Easy Pieces in Geometry

Thomas C. Hales

October 22, 2007

Figure 1.

Contents

Preface

This book is a collection of some elementary results in geometry. The results are elementary in the sense that this book is almost entirely self-contained, and does not require require anything beyond basic concepts at an undergraduate level. Many of the pieces are accessible to a high-school student with an interest in mathematics. I have even tried to avoid calculus wherever possible: only a few of the easy pieces require any knowledge of integration. The most difficult fact we use about integration is the formula for the area of a surface of revolution.

Although the results may be elementary, many results in this collection have fundamental significance. Anyone who wishes to possess a certain level of cultural literacy in mathematics might find it beneficial to commit proofs of these results to heart. There is an easy proof of the Jordan Curve theorem, the Jordan-Schönflies theorem, of the isoperimetric inequality, the Brouwer fixed point theorem, the triangulation theorem for surfaces, and the main part of the classification theorem for surfaces. This book differs from many textbooks that cover these topics in that it gives complete elementary proofs of these results.

Many books that reach students today feed students a disfigured form of these results. For example, books claim to prove the the classification of surfaces, but omit the proof of the core part, which is the existence of a triangulation. Or a book claim to prove Euler's theorem $V - E + F = 2$, but make covert use of the Jordan Curve theorem, which is never proved. A survey of textbook coverage of the isoperimetric inequality reveals that disappointingly many of them lack a complete proof. This book aims to prove everything that is claimed and to do it in an easy way.

This book was used in a course at the University of Pittsburgh as a supplementary text that accompanied a course in Discrete Geometry. I presented an easy piece to the students at the beginning of every lecture. I asked the students to learn the statements and proofs of any thirty of the pieces. The students demonstrated their mastery of the pieces by practicing presenting the proofs to other students without the use of notes, and then by an oral examination at the end of the semester. This book might be used in a similar way as a supplementary text in courses in various courses in Geometry or Topology.

In recent years, I have become involved in the formalization of mathematics. There, the aim is to convince a computer of the correctness of mathematical proofs. This technology is still under development, and it requires a great deal of effort to transcribe even simple proofs into machine checkable form.

I am currently supported by an NSF grant on the "Formal Foundations of Discrete Geometry." As part of that project, I implemented the first formal proof of Jordan Curve Theorem. This is a proof of the Jordan Curve theorem that can be executed on a computer. It consists of some 50,000 lines of computer code. The fully expanded proof of the theorem (at the level of the primitive axioms and rules of inference of mathematics) consists of an estimated 80 million logical inferences.

With large projects such as these, at first I thought that the formal theorem proving community would be mired down in heavy technicalities. Just the opposite turns out to be the case. Anyone who spends time in this area begins to look for elegant proofs – things that even a computer can understand. My colleagues in

the formal theorem proving community take particular delight in finding ever more elegant proofs of theorems. An excellent example is Gonthier's formal proof of the Four Color Theorem. Not only did Gonthier formalize the proof; in the process of doing so, he made the proof far more comprehensible than ever before.

This book reflects how I have been overcome by the enthusiasm of this community for easy pieces. Freek Wiedijk maintains a web page on the formalization of "100 theorems" in mathematics. This book might be viewed as a supplement to that list of 100 theorems.

Another aim has been to present visual proofs that do not sacrifice rigor. This book has an extraordinary number of color illustrations. The proofs are often based on these illustrations; the text is secondary. Although these results are fully rigorous, I have maintained an informal style. Definitions and notation have been omitted, or have been moved to the glossary, when they can be inferred from the pictures.

Above all, I have intended for the proofs to be enjoyable. Please enjoy!

I would like to thank Thanh Phuoc, David Green, Alexander Borisov, Chris Jones, Bojana Pejic, Johnny Kwong, and all the students of my Discrete Geometry course for helping this project come to fruition.

Thomas Hales

Chapter 1

The Game of Hex

Problems worthy of attack, prove their worth by hitting back – a Grook of Piet Hein

The four-color problem asks whether it is possible to color every map with four colors in such a way that countries that share a border have different colors? This problem was not resolved until 1976 (See Chapter XX.)

During the Nazi occupation of Denmark, Piet Hein invented the board game Hex, while working on the four-color problem. Piet Hein drew four countries along the sides of a diamond and a fifth country surrounding the four. He wondered if it might be possible to find a way to extend the four countries into the diamond in such a way that the the countries on opposites sides share a border. If such a map could be found, all five countries would share borders with one another, and five colors would be required to color the map. This would give a counterexample to the famous problem.

Why not turn the search for a counterexample of a famous problem into a board game? A diamond-shaped country is invaded by two opposing players who take turns enlarging their territories. The victor is the first player to connect their divided homelands at opposite ends of the board.

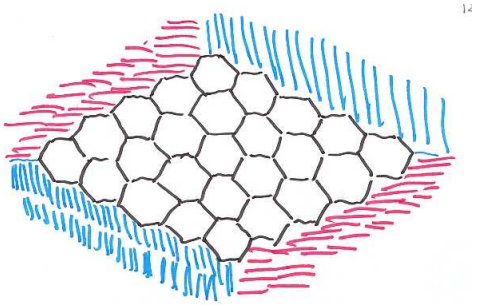


Figure 1.1. *A hex board*

The rules of the game are simple. A hex board is a diamond shaped grid of hexagons surrounded by countries of blue and red. There are two players, red and blue, who take turns placing hexagonal (red and blue) tokens on the hex board. The player wins who first builds an unbroken bridge between the two oceans of that player's color. We start out our collection of easy pieces with a theorem about Hex.

Theorem 1.1. *'There is no stalemate in Hex: by the time the board is filled (even if red and blue play unequal numbers of tokens), either red or blue builds a winning bridge.*

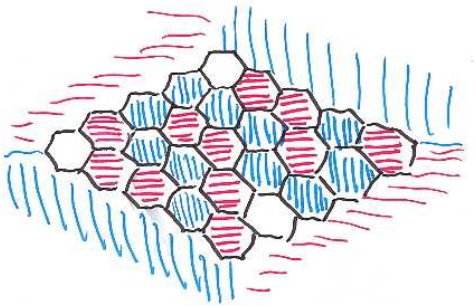


Figure 1.2. *Blue wins*

Proof. Start at point A in a filled hex board and follow the 'fault line' (*interface* is the mathematical term) between the blue and red (shown in black). The fault line exits the hex board at B, C, or D. (In fact, the exit is never at C. The fault line carries blue to its left and red to its right, but C carries red to its left, a 'parity violation'.) The tokens along the fault line form a winning bridge for blue if the exit is at B and for red if the exit is at D. □

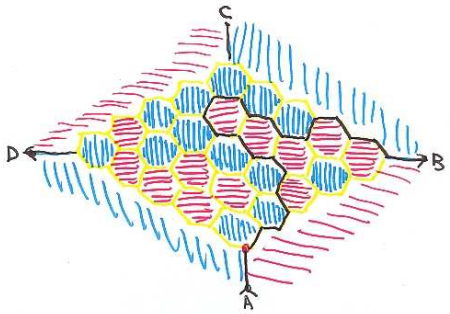


Figure 1.3. *The interface between red and blue*

References:

Thomas Maarup, Hex: everything you always wanted to know about hex but were afraid to ask, thesis, 2005. http://maarup.net/thomas/hex/hex3.pdf

Chapter 2

The Brouwer Fixed Point Theorem

The Brouwer Fixed Point Theorem is one of the most striking theorems in mathematics. Take two crisp dollar bills. Crumble one of them into a wad and set it on the crisp bill. Brouwer found that it is always possible to drive a nail into the crumbled bill and through the bill underneath in such a way that when the crumbled bill is opened again, the nail hole of the crisp bill is perfectly aligned with a hole in the other bill.

After proving his fixed-point theorem, Brouwer went on to found the intuitionist school of mathematics. Brouwer later renounced his fixed-point theorem when he was unable to give an intuistionist proof. His renunciation was unnecessary, because intuitionist proofs have since been found.

The Brouwer Fixed Point theorem holds not only for two-dimensional surfaces such as a dollar bill, but also in higher dimensions. This general theorem is one of the most fundamental theorems in mathematics. Arrow and Debreu were awarded the Nobel Prize in Economics for their proof of the existence of equilibrium market prices, by showing that equilibrium prices are the Brouwer fixed point of a nonlinear system of equations. In work that garnered another Nobel Prize, John Nash found that Brouwer's fixed point can be used to show that a game has a Nash equilibrium.

In the following theorem, $[0,1]^2$ represents the dollar bill and f maps each point of the wadded bill to the point of the crisp bill directly under it. The proof of this amazing dollar parlor trick follows directly from the game of Hex.

Theorem 2.1. *A continuous function $f : [0,1]^2 \to [0,1]^2$ has a fixed point $f(x,y) = (x,y)$.*

Proof. For a contradiction, suppose $f : [0,1]^2 \to [0,1]^2$ has no fixed point. If we had arbitrarily close approximations to a fixed point, we could find a suitable sequence of approximations converging to a fixed point. Thus, we can assume that $f(x,y)$ is bounded away from (x,y). If we write this out in detail, either the x-coordinate of

Figure 2.1. *Brouwer illustrated*

$f(x, y)$ is too small or too large by some $\epsilon > 0$:

$$(R_1): \ f_1(x,y) - x > \epsilon \quad \text{or} \quad (R_2): \ f_1(x,y) - x < -\epsilon \tag{2.1}$$

or the y-coordinate of $f(x, y)$ is too small or too large by some epsilon.

$$(B_1): \ f_2(x,y) - y > \epsilon \quad \text{or} \quad (B_2): \ f_2(x,y) - y < -\epsilon \tag{2.2}$$

That is, $f(z)$ always strays to the right, left, above, or below the source z.

