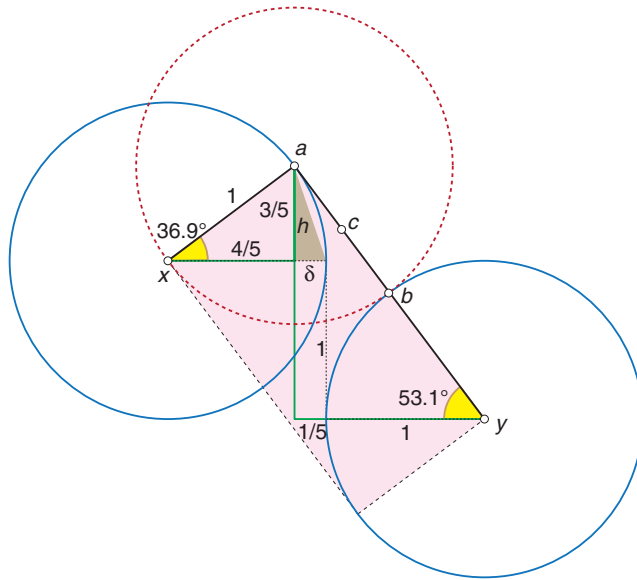


Figure Ans.9. Extreme configuration for Watt linkage formed of three unit-length links.

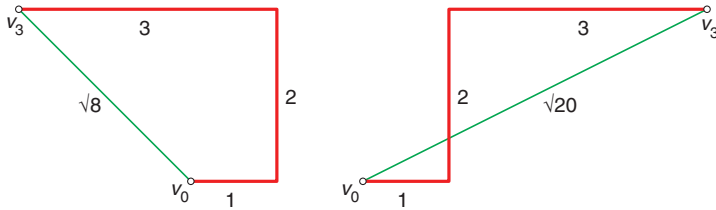


Figure Ans.10. The two planar configurations of a $(1,2,3)$ 90° -chain.

Chapter 3

Answer to Exercise 3.1. There are only two distinct planar configurations of a 3-link 90° chain: either two left turns at the joints, or a left and right turn; see Figure Ans.10. The first case leads to a span of $\sqrt{2^2 + 2^2} = \sqrt{8} \approx 2.8$, whereas the second staircase configuration achieves the maxspan of $\sqrt{4^2 + 2^2} = \sqrt{20} \approx 4.5$. Note that this accords with the Piercing Theorem (Theorem 3.2).

Answer to Exercise 3.2. No, the span cannot be increased by rotating the third link out of the plane. Figure Ans.11 shows why. If the first two links are kept fixed and the third rotated through its full freedom of motion, v_3 , the tip of the third link, traces out a circle of radius 3 centered on v_2 , with the disk of the circle perpendicular to the $v_1 v_2$ middle link. The question then becomes: Are any of

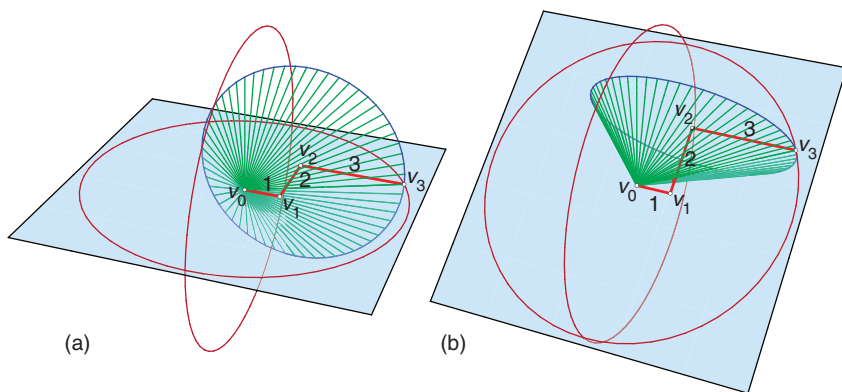


Figure Ans.11. Two views of rotating v_3 out of the plane of $\{v_0, v_1, v_2\}$.

the points on this circle (blue in the figure) further away from v_0 than is the planar span of $\sqrt{20}$? Those distances from v_0 are drawn green in the figure. That they are all shorter than $\sqrt{20}$ becomes more plausible if we imagine a sphere of radius $\sqrt{20}$ centered on v_0 , indicated by the two red great circles in the figure. The blue circle of possible v_3 positions lies entirely inside the $\sqrt{20}$ sphere, and just touches it at the one point that corresponds to the maxspan, achieved by the staircase configuration.

In fact it is a theorem (due to Nadia Benbernou) that the maxspan of any 3-link fixed-angle chain is achieved in a planar configuration.

Answer to Exercise 3.3. The 5-link chain in Figure 3.8(a) has span $\sqrt{3^2 + 2^2} = \sqrt{13} \approx 3.6$, whereas the 4-link chain in (b) has span $\sqrt{2^2 + 2^2} = \sqrt{8} \approx 2.8$. For n even, the maxspan is

$$\sqrt{(n/2)^2 + (n/2)^2} = n/\sqrt{2},$$

and

$$\sqrt{((n+1)/2)^2 + ((n-1)/2)^2} = \sqrt{n^2 + 1}/\sqrt{2}$$

for n odd.

Answer to Exercise 3.4. As the card opens, point c' revolves around a circle that is centered on the segment $a'b'$, which lengthens in 3D with opening. Segment $a'b'$ is perpendicular to the disk bounded by the circle, which remains in the card “midplane” throughout. Point c' traverses half of the circle, 180° . Point c also moves on the same circle, but only covers 90° .

The same motion holds for spinners with longer spirals: The central tip c' spins on a circle about $a'b'$.