Label an $N + 1$ by $N + 1$ hex board with the rational numbers

$$x, y = 0/N, 1/N, \ldots, N/N$$

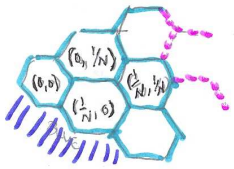


Figure 2.2. *Hex board labels*

The labels (x, y) of each hexagon satisfy Inequalities 2.1 or Inequalities 2.2. If N is sufficiently large, it follows by uniform continuity that a hexagon satisfying R_1 is not adjacent to a R_2 hexagon, and that B_1 hexagons are not adjacent to B_2 hexagons.

Play a game of hex, restricting red to play on R_1 and R_2 hexagons, and blue to play on B_1 and B_2 hexagons. (If one player runs out of places to move, the other player is allowed to fill the rest of the board.)

Blue cannot win. A blue bridge, if it exists, can be built entirely from B_1 hexagons or entirely from B_2 hexagons, since no step of the bridge can move back or forth from B_1 to B_2. A B_1 bridge can never reach the upper right: because the upper right does not meet the B_1 condition: $f(x, 1) - 1 > \epsilon$. Similarly, a B_2 bridge can never reaches the lower left.

Similarly, red cannot win either, and stalemate follows. We have found a contradiction with the GAME OF HEX. □

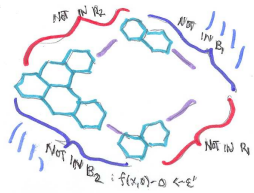


Figure 2.3. *Stalemate in Hex*

Brouwer's fixed point theorem holds for also implies that a continuous map $f : D \to D$ has no fixed point for any set D that is homeomorphic to the square D. In detail, we need to pick a homeomorphism $g : D \to [0,1]^2$, then apply Brouwer to the corresponding map $g \circ f \circ h \circ g^{-1}$ from the square to itself.

Chapter 3

Disk Retraction

Theorem 3.1. *The circle is not a retract of the disk it bounds. In other words, there does not exist a continuous function $f : D \to S^1$ from the unit disk to the unit circle that is the identity map on $S^1 \subset D$.*

Proof. Let $h : D \to D$ be the rotation by angle π. Suppose for a contradiction that f exists, then $f \circ h : D \to S^1 \subset D$ is continuous and fixed point. This is prohibited by (the follow-up comment to) Brouwer's Fixed Point Theorem. □

Chapter 4

Fashoda Meet Theorem

The following is a generalized version of the intermediate value theorem.

Theorem 4.1. *Let $D = [-1, 1]^2$ be a square. Assume $f(s)$, for $-1 \leq s \leq 1$ is a path in D from the left edge of the square to the right edge. Similarly, assume $g(t)$, for $-1 \leq t \leq 1$, is a path from the bottom edge to the top edge. Then for some s, t, the paths meet: $f(s) = g(t)$.*

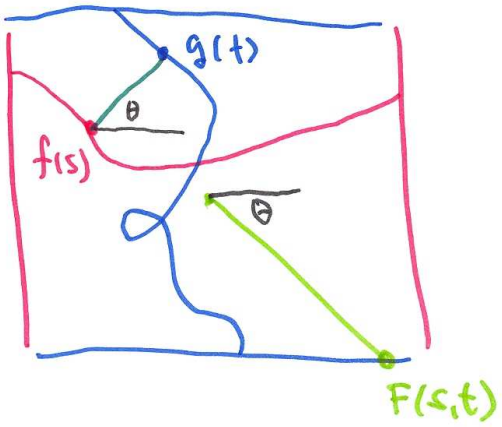


Figure 4.1. *Crossing paths*

Proof. Suppose for a contradiction that $f(s) \neq g(t)$, for all s, t. Let $F : D \to D$ by given by $F(s, t) = (a, -b)$, where (a, b) is the point on the boundary of D crossed by the ray from the center of the square in the direction $g(t) - f(s)$.

The function F is continuous. Let (s', t') be the Brouwer fixed point for F. It, like the image of F, lies on the boundary of the square.

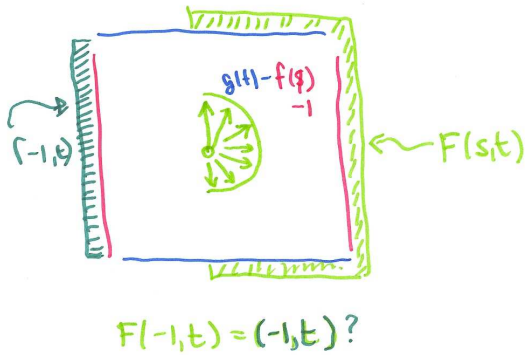


Figure 4.2. *No fixed points on the left edge*

The fixed point cannot be on the left edge $F(-1, t) = (-1, t)$, because the values of $F(-1, t)$ land in right half-plane (as the vector $g(t) - f(s)$ has non-negative first coordinate). By similar reasoning, the fixed point cannot land on the other edges either. □

Chapter 5

Hex Board Jordan Curves

A hex board Jordan curve is a simple closed curve that follows the edges of the hexagons in a hex board.

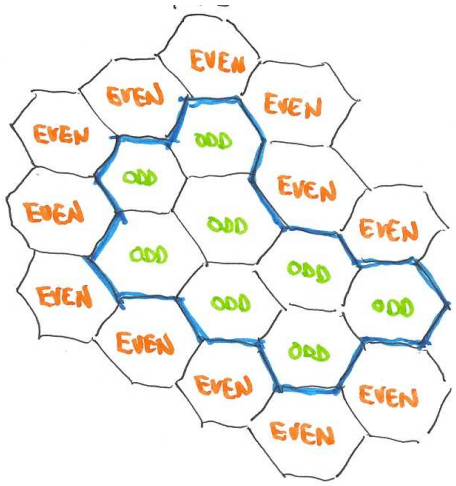


Figure 5.1. *Hex board Jordan curve*

Given a hex board Jordan curve, assign a parity to each hexagon by the parity of the number of horizontal edges on the Jordan curve below the hexagon.

Theorem 5.1. *A hex board Jordan curve separates the board accordiing to parity: the parity differs on adjacent hexagons exactly when the Jordan curve passes between the two hexagons.*

Proof. This is obvious for vertically stacked adjacent hexagons: precisely the

horizontal edges on the Jordan curve lie beneath both, unless the curve passes between the two hexagons.

Assume now that the adjacency between the two hexagons is not vertical. Assume that the Jordan curve does not pass between the hexagons A and B.

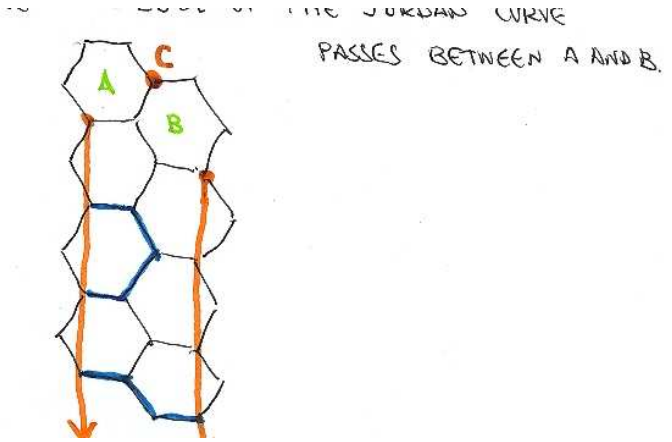


Figure 5.2. *Non-vertically stacked hexagons*

Each horizontal edge below A or B on the curve is part of a segment of the curve with both endpoints on the orange lines. A segment has two endpoints. The horizontal edges on the Jordan curve are in bijection with these endpoints, so the number is even. Thus, A and B have the same parity.

If an edge of the Jordan curve passes between A and B at C, then the segment of the Jordan curve at C has only one endpoint on the orange lines, reversing the parity.

□

Chapter 6

The Complement of an Arc

Theorem 6.1. *The complement of a simple arc in the plane is connected.*

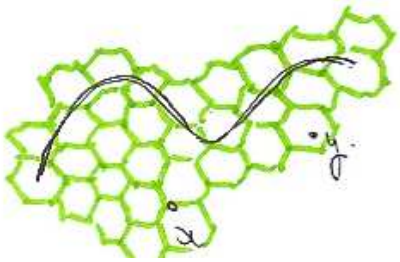


Figure 6.1. *Covering an arc with hexagons*

Proof. The simple arc is the continuous image of a 1-1 function $\gamma : [0, 1] \to \mathbb{R}^2$.

Pick x, y in the complement of the arc. Superimpose a hex board on this picture. Assume that the closed hexagons are small enough that if hexagons meeting the arc at $\gamma(s)$, $\gamma(s')$, with $s < s'$ and if both hexagons are adjacent to a third, then all the hexagons meeting $\gamma[s, s']$ stay far from x and y in the sense that there is a hexagonal path from x to y avoiding all these hexagons. (The existence of such small enough hexagons requires uniform continuity, etc.)

Let $\mathcal{B}$ (blue hexagons) be those meeting the simple arc. Enumerate them $B_1, \ldots, B_m$, ordered in the order encountered on the path γ (breaking ties arbitrarily when the path passes through the vertex of a hexagon).

Lemma 6.2. *For $k = 0, \ldots, m$, there is a (red) path of hexagons from x to y that avoids the blue hexagons $\{B_1, \ldots, B_k\}$.*

Proof. If $k = 0$, the blue set to avoid is empty, so any red hexagonal path will

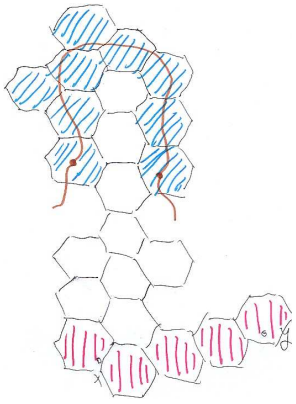


Figure 6.2. *Small hexagons*

work. Use this as the base case for induction. Assume a red hexagonal path exists for $\{B_1, \ldots, B_k\} = \mathcal{B}_k$. Take the shortest possible red hexagonal path P (measured by the number of hexagons). We construct a path avoiding $\mathcal{B}_{k+1}$.

If P avoids $\mathcal{B}_{k+1}$, we are done. If P can be rerouted through hexagons adjacent to $\mathcal{B}_{k+1}$, we are done. (If the path can be extended by rerouting as needed through adjacent hexagons, we say that the path follows a *simple detour*.)

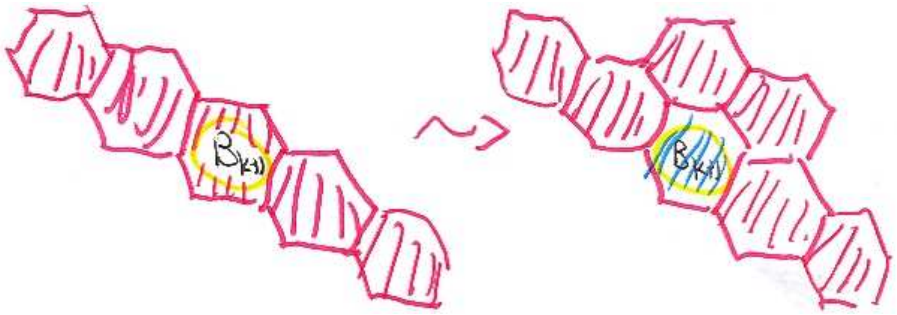


Figure 6.3. *Rerouting around* B_{k+1}

We assume that the route around B_{k+1} is blocked in both directions by blue hexagons in $\mathcal{B}_k$. Call these adjacent blocking hexagons A and C.

By the smallness hypothesis, and the given ordering on hexagons, there is an interval $[s, s']$ such that

$$\gamma[s, s'] \subset B_1 \cup \cdots \cup B_k,$$

with $\gamma(s) \in A$ and $\gamma(s') \in C$. We can take the shortest edge path Q (black) from A to C along edges of hexagons in $\mathcal{B}_k$. Then complete it to a hex board Jordan curve Q' by adding edges from $\{A, C, B_{k+1}\}$ (green). The red hexagon path crosses

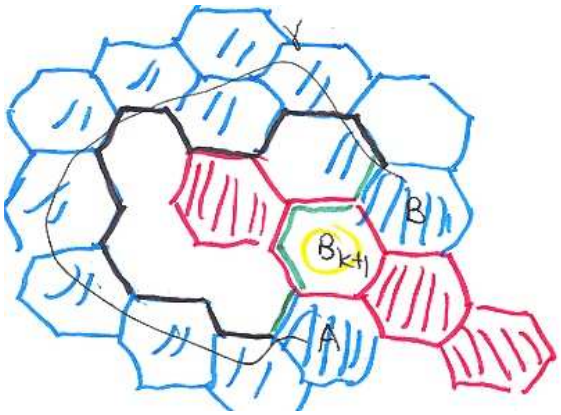


Figure 6.4. *Avoiding the trap at* B_{k+1}

the Jordan curve exactly once (at B_{k+1}) because it cannot cross along Q, as those edges are blocked by hexagons in $\mathcal{B}_k$. So x and y have opposite parity with respect to the Jordan curve Q'. This is contrary to $(*)$. We have completed the proof of the induction.

(By way of terminology, we call the Jordan curve Q' a *small trap*. What we have shown is that the shortest red hexagonal path avoids *small traps*.)

Thus, a red hexagonal path exists from x to y that avoids $\mathcal{B}_m$. Join x and y by a path in red that avoids all blue hexagons and that consequently avoids the simple arc. □

□

Chapter 7

Components

Chapter 8

Detours and Traps

Note: this chapter is not complete. Eventually, it will give a proof that a Jordan curve can be approximated by a polygon whose length is no greater than the Hausdorff length of the Jordan curve and whose enclosing area is at most ϵ more than the area enclosed by the Jordan curve. For now, we will assume this result as given.

The last chapter gives a proof that the complement of a simple arc in the plane is connected. The proof proceeds by superimposing a hex board grid over the plane, which is used to approximate paths. Two strategies are used in drawing a path between two points in the complement. The first is called *simple detours*, which involves rerouting a path of hexagons around a given hexagon through the adjacent hexagons. The second strategy is to avoid *small traps* by taking the shortest possible path (measured by the number of hexagons traversed) between two points. A small trap, if we continue with the traffic analogy, is a small *cul de sac* or dead end that leads nowhere.

Chapter 9

The Jordan Curve Theorem

Theorem 9.1. *The complement of a simple closed curve in the plane consists of exactly two connected components.*

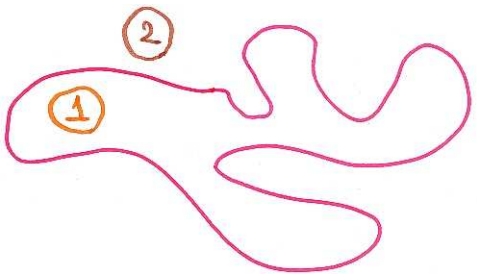


Figure 9.1. *A Jordan Curve*

Proof. We make use of basic topological properties such as the fact that each connected component is open an that the connected components are the same as path components. (Appendix to be added about this.)

Let A and B be points on the simple closed curve J at maximal distance from one another. For ease of illustration we assume that the coordinate system has been chosen so that A and B lie along the x axis. Draw a rectangle touching J only at A and B with edges parallel to the coordinate axes.

J breaks into two simple arcs U (up) and D (down) with endpoints A and B. The vertical line through the center of the rectangle intersects both arcs U and D. U is chosen as the one first encountered moving from top to bottom along the central vertical line. We need a couple of facts before completing the proof.

(*) Every path in the rectangle from the top edge to the bottom edge meets both U and D (CROSSING PATHS).

(**) Every path in the rectangle from the top to the bottom of the rectangle meets every bounded connected component C in the complement of J. (Note: the

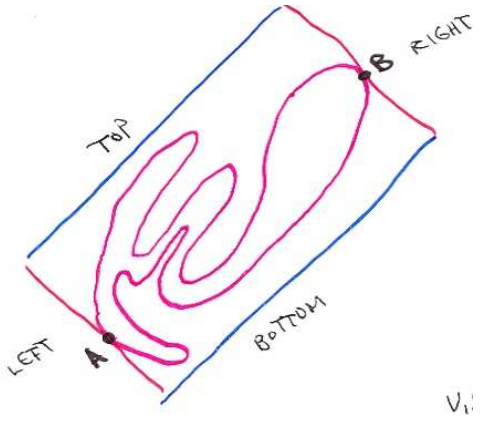


Figure 9.2. *A bounding rectangle*

exterior of the rectangle is contained in the unbounded component of the complement of J, and so C lies in the interior of the rectangle.)

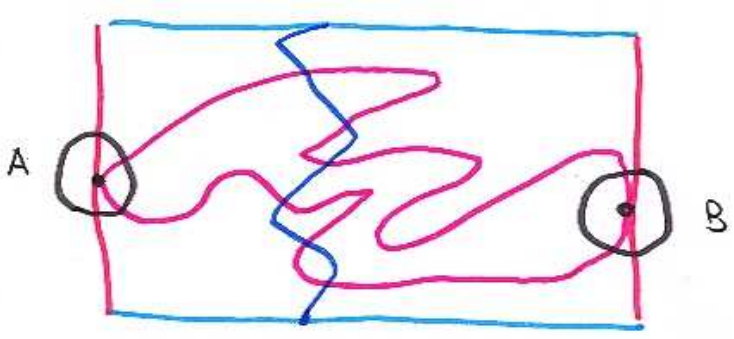


Figure 9.3. *A top-to-bottom path meets components.*

We prove (**). Assume for a contradiction that (**) is false. Pick a small disk D_A around A that that does not meet the path. Pick a simple arc $J' \subset J$ with $J \setminus J' \subset D_A$. By COMPLEMENT, there is a path from C to the rectangle's exterior avoiding J'. This implies that C meets the disk D_A. Similarly, C meets a corresponding disk D_B around B. We can then draw a path in $C \cup D_A \cup D_B$ from the left of the rectangle to the right of the rectangle, avoiding the top to bottom path (blue). This is contrary to CROSSING PATHS.

Draw the following path E in the rectangle from top to bottom.

- Drop along the central vertical line from the top edge of the rectangle to the highest point on U.

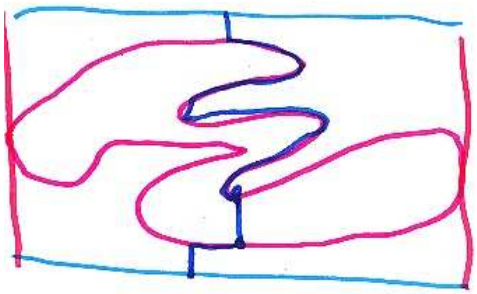


Figure 9.4. *Constructing the top-to-bottom path*

- along U to the lowest point p on U along the central vertical line. (By (*), the vertical segment from p to the bottom edge meets D.)
- From p, a vertical drop through a component C to the first point encountered on D.
- Along D to the lowest point of D on the central vertical line.
- Vertical drop to the bottom edge of the rectangle.