Answer to Exercise 3.5. Weave four links of a regular 5-pointed star as shown in Figure Ans.12. Both v_0 and v_4 are above the plane containing $\{v_1, v_2, v_3\}$, and both

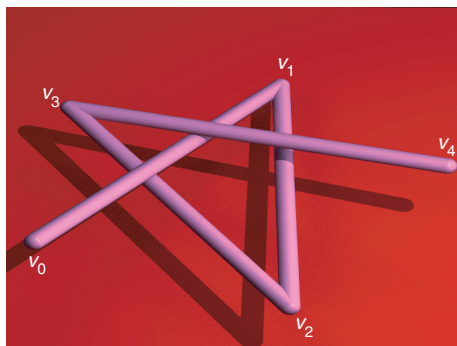


Figure Ans.12. A locked 4-chain with fixed 36° angles.

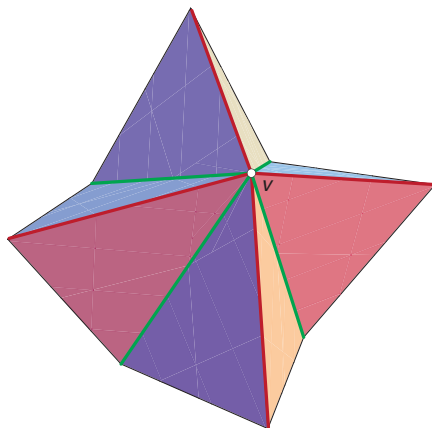


Figure Ans.13. Four mountain creases want to create four interleaved valley creases.

the first and last links, v_0v_1 and v_3v_4 , are slightly longer than the middle three links. Thus spinning the last link v_3v_4 link about v_2v_3 results in it banging into the first link v_0v_1 near its tip v_0 . And so the configuration cannot be unraveled into a planar zig-zag staircase.

There is nothing special about 36° ; it just makes it easier to see. Any fixed angle smaller than 60° leads to a similar locked-chain configuration (although that is by no means obvious).

Chapter 4

Answer to Exercise 4.1. Manipulation of the paper in 3D seems to force valley folds between the four mountain creases, as depicted in Figure Ans.13. Flattening this shape into the plane “erases” two of the four valley creases, a consequence of the Maekawa-Justin Theorem (Theorem 4.2).

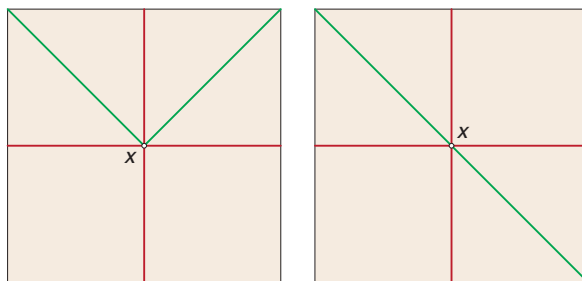


Figure Ans.14. Add two valley creases to permit four mountain creases to flatten.

Answer to Exercise 4.2. Adding two valley creases at 45° between any pair of mountain creases allows the construction to be flattened. The two distinct ways to do this are shown in Figure Ans.14. The Maekawa-Justin Theorem is verified: $M - V = 4 - 2 = 2$. Note that flattening the 3D shape in Figure Ans.13 leads one of the two patterns in Figure Ans.14.

Answer to Exercise 4.3. Figure 4.11 satisfies the Even-Degree Theorem (Theorem 4.1) because the vertex had even degree: 8. It satisfies the Maekawa-Justin Theorem (Theorem 4.2) because $M - V = 5 - 3 = 2$. However, it fails to satisfy the Local Min Theorem (Theorem 4.3) because the 30° wedge between five- and six-o'clock is a local minimum – being surrounded on both sides by 60° angles – but it is delimited by two mountain folds. Similarly the 30° wedge between eight- and nine-o'clock violates the Local-Min Theorem.

Answer to Exercise 4.4. The alternating angle sum is:

$$\begin{aligned}
 30^\circ - 60^\circ + 30^\circ - 60^\circ + 30^\circ - 60^\circ + 30^\circ - 60^\circ &= \\
 30^\circ + 30^\circ + 30^\circ + 30^\circ - 60^\circ - 60^\circ - 60^\circ - 60^\circ &= \\
 120^\circ - 240^\circ &= \\
 -120^\circ &\neq 0
 \end{aligned}$$

Therefore, the pattern in Figure 4.11 cannot fold flat, which the previous exercise established via the Local-Min Theorem (Theorem 4.3).

Answer to Exercise 4.5. Indeed the Kawasaki-Justin Theorem (Theorem 4.4) is satisfied, and so it must be possible for those creases to fold flat. But not necessarily with any given mountain-valley folding. And we have seen those in Figure 4.6 fail. Theorem 4.4 only guarantees that there is *some* mountain-valley labeling of those creases that folds flat, and indeed there is, as shown by the first example of the chapter, Figure 4.3, labeling three as mountain folds and one as a valley fold.

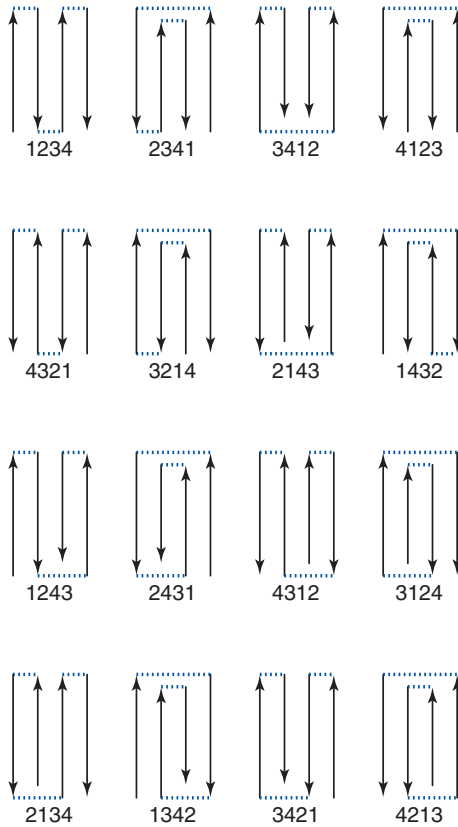


Figure Ans.15. The 16 different permutations of 1234 achievable by folding a stack of labeled stamps. (Based on <http://theory.cs.uvic.ca/inf/perm/StampFolding.html>)

Answer to Exercise 4.6. 16 of the 24 permutations of 1234 are achievable. They are listed in Figure Ans.15, with the stack top to the left, and the convention that the stamp label is to the left of the stamp's arrow. The stamp-1 arrow points upward in each diagram by the stipulated counting convention. As an example of what's missing here, the permutation 1324 would bury the 3-stamp between the 1- and 2-stamps without any way to connect to stamp 4.

There is no formula known for the number of permutations achievable in this fashion for a strip of n stamps, although the number has been calculated with computers. For example, for $n = 20$, more than 2 billion permutations are achievable: 2,050,228,360 to be precise. However, the number of possible permutations of 20 numbers is much, much larger: $20! \approx 2 \times 10^{18}$, compared to 2×10^9 —a billion times larger!

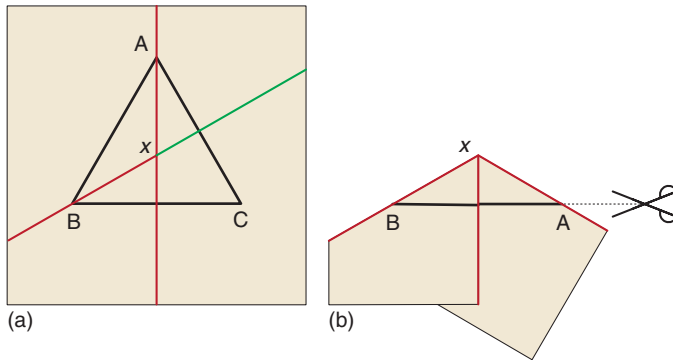


Figure Ans.16. Folding an equilateral triangle for one-cut.

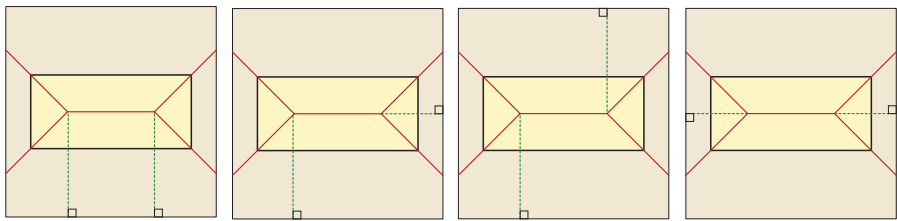


Figure Ans.17. Different ways to draw perpendiculars to fold and one-cut a rectangle.

Chapter 5

Answer to Exercise 5.1. Mountain-fold the sheet of paper in half through the apex A of the triangle, placing right corner C underneath left corner B . Now fold to bisect the angle at B ; see Figure Ans.16(a). These foldings align all three edges below the vertex x , the centroid of the triangle (b), and one straight cut through those aligned edges cuts out precisely the original triangle.

Answer to Exercise 5.2. There are four distinct ways do draw in perpendiculars, as illustrated in Figure Ans.17. All the other possibilities (e.g., two perpendiculars straight up), are symmetric with one of these (e.g., two perpendiculars straight down). They all lead to a one-cut fold similar to Figure 5.5(b). These are not the only ways fold a rectangle for fold and one-cut, incidentally, as the next exercise demonstrates.

Answer to Exercise 5.3. Yes, the rectangle crimp method works – try it! This is the typical mathematician’s strategy of reducing a new problem to a previously solved problem: in this case, rectangle to square. However, one must be cautious here, because the method does not work on a longer rectangle, such as that illustrated in Figure Ans.18. Although the crimp produces a square from the top

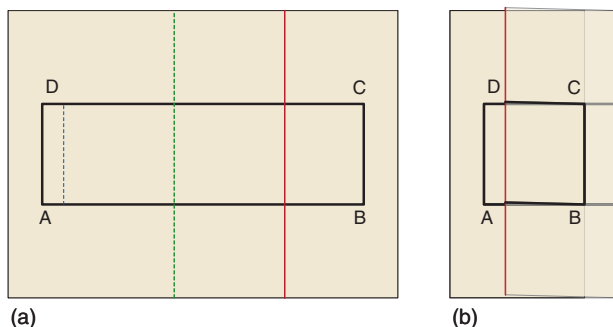


Figure Ans.18. A rectangle more than twice as wide as it is high fails fold and one-cut via crimping.

view, underneath there remains a central portion of the horizontal sides that are not aligned with other edges, as indicated in (b). Now cutting out the square as before will cut out two squares rather than a rectangle, leaving behind what's left over in the middle of the rectangle. The problem can be surmounted by crimping twice. I don't pursue this idea further in the text because it seems too specialized a method to lead to the general fold and one-cut theorem.

Answer to Exercise 5.4. Counting mountains and valley creases entering each vertex (including perpendiculars) verifies that M and V always differ by 2. Twenty-three of the vertices have degree 4, so there are either three mountains or three valleys incident to each of these. One vertex (in the head of the turtle) has degree 6, $M = 4$ and $V = 2$.

Answer to Exercise 5.5. One way to flatten a cube is shown in Figure Ans.19(a). The top and bottom faces are uncreased. The left and right faces get dented in half with valley folds. The front face gets six creases meeting at the face center x , which becomes a vertex in the flat folding (b); the back face is creased in an identical pattern meeting at y . (Note that the Maekawa-Justin Theorem [Theorem 4.2] is satisfied at x and y .) The cube squashes vertically.