By (**), the every bounded component is encountered along E. Thus, there is at most one bounded component: C.

If C is actually a unbounded component, there is a path F from C in the rectangle to its top or bottom edge that does not meet J. If it reaches the bottom edge, we get a top to bottom path that does not meet D. (Follow E from the top edge as before down to C, then follow F to the bottom.)

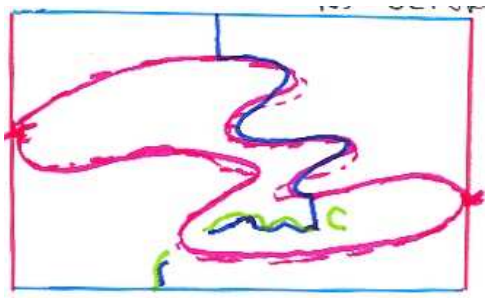


Figure 9.5. *A path avoiding D*

Similarly, if F reaches the top edge, we get a top to bottom path F that does not meet U. (Follow F from the top edge into the component C, then transfer to E to reach the bottom edge.) This is contrary to (*). Hence C is in fact bounded.

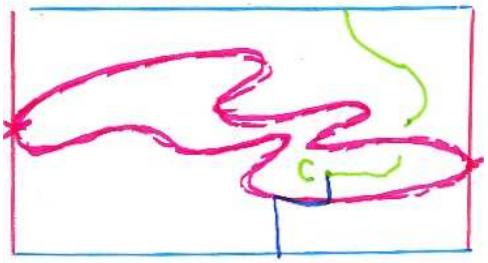


Figure 9.6. *A path avoiding U*

Thus, there are precisely two connected components: C and the unbounded component. □

Chapter 10

The Utility Puzzle

A children's puzzle is to connect three houses to three utility plants without any crossing paths. In mathematical terms, this asks for the embedding of the K_{33} graph in the plane.

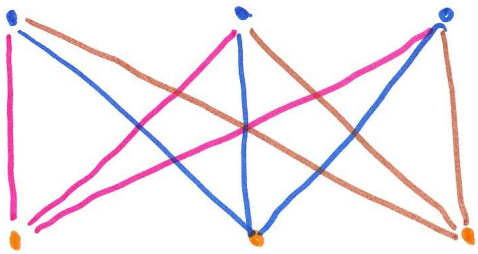


Figure 10.1. *The utility puzzle is to realize this graph (called K_{33}) in the plane with no crossing paths.*

Theorem 10.1. *The utility puzzle cannot be solved in the plane with polygonal paths. That is, K_{33} is not a polygonal plane graph.*

Lemma 10.2. *Let J be a Jordan curve that contains a linear segment $J' = (P_1, P_2)$. Suppose I is an open segment that crosses J' at a single point $P \neq P_i$, then P separates I into two pieces lying in different components of the complement of J.*

Proof. Pick a small disk D around P such that J and I are crossing lines inside D. The complement $D \setminus J$ has two connected components: the two halves of the

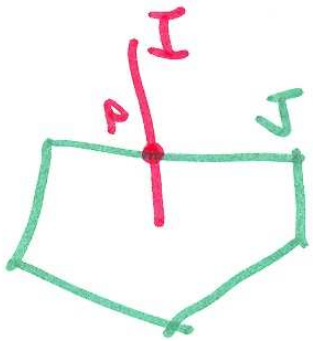


Figure 10.2. *The segment I crosses the polygonal Jordan curve J.*

disk. By XXX, each connected component of $\mathbb{R}^2 \setminus J$ meets $D \setminus J$, and therefore coincides with a half-disk. Since I meets both halves, and the parts are separated by P, the result follows. □

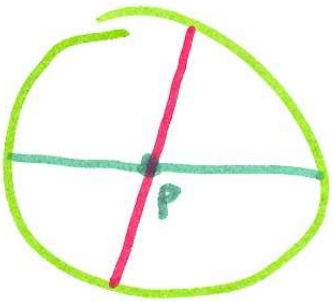


Figure 10.3. *The intersection of I and J inside a small disk.*

Proof. The K_{33} graph can be described as a hexagon J with three diagonals. (It does not matter that J is precisely a hexagon. It can by any polygonal Jordan curve.) By the Jordan curve theorem, the complement of J has only two connected components. Hence, at least two of the three diagonals land in the same component of the complement of J. Let two diagonals that land in the same component be called AB and CD. Let J_2 be a polygonal Jordan curve formed by the diagonal AB and an arc of J from the endpoint A to the endpoint B.

Let C' near C and D' near D lie on the diagonal CD. We can construct a polygonal path C' to D' crossing J_2 exactly once in the fashion prescribed by the

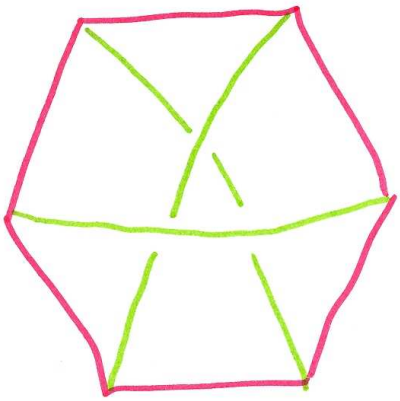


Figure 10.4. *The utility puzzle is to draw the three diagonals of a hexagon in such a way that they do not cross.*

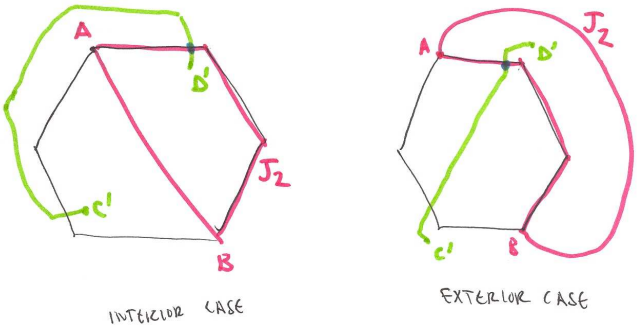


Figure 10.5. *We cross J_2 exactly once on the path from C' to D'.*

lemma. (Just cross J near C and cross J again near D. Only one of these crossings of J is on J_2.) Hence by the lemma, C' and D' lie in different connected components of the complement of J_2. So there is no diagonal joining them avoiding J_2. □

Chapter 11

The Jordan-Schönflies Theorem

Theorem 11.1. *If J is any Jordan curve in the plane, then any homeomorphism of J to the unit circle, extends to a homeomorphism of J together with its interior region to the the unit disk.*

This chapter is not complete. The proof will be based on Thomassen's proof.

Chapter 12

Triangulating Polygons

Theorem 12.1. *Every polygon admits a triangulation.*

Figure 12.1. *A triangulation*

Proof. We argue by induction on the number of vertices. The base case of the induction (the three vertex case) is trivial. Pick the coordinate axis so that no two vertices have the same y coordinate. Let B be the lowest vertex on the polygon, and let A and C be adjacent to B.

If AC is an interior diagonal, we draw the diagonal AC, forming a triangle ABC and a polygon without the vertex C. This smaller polygon we triangulate by induction.

Otherwise, let D be the lowest vertex interior to the triangle ABC. Cut the polygon into two along the edge BD. Both of the resulting polygons have fewer vertices, and can be triangulated by induction. □

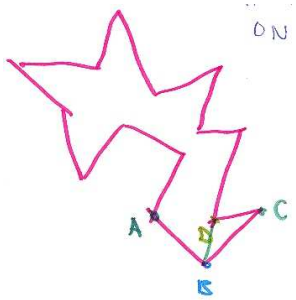


Figure 12.2. *Second case of the proof*

Chapter 13

Triangulating Surfaces

Theorem 13.1. *Every compact surfaces admits a triangulation.*

This chapter is not complete. The proof will be based on Ahlfors and Sario. It is a consequence of Jordan-Schönflies.

Chapter 14

Ears of Polygons

Theorem 14.1. *Every triangulation of a polygon (with at least four sides) has at least two ears; that is, triangles with two edges on the polygon.*

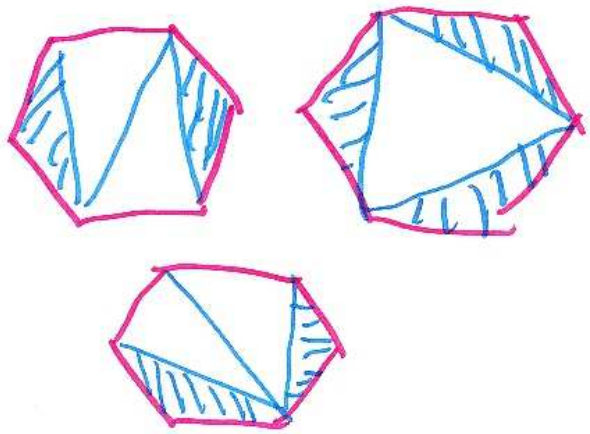


Figure 14.1. *Blue ears of triangulations*

Proof. Create a graph whose nodes are the centers of gravity of the triangles and whose edges are segments for each pair of triangles sharing an edge.