Chapter 6

Answer to Exercise 6.1. Each interior vertex has either three mountain creases and one valley crease incident to it, or the reverse: three valley and one mountain. So $M - V = 3 - 1 = 2$ or $V - M = 3 - 1 = 2$, and the Maekawa-Justin Theorem is verified. Note that this theorem only applies to interior vertices, not to points where creases reach the paper boundary.

Answer to Exercise 6.2. The reader may recognize this from Exercise 4.1 and Figure Ans.13. It folds to the shape indicated in Figure Ans.20, with all eight isosceles right triangles sharing a vertical edge.

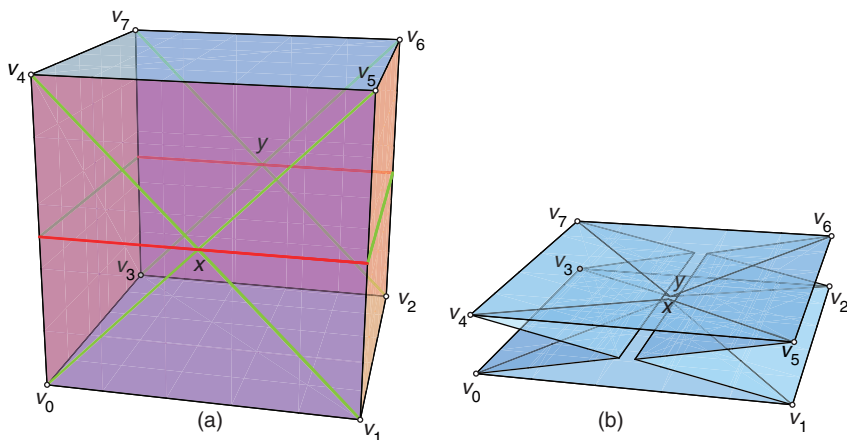


Figure Ans.19. Flattening a cube.

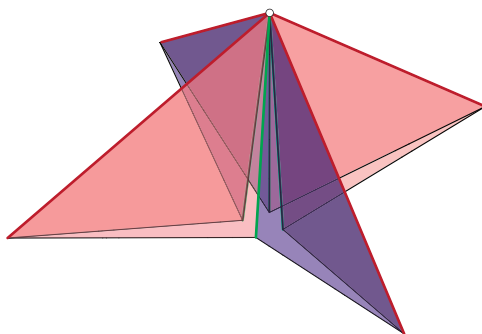


Figure Ans.20. A rigid folding of Figure 6.6.

Answer to Exercise 6.3. (a) Cube faces meet at right angles, so the dihedral angle between adjacent faces is 90° . (b) Each pair of faces of a regular tetrahedron share one edge, at which they form the same dihedral angle: $\approx 70.5^\circ$. This is the angle whose cosine is $\frac{1}{3}$, that is, it is $\cos^{-1} \frac{1}{3}$. This can be derived as follows. If the length of each edge of the tetrahedron is 1, then the altitude of each face is $\frac{\sqrt{3}}{2}$. The centroid of an equilateral triangle face falls at $\frac{1}{3}$ of the altitude. So the triangle illustrated in Figure Ans.21 has base $\frac{\sqrt{3}}{6}$ and hypotenuse $\frac{\sqrt{3}}{2}$. So $\cos \theta = \frac{1}{3}$. And θ is the dihedral angle.

Answer to Exercise 6.4. No, labels 1 and 4 cannot end up diagonally opposite. No matter how the rectangle in Figure 6.15 is hinged to a square, one can read off the labels 1234 clockwise, because the hinges are never broken so the squares must retain their original order. See Figure Ans.22. With such an ordering, the

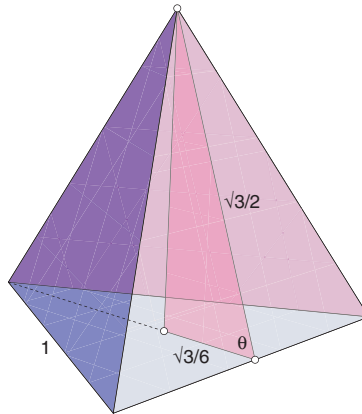


Figure Ans.21. Construction shows that the dihedral angle θ satisfies $\cos\theta = \frac{1}{3}$.

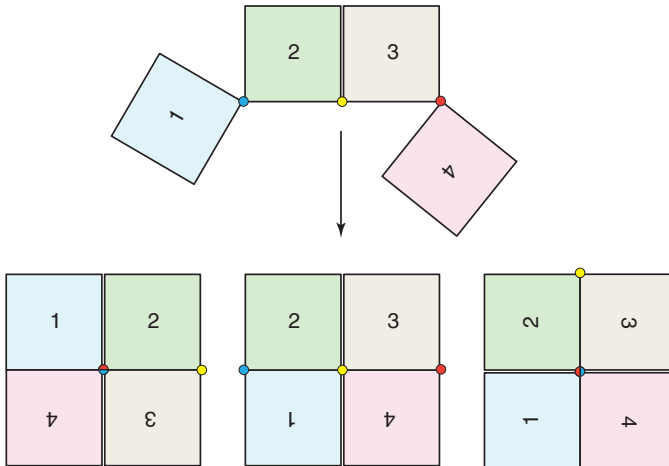


Figure Ans.22. Hinging of the 4×1 rectangle in Figure 6.15 to a square. One intermediate position is shown.

odd numbers 1 and 3 are diagonally opposite, and the even numbers 2 and 4 are diagonally opposite.

Answer to Exercise 6.5. It is clear from Figure 6.16(b) and (f) that the base is $1 \times \sqrt[3]{2}$. In order to have volume 2, the height must be $(\sqrt[3]{2})^2$. Using exponents to represent the roots, its dimensions are:

$$1 \times 2^{\frac{1}{3}} \times 2^{\frac{2}{3}},$$

roughly $1 \times 1.26 \times 1.59$.

Chapter 7

Answer to Exercise 7.1. Although the polyhedron has the same structure as a cube with two diagonally opposite vertices truncated by parallel planes – producing the visible top triangular face and another triangular bottom face on which the polyhedron is resting – the consensus of scholars (yes, this is the focus of scholarly debate!) runs against the truncated cube hypothesis. It appears that the faces prior to truncation are not quite squares as they would be for a cube, but rather *rhombi*, shapes all of whose four sides have the same length but not necessarily meeting at right angles. So the polyhedron is a “truncated rhombohedron.” There is less agreement on the rhombus angle marked in Figure Ans.23, with estimates ranging from 72° to 82° .

Answer to Exercise 7.2. Yes, the up-and-down path shown in Figure Ans.24(a) cuts open the cube to the net in (b). Such a single-path cut is not always available,

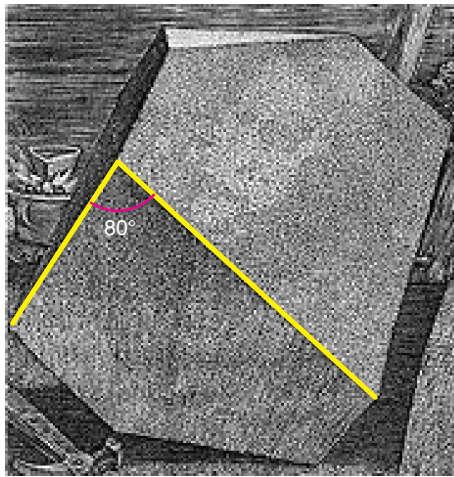


Figure Ans.23. The angle measured on the 2D drawing itself is approximately 80° .

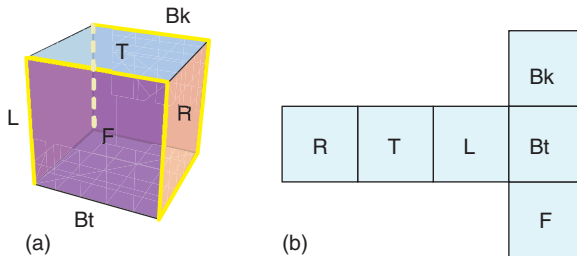


Figure Ans.24. Net for a cube produced by a single cut path.

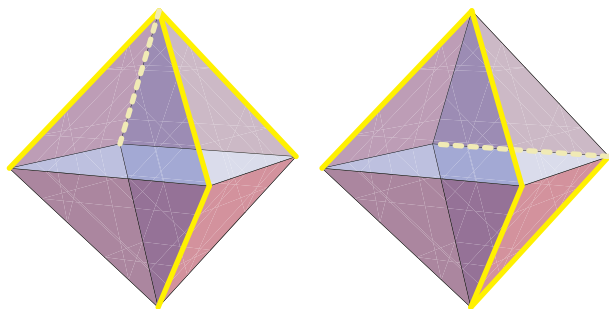


Figure Ans.25. Every spanning tree for an octahedron uses five edges.

because some polyhedra have no path of edges that touch each vertex once. Such a path is known as a *Hamiltonian path* in the literature. For the cube there are three different Hamiltonian path nets. You'll find the others if you work on Exercise 7.4.

Answer to Exercise 7.3. This is a bit of a trick question, because every spanning tree of an octahedron uses five edges. Figure Ans.25 shows two. In fact, every spanning tree of any polyhedron has the same number of edges! If the polyhedron has n vertices, every spanning tree has $n-1$ edges, because every tree of n vertices has $n-1$ edges. The octahedron has six vertices, and so all of its spanning trees have five edges.

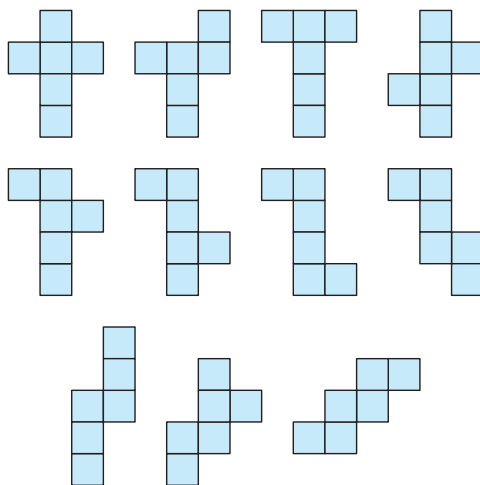


Figure Ans.26. The 11 distinct nets for a cube.

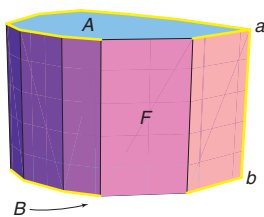


Figure Ans.27. An alternate unfolding of a right prism.

Answer to Exercise 7.4. There are 11 different (incongruent) nets for a cube, as listed in Figure Ans.26. The challenge here is make a complete list without missing any nets, or listing any more than once.

Answer to Exercise 7.5. At least as natural as the petal unfolding, which cuts between and separates all the vertical, lateral faces, is to keep all the lateral faces of the prism together. This can be accomplished by selecting some vertical edge ab to cut (see Figure Ans.27), and then cutting all but one edge of the top A and all but one edge of the bottom B . One could select the top and bottom uncut edges to be edges of the same face F as illustrated. Now all the side faces unfold to a rectangular strip bounded on the right and left by the two sides of the ab cut, with A and B attached above and below the strip at face F .

Answer to Exercise 7.6. Only the path from x to a is unique, along the diagonal of the top face. Both b and c have two shortest paths from x , passing to either side of the corner a . See Figure Ans.28. The paths may be constructed by unfolding the front and right faces attached to the top, and drawing straight lines from x to the points b and c .