A path in the graph cannot loop back on itself (that is, the graph is a tree). Otherwise, the loop in the graph is a Jordan curve. A triangle corresponding to a node of the graph has a vertex inside the Jordan curve and a vertex outside. But the path along the polygon between the vertices of the polygon puts them in the same connected component of the complement of the Jordan curve. This contradicts the

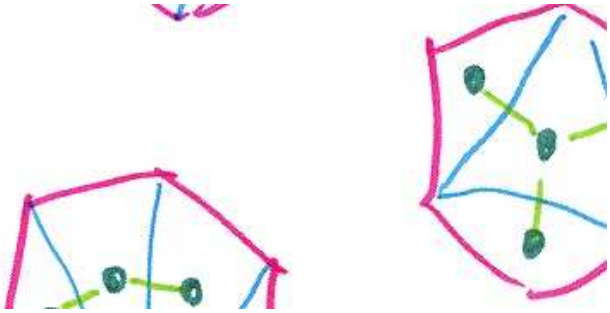


Figure 14.2. *The graph of a triangulation*

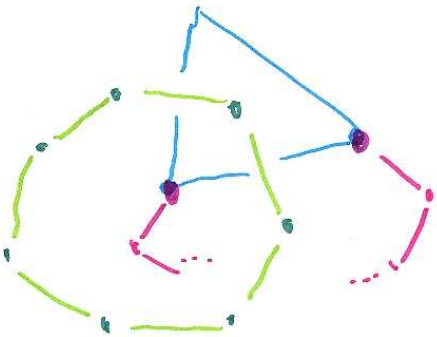


Figure 14.3. *A cycle in the graph*

Jordan curve theorem. This contradiction proves that the graph really is a tree.

Start from any node and follow a path to an end of the graph. The ending node is an ear. Follow in the opposite direction to a second ear. □

Chapter 15

Three Coloring a Triangulation

Theorem 15.1. *Fix a triangulation of a polygon. The vertices of the polygon can be three colored so that the vertices of each triangle receive distinct colors.*

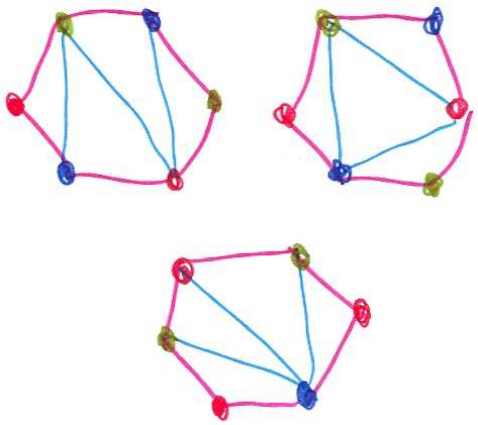


Figure 15.1. *A three-colored triangulation*

Proof. Remove an ear from the triangulation. The smaller triangulation has a good three coloring by induction. The induction assigns colors to two vertices on the ear. Give the third vertex on the ear the third color. □

Chapter 16

Guarding a Polygon

Theorem 16.1. *The interior of a polygon with n vertices can be guarded with $\lfloor n/3 \rfloor$ (integer part) guards.*

Proof. Triangulate the polygon and three color the vertices. Place guards at the vertices that are colored by the least frequent color. There are at most $\lfloor n/3 \rfloor$ vertices of that color. □

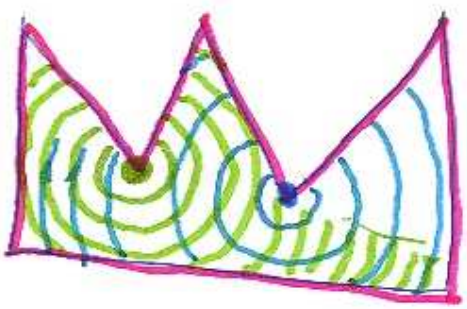


Figure 16.1. *Guards*

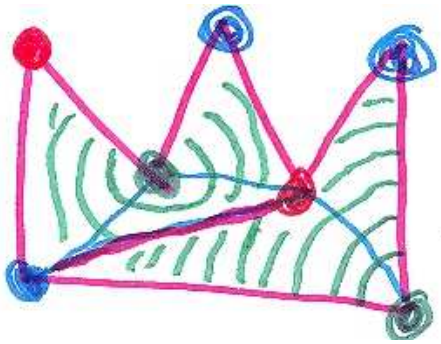


Figure 16.2. *Guards stationed at red vertices*

Chapter 17

Platonic Solids

A bounded convex polyhedron in Euclidean space is a bounded region defined by a finite number of linear inequalities:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \leq b.$$

Each linear inequality defines a half-space

$$H_{a,b} = \{x : ax \leq b\}, \quad a = (a_1, \ldots, a_n).$$

Each half-space is convex, so a bounded convex polyhedron is an intersection of finite sets (the half-spaces), and is consequently a convex set.

A *face* of a bounded convex polyhedron P is the intersection of P with a supporting hyperplane

$$F = P \cap \{x : ax = b\}$$

such that

- $P \subset H_{a,b}$, and
- F is not empty.

Now consider bounded convex polyhedra in three dimensions. If a face is not contained in a line, it is called a facet. It can be shown that a facet is a convex polygon together with its interior. A *vertex* is a face that is a point.

A regular bounded convex polyhedron is a bounded convex polyhedron for which every facet is congruent to every other, every face is a regular polygon (and interior), and each vertex meets the same number of faces.

Theorem 17.1. *Every regular bounded convex polyhedron in three dimensions is one of the following:*

- *a tetrahedron,*
- *an octahedron,*

- *a cube,*
- *an icosahedron, or*
- *a dodecahedron.*

Figure 17.1. *The Platonic solids*

Proof. The exterior angles of a polygon sum to 2π.

The sum of the interior angles θ of a regular polygon with n sides is $n\theta = \pi(n-2)$ [EULER TORUS].

Let m be the number of faces at each vertex v. Let L be a supporting hyperplane that meets P in a single point, a vertex v. A sufficiently small parallel translate L' of L meets P in a convex polygonal region C, and the part of P between L and L' is a bounded convex polyhedron, a cone P' with base C and vertex v. The facets of P' are C and m triangles, each with angle θ at v.

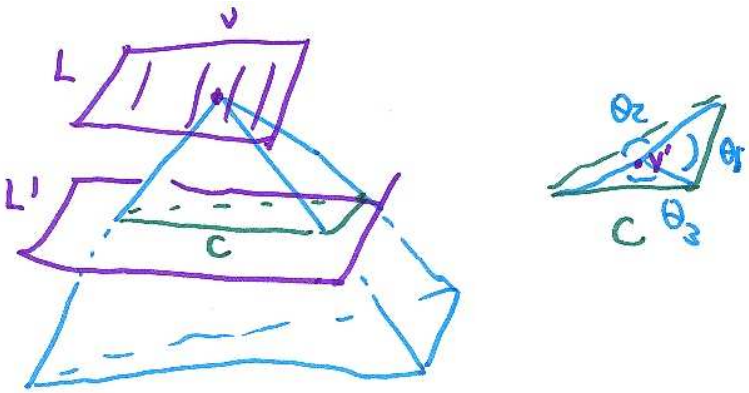


Figure 17.2. *Near a vertex, the polyhedron is a cone over a convex polygon.*

Under orthogonal projection to L', the vertex v projects to a point v' and the m triangles project to m triangles with vertex v', giving a triangulation of C.

Under projection, the angles θ increases to angles θ_i, and

$$m\theta < \sum_{i=1}^{m} \theta_i = 2\pi.$$

This gives

$$m\pi(1 - 2/n) < 2\pi.$$

This is equivalent to

$$(n-2)(m-2) < 4.$$

The only positive integers less than 4 are 1, 2, and 3, and these integers have few factorizations. In fact, there are five possible factorizations of 1, 2, or 3, and five corresponding solids.

$(n-2)(m-2)$	$n-2$	$m-2$	n	m	solid
1	1	1	3	3	tetrahedron
2	1	2	3	4	octahedron
2	2	1	4	3	cube
3	1	3	3	5	icosahedron
3	3	1	5	3	dodecahedron

□

Chapter 18

Maxwell Cremona

Note: This chapter will be rewritten. The homotopy argument that is used below is not very rigorous without further argument. Instead, I will substitute a simple induction argument similar to that used in the five-color theorem. The region will be cut into two smaller pieces, the Maxwell-Cremona function will be constructed on each piece, and the function on the two pieces can be made to agree along the boundary.

Let G be a finite plane polygonal graph. Let E be the set of edges of G. Let $\omega : E \to \mathbb{R}$ be an assignment of real numbers to each edge. If $\{i, j\} \in E$, write $\omega_{ij} = \omega ji$ for the value of ω on $\{i, j\}$.

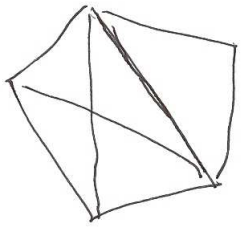


Figure 18.1. *A plane polygonal graph*

We write $(x, y)^{\perp} = (-y, x)$, for a vector orthogonal to (x, y). Fix $a_0 \in \mathbb{R}^2$ and $b_0 \in \mathbb{R}$. Fix a face F_0 of the graph.

Theorem 18.1. *The following two conditions are equivalent:*

1. *There is a unique continuous function $g : \mathbb{R}^2 \to \mathbb{R}$ such that*
   - *It is linear on each face: $g|_F(x) = a_F x + b_F$ for some a_F and b_F.*

- $g|_{F_0}(x) = a_0 x + b_0$.
- *On adjacent faces F, F' separated by edge $\{i,j\}$ with endpoints $p_i, p_j \in \mathbb{R}^2$, we have $a_{F'} = a_F + \omega_{ij}(p_i - p_j)^\perp$. (Here we assume that j immediately follows i in the counterclockwise traversal of the vertices of F.)*

2. *For each i, we have*

$$\sum_{j:\{i,j\}\in E} \omega_{ij}(p_j - p_i) = 0.$$

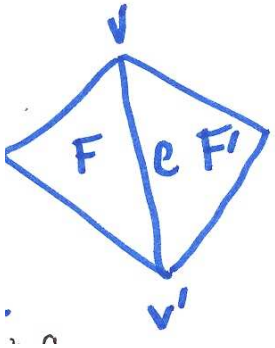


Figure 18.2. *Adjacent faces in the graph.*

Proof. We show that the first condition implies the second. Assume that the piecewise linear function g exists. As we circle the vertex i, label the faces $F, F', F'', \ldots$. Label that edges between consecutive faces $\{i,j\}, \{i,j'\}, \ldots$. We find that

$$\begin{array}{ll} a_{F'} & = a_F + \omega_{ij}(p_j - p_i)^\perp \\ a_{F''} & = a_{F'} + \omega_{ij'}(p_{j'} - p_i)^\perp \\ \ldots & = \ldots \\ (a_{F'} + \cdots a_F) & = (a_F + a_{F'} + \cdots) + \sum_{j:\{i,j\}\in E} \omega_{ij}(p_j - p_i)^\perp. \end{array}$$

It follows that the sum over j of $\omega_{ij}(p_j - p_i)$ is zero.

Now we show that the second condition implies the first. To define g on F, pick a path $F_0, F_1, \ldots, F_N = F$ of faces, each adjacent to the next. We will abbreviate a_{F_i}, b_{F_i} and $g|_{F_i}$ to a_i, b_i, and g_i. Define $g_i = a_i x + b_i$ recursively in terms of the function g_{i-1} by using the relation $(*)$ to define the slope a_i. Define b_i to be the unique constant term to make g continuous along the intersection of the faces $F_i \cap F_{i-1}$. Explicitly, if p is a shared vertex of F_i and F_{i-1}, the defining relation for b_i is

$$a_i p + b_i = a_{i-1} p + b_{i-1}.$$

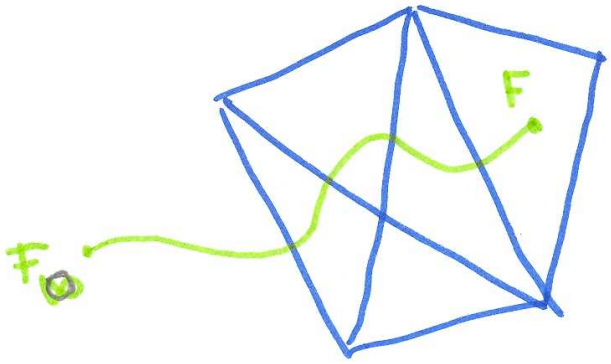


Figure 18.3. *A path from one face to another.*

(If p and q are the two shared vertices, it is a consequence of the fact that the change in slopes is proportional to $(p-q)^{\perp}$ that the value for b_i does not depend on which of the two vertices p or q is used.)

The function defined in this way is independent of the path that was chosen from F_0 to F. In fact, any two paths can be deformed one into the other by a succession of two changes.

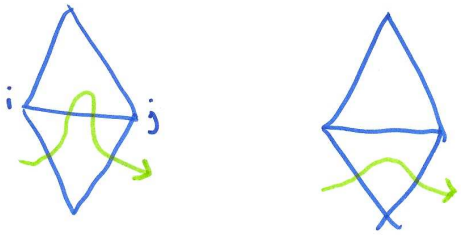


Figure 18.4. *Deforming a path through an edge.*

(1) We can fold along an edge. The definition of g is independent of such a change because of the relation

$$(p_j - p_i)^{\perp} + (p_i - p_j)^{\perp} = 0.$$

(2) We can collapse along a vertex. Here we have path independence for g by the relation

$$\sum_j \omega_{ij}(p_i - p_j) = 0.$$

So g is a well-defined function. □

Figure 18.5. *Deforming a path through a vertex.*

Chapter 19

The Area of a Triangle

This chapter assumes a knowledge of determinants. We will derive a formula for the area of a triangle as a special case of a formula that gives the volume of a general n-dimensional simplex (a triangle in dimension two, a tetrahedron in dimension three, and so forth).

Let $v_i \in \mathbb{R}^n$ be row vectors $i = 0, \ldots, n$. If number a_{ij} are given, write $[a_{ij}]$ for the matrix whose entries are a_{ij}. Write $\underline{c}$ for a row vector whose entries are all $c \in \mathbb{R}$. (The size of the row will be clear in every case from the context in which it appears.) Let $k_{ij} = \|v_i - v_j\|^2$. We write $|A|$ for the determinant of the matrix A. Let

$$D = |v_1 - v_0, \ldots, v_n - v_0|.$$

The absolute value of D is the volume of the parallelopiped spanned by $v_1 - v_0, v_2 - v_0, \ldots, v_n - v_0$.

Theorem 19.1.

$$D^2 = \frac{(-1)^{n-1}}{2^n} \begin{vmatrix} [k_{ij}] & {}^t\underline{1} \\ \underline{1} & 0 \end{vmatrix}.$$

Proof. Let V be the $n+1$ by n matrix whose ith row is v_i. The key step in the proof will be the matrix product

$$V^t V = [v_i \cdot v_j].$$

We have

$$k_{ij} = \|v_i - v_j\|^2 = \|v_i\|^2 + \|v_j\|^2 - 2v_i \cdot v_j.$$

The right-hand side of the statement can be simplified by row and column operations. The first of these is two add the multiple $\|v_i\|$ of the last column to the ith

column and a multiple $\|v_j\|$ of the last row to the jth row:

$$\begin{aligned}
&\tfrac{(-1)^{n-1}}{2^n}\begin{vmatrix}[k_{ij}] & {}^t\underline{1}\\ \underline{1} & 0\end{vmatrix}\\
&\qquad = \tfrac{(-1)^{n-1}}{2^n}\begin{vmatrix}[k_{ij}-\|v_i\|^2-\|v_j\|^2] & {}^t\underline{1}\\ \underline{1} & 0\end{vmatrix}\\
&\qquad = \tfrac{(-1)^{n-1}}{2^n}\begin{vmatrix}[-2v_i\cdot v_j] & {}^t\underline{1}\\ \underline{1} & 0\end{vmatrix}\\
&\qquad = -\begin{vmatrix}[v_i\cdot v_j] & {}^t\underline{1}\\ \underline{1} & 0\end{vmatrix}\\
&\qquad = -\begin{vmatrix}V & {}^t\underline{1} & {}^t\underline{0}\\ \underline{0} & 0 & 1\end{vmatrix}\begin{vmatrix}{}^tV & {}^t\underline{0}\\ \underline{0} & 1\\ \underline{1} & 0\end{vmatrix}\\
&\qquad = \begin{vmatrix}V & {}^t\underline{1}\end{vmatrix}\begin{vmatrix}{}^tV\\ \underline{1}\end{vmatrix} = D^2.
\end{aligned}$$

(To get to the last line, we expand the first determinant along the last row, and the second determinant along the next to last row, picking up a negative sign in the process.) □

Now we specialize to two dimensions. The area of a triangle is half the area of the parallelogram formed by two of its sides. The formula gives an expression for the area of a triangle in terms of the squares $k_{12} = a^2, k_{13} = b^2, k_{23} = c^2$ of the three sides. Expanding the determinant, we get precisely the classical Heron formula for the area of a triangle with sides a, b, c:

$$A^2 = 4D^2 = s(s-a)(s-b)(s-c),$$

where $s = (a+b+c)/2$. When $n = 3$, we get a formula for the volume of tetrahedron as a function of the squares of its edge lengths. This theorem gives the n-dimensional generalization of Heron's formula.

Chapter 20

The Area of a Polygon

In this chapter, in the interest of keeping everything easy, we assume that we do not have any knowledge of calculus, measure theory or related topics. We give a direct construction of area for polygons in the plane.

If $u, v \in \mathbb{R}^2$, we set $\det(u, v) = u_1 v_2 - u_2 v_1$. (We use $\det(A)$ rather than $|A|$ in this chapter to signify the determinant, to avoid confusion with the absolute value.)

Define the area of a triangle T with vertices u, v, w by

$$m(T) = (1/2)|\det(v - u, w - u)|.$$

Let P be a polygon. For any triangulation $\mathcal{T}$ of P, let

$$a(\mathcal{T}) = \sum_{T \in \mathcal{T}} m(T).$$

Theorem 20.1. *$a(\mathcal{T})$ is independent of the triangulation $\mathcal{T}$ of P.*

Thus, we have a well-defined area function, defined by $m(P) = a(\mathcal{T})$ for any triangulation $\mathcal{T}$ of P.

Proof. We prove this by induction on the number of sides N of P. We start with the base case $N = 4$. (We will see that there are reasons not to start with the trivial base case $N = 3$.)