Answer to Exercise 7.7. There is no “answer” here, but I have folded this myself, and it indeed works as described. It is very difficult to see the polyhedral structure in the vicinity of x , where 11 “images” of x come together, but the more

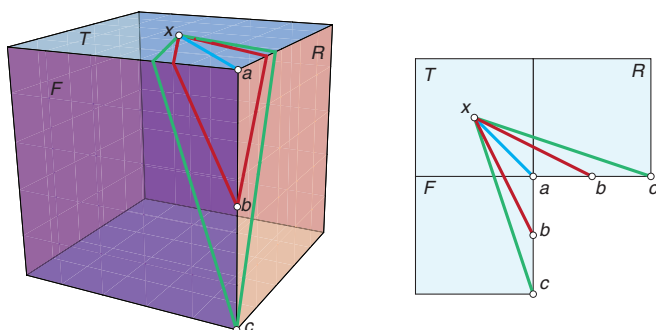


Figure Ans.28. Shortest paths from x to a , b , and c . T , F , and R indicate Top, Front, and Right faces respectively.

prominent edges on the back in Figure 7.22(b) (connecting the vertices 6, 7, 8, 9, 10, 11) do emerge with some nudging.

Chapter 8

Answer to Exercise 8.1. We need to count the rectangles visible in Figure 8.3. There are 10 blue top rectangles, because the example terrain base is 10×10 . The heights of the orange left and right side rectangles can be determined by taking the difference in adjacent tower heights. This yields this sum for the side rectangles:

$$1 + 2 + 2 + 3 + 1 + 3 + 2 = 14$$

so indeed the total length is $10 + 14 = 24$, which can be verified in the final unfolding Figure 8.5(a).

Answer to Exercise 8.2. Figure Ans.29 shows the unfolding. The bridge between strips is of height 2, a tie between the towers of heights 1 and 3, and the towers of heights 2 and 4.

Answer to Exercise 8.3. Figure Ans.30 shows the unfolding for $n=10$. Because all strips are identical, all bridges are “empty bridges.” So all x -strips unfold directly abutting against one another.

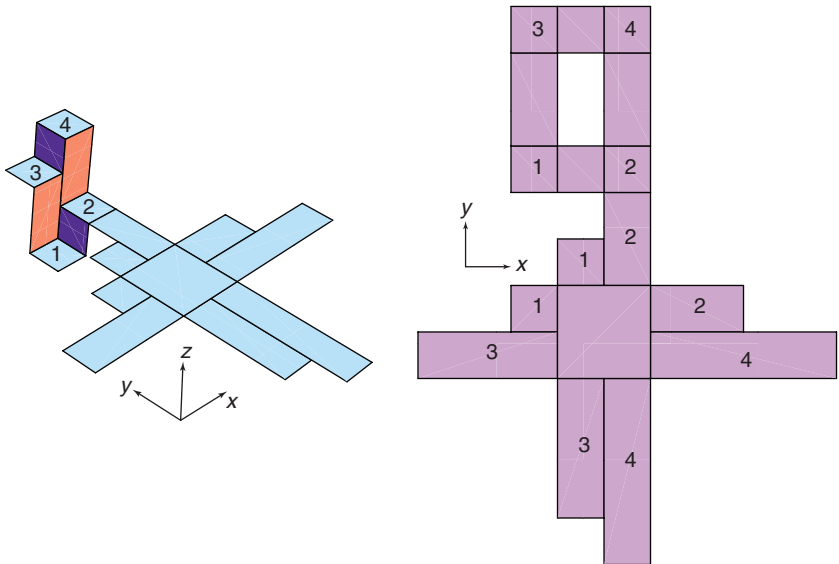


Figure Ans.29. Unfolding of 2×2 terrain. Several heights are marked.

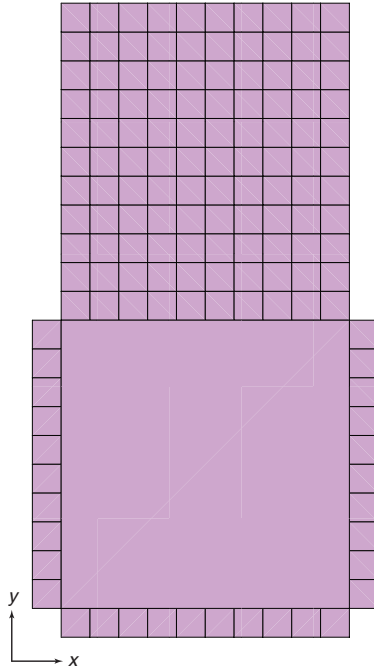


Figure Ans.30. Unfolding of a $10 \times 10 \times 1$ box.

Answer to Exercise 8.4. There are many ways to unfold this torus, one of which is shown in Figure Ans.31. The main challenge is to find spots for attaching the inner ring of faces a, b, c, d .

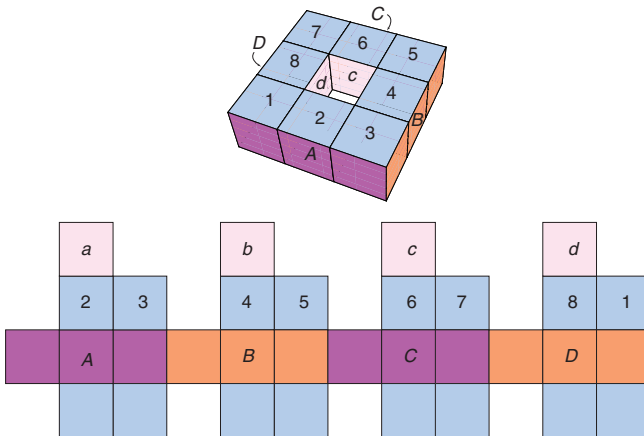


Figure Ans.31. Unfolding of an orthogonal torus. Selected faces are labeled. The inner ring of faces are a and b (obscured), and c and d .

Chapter 9

Answer to Exercise 9.1. No, Figure 9.2 cannot fold to a cube, as you already knew if you worked through Exercise 7.4: It does not appear in that list of 11 nets in Figure Ans.26. Attempts with paper result in two faces overlapping, leaving one cube face as a hole. However, that polygon can fold to several other polyhedra by creasing differently, including polyhedra with five, six, and eight faces. Warning: These are very difficult to find! The curious may look ahead to Exercise 9.3.

Answer to Exercise 9.2. It folds to a cube, as shown in Figure Ans.32.

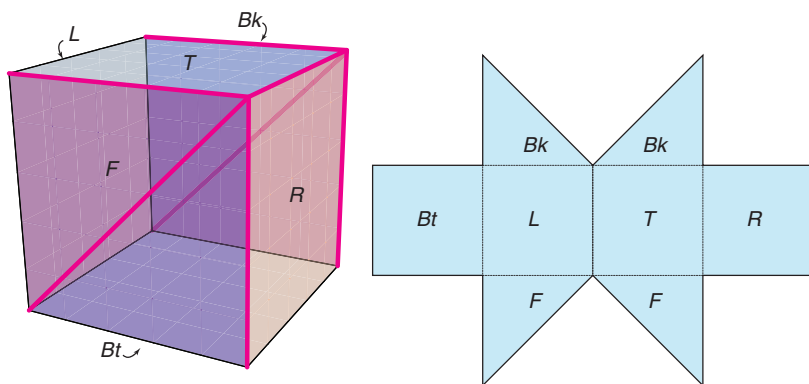


Figure Ans.32. The bow-tie polygon folds to a cube.

Answer to Exercise 9.3. There are several ways to achieve an Alexandrov gluing, none straightforward. Perhaps the easiest to see is the gluing shown in Figure Ans.33. The two problematic reflex corners, labeled 4 and 8 in the figure, are “zipped” closed, by gluing together the two corners labeled 1, and the two

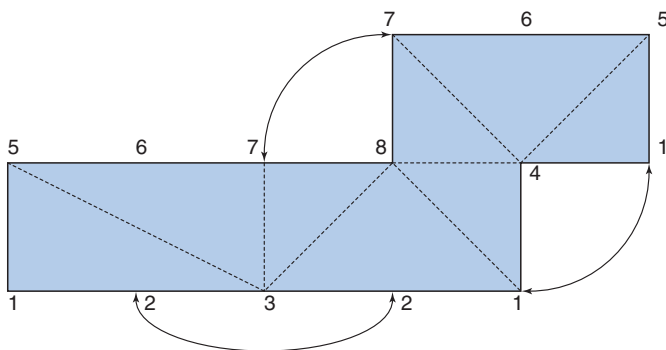


Figure Ans.33. An Alexandrov gluing of Figure 9.2. The dashed edges are creases of the resulting polyhedron.

points labeled 7. This ensures that the final angles glued in at 4 and 8 are each 270° , satisfying criterion (b) that the angle must be at most 360° . If one then creases at point 3, the remaining identifications shown are forced, using up the perimeter as per criterion (a). Cutting this out of paper and taping the identified boundary sections indeed results in a type of bag, criterion (c). If you crease along the dashed lines, you will see an octahedron emerge, a polyhedron with eight triangular faces, and six vertices at points 1, 3, 4, 5, 7, and 8.

Acknowledgments

Figure Permissions

Figure 1.1(a) (Robot Arm)	<i>Machines Automation Robotic Systms Ltd.</i> , by permission.
Figure 2.11 (Pantograph)	<i>Gravograph Ltd.</i> , by permission.
Figure 2.12 (Jefferson Polygraph)	<i>Special Collections, University of Virginia Library</i> , by permission.
Figure 2.14 (The Turk)	<i>American Heritage Publishing</i> , by permission.
Figure 3.10 (Pop-up Spinner)	<i>Akira Nishihara</i> , by permission.
Figure 4.13 (Oval Tessellation)	<i>Robert Lang's Oval Tessellation, 1999</i> , by permission.
Figure 5.15 (Elegant Head)	<i>Chris Perry and Mark Cheng</i> , by permission.
Figure 6.5 (Miura Maps)	<i>Photo by Erik Demaine</i> , by permission.
Figure 6.4 (Miura Animation)	<i>Based on a Mathematica animation by Thomas Hull</i> , by permission.
Figure 6.7 (Square Twist)	<i>Based on a Mathematica animation by Thomas Hull</i> , by permission.

Acknowledgments

I used three wonderful pieces of software to create figures: *Cinderella* for the linkage loci in Figures 2.2, 2.3, and 2.16, *Mathematica* for the 3D figures throughout, and *Adobe Illustrator* for everything else.

I thank Ryuhei Uehara for showing me the Pop-up Spinner and tracking down information on it. Lara Ramsey and Theodora Nikolva are great grade-school teachers who worked with me on making pantographs accessible to their students. I learned a great deal working with Nadia Benbernou on the material in Chapter 3, and I thank her for reviewing that chapter. I thank Thomas Hull for permission to use his rigid origami animations, and for answering my questions. The linkage to sign ‘J’ (Figure 2.16) is based on a construction of Don Shimamoto, who kindly allowed me to include it. I thank Vera Sacristan and Eduard Llamas for correcting an error in Section 1.2. I thank Erik Demaine for letting me draw heavily from *Geometric Folding Algorithms*.

Brad Ballinger read the whole first draft and gave me great advice. I owe special thanks to Joseph Malkevitch, who also read the entire manuscript and gave sage advice throughout, and to my editor at Cambridge University Press, Lauren Cowles, whose “blue pencil” improved the book immeasurably.