A quadrilateral has at most two triangulations. Let us consider two triangulations of a triangle. Orienting the triangles so that the determinants are positive, we have

$$\begin{aligned} a(\mathcal{T}) &= \det(v_1 - v_0, v_2 - v_0) + \det(v_2 - v_0, v_3 - v_0) \\ &= \det(v_2 - v_0, (v_3 - v_0) - (v_1 - v_0)) \\ &= \det(v_2 - v_0, v_3 - v_1). \end{aligned}$$

This expression is the determinant formed by the two diagonal vectors of the quadrilateral. Keeping in mind the skew symmetry properties of the determinant, this

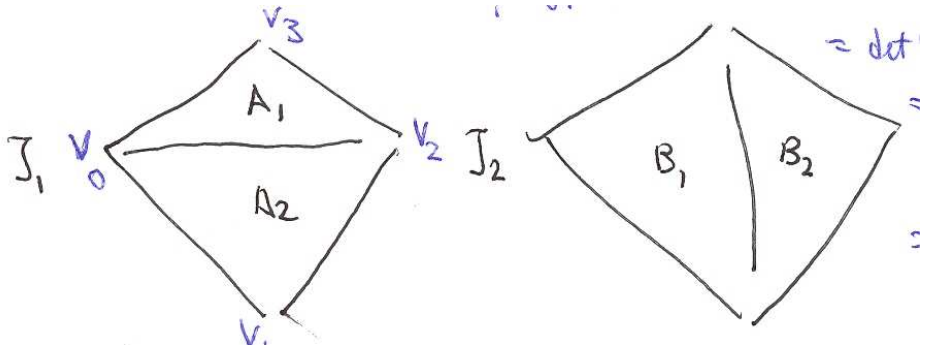


Figure 20.1. *The two triangulations of a quadrilateral*

determinant is symmetrical in the two diagonals. Thus, $a(\mathcal{T})$ does not depend on which diagonal is chosen for the triangulation. This finishes the proof of the base case.

Now assume the theorem for polygons with N sides. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be triangulations of an $N+1$-gon. We distinguish three cases:

1. If $\mathcal{T}_1$ and $\mathcal{T}_2$ have a common ear $T_0 \in \mathcal{T}_1 \cap \mathcal{T}_2$, then lopping off the ear, we see that $\mathcal{T}_i' = \mathcal{T}_i \setminus \{T_0\}$ is a triangulation of an N-gon. By the induction hypothesis, the right hand side of the equation

$$a(\mathcal{T}_i) = m(T_0) + \sum_{T \in \mathcal{T}_i'} m(T).$$

   is independent of i. Hence $a(\mathcal{T}_1) = a(\mathcal{T}_2)$.

2. If $\mathcal{T}_1$ and $\mathcal{T}_2$ have nonadjacent ears T_1 and T_2, respectively. Let $\mathcal{T}_3$ be a triangulation with both ears T_1, T_2. Then by the previous case,

$$a(\mathcal{T}_1) = a(\mathcal{T}_3) = a(\mathcal{T}_2).$$

3. Finally, assume the two triangulations have no common ears, and that every ear of one is adjacent to every ear of the other. Recall that each triangulation has at least two ears (TWO EARS). So two ears of one triangulation are each adjacent to two ears of the other triangulations. So much adjacency among so many ears forces the polygon to be a quadrilateral. This is our base case, and we are done.

□

Chapter 21

Scissors and Polygons

Theorem 21.1. *Any polygon in the plane can be cut into finitely many triangles that can be reassembled into a rectangle of unit width.*

Proof. Let P be a polygon. If we can prove the result for triangles, then we get the result in general, by triangulating the polygon P.

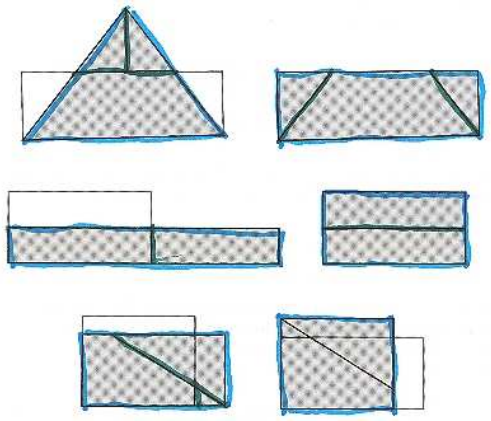


Figure 21.1. *Scissor operations transform a triangle to a unit width rectangle.*

Each triangle can be cut into three pieces that can be reassembled into a rectangle. Make a cut parallel to the longest edge AB passing through the midpoints of the other sides. Make a second cut perpendicular to the first, from C to the parallel cut.

Rectangles can be cut into two halves and stacked, until the height falls in the interval $[1/2, 1]$.

Finally, rectangles (other than the long skinny ones we have eliminated) can be rescaled by making two cuts, so that the width of the rectangle is exactly 1. □

Chapter 22

The Dehn Invariant

It follows from the piece on scissors and polygons that a polygon can be cut into pieces and reassembled into a rectangle of the same area. The obvious generalization of this statement to three dimensions is false in a dramatic way.

Theorem 22.1. *A regular tetrahedron of side* 1 *cannot be cut into finitely many polyhedra and reassembled into a cube of the same volume.*

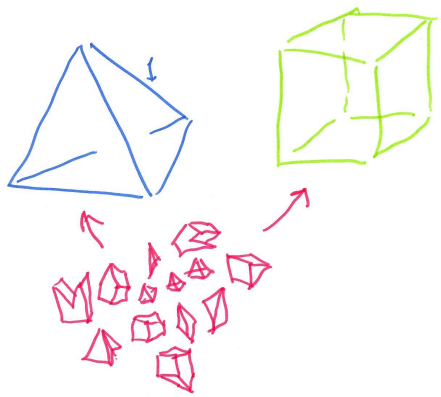


Figure 22.1. *Polyhedral pieces cannot be assembled into both a cube and a regular tetrahedron.*

We start with a couple of simple lemmas.

Lemma 22.2. *For all* $a, b \in \mathbb{Z}$*, with* $b > 0$ *we have*

$$1/3 \neq \cos(2\pi a/b).$$

Proof. Suppose for a contradiction that this can be done. Replacing a with $a \pm kb$, we can assume that $0 \leq a < b$. Replacing a with $b - a$ if necessary, we can assume that $0 \leq a \leq b/2$. Then $\sqrt{8}/3 = \sin(2\pi a/b)$. Set $y = 1 + i\sqrt{8}$.

DeMoivre's theorem states that for all $n \in N$, we have

$$(\cos x + i \sin x)^n = \cos(nx) + i \sin(nx).$$

So

$$y^b = 3^b(\cos(2\pi a/b) + i \sin(2\pi a/b))^b = 3^b(\cos(2\pi a) + i \sin(2\pi a)) = 3^b.$$

This means that

$$y^b = 3^b \equiv 0 \mod 3,$$

where we write $r + is\sqrt{8} \equiv 0 \mod 3$ to mean that both r and s are multiples of 3.

We compute

$$y^2 = -7 + 2i\sqrt{8} \equiv -y \mod 3.$$

By induction $y^b \equiv (-1)^{b-1}Y \mod 3$, which is not $0 \mod 3$. This contradiction proves the result. □

Lemma 22.3. *The dihedral angle of a regular tetrahedron is* $\arccos(1/3)$.

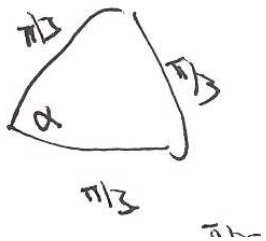


Figure 22.2. *The calculation of the dihedral angle of a regular tetrahedron*

Proof. The dihedral angle of a regular tetrahedron is equal to the angle of the triangle formed by the perpendicular bisector of a side. This triangle has sides $\sqrt{3}/2$, $\sqrt{3}/2$, and 1. By the law of cosines (Law of Cosines):

$$1 = c^2 = a^2 + b^2 - 2ab\cos\theta = 3/2 - 3/2\cos\theta.$$

We solve for θ to get the result. □

Proof. [Theorem] Let $\alpha = \arccos(1/3)$. By the first lemma, α is not a rational multiple of π.

Suppose for a contradiction that this can be done. Let $\alpha_1, \ldots, \alpha_n$ be the dihedral angles of all of the polyhedral pieces into which the tetrahedron is cut. Let

$$V = \text{span}_{\mathbb{Q}}\langle \alpha, \pi, \alpha_1, \ldots, \alpha_n \rangle$$

be the set of all rational linear combinations of α, π and the angles α_i. There is a subset

$$S = \{\alpha, \pi, \ldots\}$$

that is a basis of V. (Since α and π are not rational multiples of one another, we can include both in the basis.)

Define a linear function $F : V \to \mathbb{Q}$ by its values on the basis S:

$$F(\alpha) = 1,\ F(\beta) = 0$$

if $\beta \in S \setminus \{\alpha\}$. In particular $F(\pi r) = 0$ for every rational number r.

Define the Dehn invariant to be the real number

$$D = \sum_i \ell_i F(\alpha_i),$$

where the sum runs over all edges of all the polyhedral pieces. Here ℓ_i is the length of the edge and α_i is the corresponding dihedral angle. Notice that the Dehn invariant does not change if an edge is subdivided into two shorter edges.