Index

- Alexandrov's theorem, 118, 133–135, 140
- algorithm, 140
 - complexity, 149
 - map folding, 70
 - polygon folding, 133, 140
 - unfolding orthogonal polyhedron, 145
- amino acid, 40–42, 149
- angle bisector, 77
- annulus, 5, 14, 147
- Archimedean solids, 101, 104, 145, 147, 150
- array, 120

- chain
 - alignment, 44, 45
 - locked, 53, 54
 - near-unit, 53, 54, 158
 - piercing, 46, 143
 - unit 90° , 42, 44, 47, 51, 54
- configuration, 147
 - degenerate, 148
 - locked, 53, 158
 - maxspan, 42–44
 - piercing, 46
 - of robot arm, 8
 - of shopping bag, 94, 147
 - space, 53, 147
 - staircase, 43, 44
- convex hull, 112, 114
- convexity, 105

- degree
 - angle measure, 147
 - number of incident creases, 61, 89, 148
- degrees of freedom, 22, 148
- dihedral angle, 39–41, 89, 90, 148, 163
- dissection, 96, 148
 - hinged, 96–98, 145

- edge
 - of drawing, 73
 - of polyhedron, 103
 - reflex, 106
 - unfolding, *see* unfolding, edge
- ellipse, 38

- face
 - of polyhedron, 81, 83, 100, 103, 149
 - in rigid origami, 84
- folding
 - flat, 57, 58
 - a map, 69, 70
 - stamps, 71, 160
- Folding@Home, 52

- induction hypothesis, 7

- joint, 8, 148, 150
 - angle, 15, 17, 18
 - fixed-angle, 39
 - kinked, 18
 - pinned, 25, 148
 - universal, 4, 39

- Kawasaki-Justin theorem, 66, 68
- Kempe's universality theorem, 36, 143
- lemniscate, 25
- line segment, 3
- link, rigid, 2, 3, 148
- linkage, 2, 24, 148
 - angle-limited, 23, 154, 155
 - Chebyshev, 27
 - pantograph, 2, 28, 30, 32, 33, 35, 149
 - Peaucellier-Lipkin, 27, 28
 - planar, 36, 37, 43
 - signing, 36, 37
 - Watt, 25, 26, 155, 156
 - wiper, 29
- locus, 15
- Maekawa-Justin theorem, 61, 63
- map, 69, 70
- mathematical model, 3
- matrix, 120
- maxspan, 42, 143
- Melencolia I, 101, 102, 106, 165
- Miura map fold, 85–87, 144
- necessary and sufficient, 65, 69, 148
- net, 101, 106, 115, 126, 148, 149
 - for cube, 112, 116
 - general, 115, 117, 119, 125, 126, 129, 149
 - Latin cross, 109, 130, 131, 138, 139
- NP-complete, 21, 68, 142, 149
- NP-hard, 68, 149
- one-cut, fold and, 72, 78, 81
- open problem
 - Planar Signing (General Case), 37
 - Planar Signing Digits, 37
 - Dürer's Problem, 106
 - Edge-Unfolding Prisms, 115
 - Unfolding Manhattan Towers, 126
 - General Unfolding, 129
 - Locked Unit 90° -Chain, 54
 - Map Folding, 70
 - Flattening Polyhedra, 83
 - Folding to Convex Polyhedra, 139
- origami fold, 57–61, 67, 70, 76, 149
 - mountain/valley fold, 50, 57, 149
- pantograph, *see* linkages, pantograph
- Platonic solids, 104, 105, 149, 150
- polygon, 63, 103, 130, 140, 149
 - convex, 105, 112, 135
 - doubly covered, 134
 - regular, 149, 150
 - unfoldable, 132
- polygonal chain, *see* chain
- polygraph, 33, 34
- polyhedron, 81–83, 103, 104, 141, 149
 - convex, 105, 106, 108, 134, 136
 - Dürer's, 101, 106
 - deltahedron, 114
 - dome, 113
 - flattening, 81–83
 - Manhattan tower, 125, 126
 - orthogonal, 119, 126
 - orthogonal terrain, 120, 121, 125
 - prism, 112, 113, 115
 - prismatoid, 114, 115
 - prismoid, 113
 - regular, 104, 149, 150
 - semiregular, 147, 150
 - skeleton, 109, 110
- pop-up spinner, 48, 49
- proof, xi
 - constructive, 36, 140, 141
 - existence, 134, 140
 - induction, 4, 6, 46
 - sketch, xi
- protein
 - backbone, 23, 40–42, 149
 - folding problem, 41, 52
 - villin headpiece, 41, 53
- PSPACE-complete, 22, 149
- reachability, 22
 - angles, 15, 16, 18
 - region, 4, 5, 8, 14
- rhombus, 27, 150, 165
- rigid origami, 84, 85, 88, 89, 144
- robot arm, 3, 4
- ruler folding, 20, 21
- shortest path, 115–117, 167
- skeleton
 - of polyhedron, 109
 - straight, 79, 80

- span, maximum, 42, 143
- spherical shell, 5, 19, 147
- square twist, 88, 89

- torus, 103, 129, 169
- tree, 110, 150, 166
 - spanning, 110–112, 166
- triangle inequality, 12, 13, 45, 150, 151
- truncation, 106, 150
- Turk, The, 33, 35, 143

- unfolding
 - edge, 101, 106, 126, 128, 148, 149
 - general, 115, 119, 129, 148
 - grid, 125, 126
 - overlapping, 107, 108
 - petal, 113, 114
 - star, 116, 117
- unsolved problem, *see* open problem

- vector, 9, 30, 62
 - addition commutative, 8, 11, 147
- vertex, 150
 - of polygonal chain, 8
 - convex, 105, 135
 - origami, 58–60, 147
 - of polyhedron, 103, 132
 - reflex, 105, 132