We compute the sum two ways, first by arranging the pieces into a cube, then by arranging the pieces into a tetrahedron. If we consider the contribution to the Dehn invariant from a segment of length ℓ that appears internally in the cube or tetrahedron, it gives

$$\sum \ell F(\alpha_i) = \ell F(2\pi) = 0.$$

Similarly, a segment of length ℓ that appears on a face of the cube or tetrahedron gives contribution

$$\sum \ell F(\alpha_i) = \ell F(\pi) = 0.$$

The only contribution comes from the edges of the cube or tetrahedron. The dihedral angles of a cube are all $\pi/2$. Thus, for a cube

$$D = \sum_{i=1}^{12} \ell F(\pi/2) = 0.$$

However, for a tetrahedron

$$D = \sum_{i=1}^{6} F(\alpha) = 6F(\alpha) = 6.$$

Since $0 \neq 6$, the same pieces cannot form both a cube and a tetrahedron. □

[SIDE NOTE] DeMoivre's theorem is an easy induction argument based on the addition laws

$$\begin{aligned} \cos(x+y) &= \cos x \cos y - \sin x \sin y \\ \sin(x+y) &= \sin x \cos y + \cos x \sin y \end{aligned}$$

In fact, the base case $n = 0$ is trivial. If we assume DeMoivre for n, we get in the case $(n + 1)$ that

$$\begin{aligned}(\cos x + i \sin x)^{(n+1)} &= (\cos x + i \sin x)^n (\cos x + i \sin x) \\ &= (\cos nx + i \sin nx)(\cos x + i \sin x) \\ &= (\cos nx \cos x - \sin nx \sin x) + i(\sin nx \cos x + \cos nx \sin x) \\ &= \cos(n + 1)x + i \sin(n + 1)x\end{aligned}$$

[/SIDE NOTE]

Chapter 23

Five Colors Suffice

Theorem 23.1. *A finite plane map can be colored with five colors in such a way that adjacent states receive different colors.*

Proof. Imagine the states on an island surrounded by a body of water (or multiple islands if the states lie on separate continents). We will prove by induction on the number of states that there is a coloring with a palette of five colors in such a way that

- each coastal state is colored from a given sub-palette of three colors (the sub-palette may vary country to country)
- two given coastally adjacent countries receive prescribed colors from the palette.

We will call such a coloring *good*.

We start with the base case of the induction: if the island has at most three states, it is a trivial task to color the map with a good coloring.

Now assume the induction hypothesis: any map with fewer than N states has a good coloring. We consider two cases.

Case 1. Assume there is an adjacency between two coastal countries that are not coastally adjacent. That is, they are connected by an inland path, but not by a path along the shore.

In this case we build a canal along the given inland path, effectively separating the island into two, and forming new shoreline along the canal.

By the induction hypothesis, the new island with D and E has a good coloring. Match the colors along the canal states, then give a good coloring on the second island.

Case 2. No such inland path exists. Let F be the state coastally adjacent to E, not equal to D. Flood the state F so it disappears from the map. Let x and y be colors from the sub-palette of F, other than the color at E. The states adjacent to F become coastal states. Other than E and G, the states adjacent to F were not coastal before flooding F (because in this case there is no inland path). We give these new coastal states the sub-palette of all five colors except x and y. By

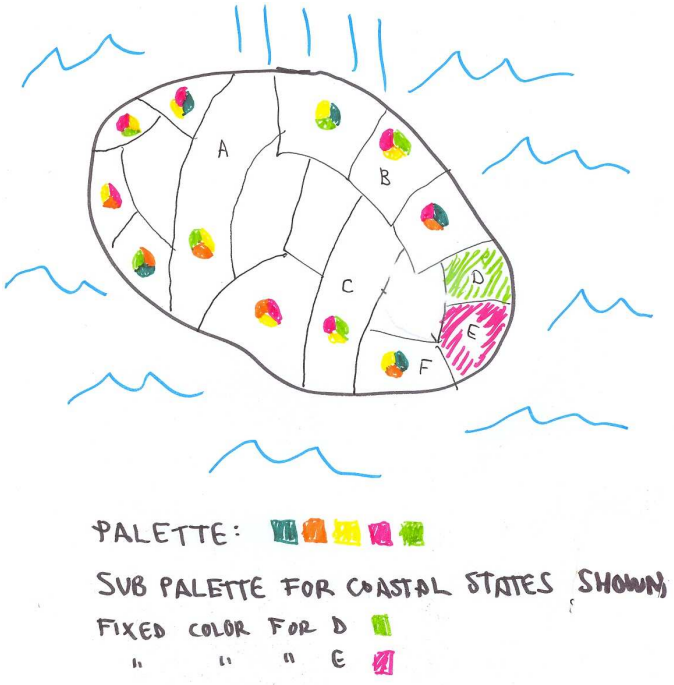


Figure 23.1. *Constraints for a good coloring*

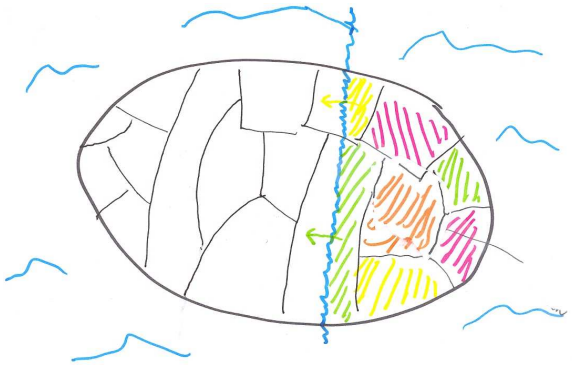


Figure 23.2. *An inland adjacency*

induction, the flooded island has a desired coloring. F can then be colored with x or y to complete the coloring. □

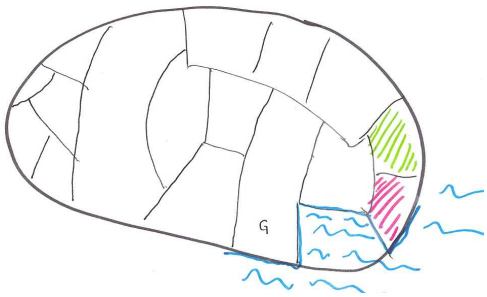


Figure 23.3. *Case 2*

Chapter 24

The Hyperplane Separation Theorem

We begin with the supporting hyperplane theorem.

Lemma 24.1. *Let C be a closed convex set in $\mathbb{R}^n$. Let $z \in C$ be a point on the boundary of C. Then there exists a unit vector $v \in \mathbb{R}^n$ such that*

$$x \cdot v \leq z \cdot v,$$

for all $x \in C$.

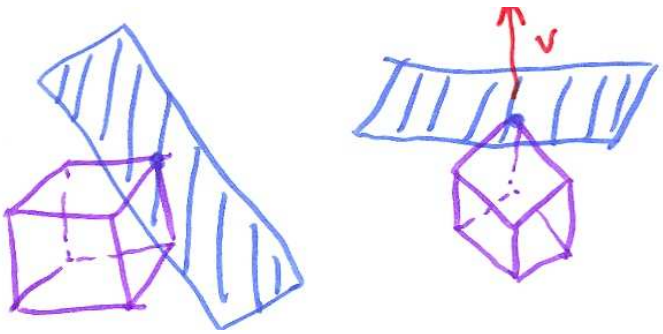


Figure 24.1. *Supporting hyperplanes*

The plane determined given by the equation $x \cdot v = z \cdot v$ is called a *supporting hyperplane*.

Proof. For each $y \in \mathbb{R}^n \setminus C$, there is a unique point $\hat{y} \in C$ closest to y. In fact, the set of points of C at least as close to y as z is closed and bounded, hence compact. Thus, the continuous function $x \mapsto \|y - x\|$ has a minimum $\hat{y}$ on C.

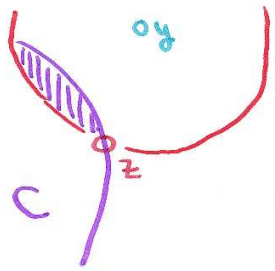


Figure 24.2. *Compactness*

If $\hat{y}$ and $\hat{y}'$ are both minima, the segment $[\hat{y}, \hat{y}']$ is also in C and has a unique point at minimal distance from y. So the minimum is unique.

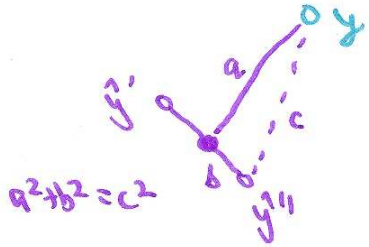


Figure 24.3. *Uniqueness of $\hat{y}$.*

Let $y_k \in \mathbb{R}^n \setminus C$ converge to z. Then also, $\hat{y}_k$ converges to z. Let

$$v_k = \frac{(y_k - \hat{y}_k)}{\|y_k - \hat{y}_k\|}.$$

Passing to a subsequence of y_k, we may assume v_k converges to some unit vector v.

Since $\hat{y}_k$ is the closest point in C to y_k, we have for $x \in C$:

$$\begin{aligned} \|y_k - \hat{y}_k\|^2 &\leq \|y_k - (tx + (1-t)y_k)\|^2 = \|(y_k - \hat{y}_k) - t(x - \hat{y}_k)\|^2, \text{ if } 0 \leq t \leq 1. \\ \|y_k - \hat{y}_k\|^2 &\leq \|y_k - \hat{y}_k\|^2 - 2t(y_k - \hat{y}_k) \cdot (x - \hat{y}_k) + t^2(\cdot). \end{aligned}$$

The linear term in t gives the inequality

$$x \cdot v_k \leq \hat{y}_k \cdot v_k.$$

In the limit as $v_k \to v$ and $\hat{y}_k \to z$, we get

$$x \cdot v \leq z \cdot v.$$

□

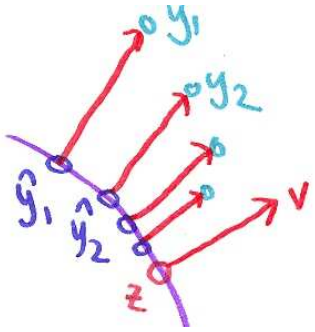


Figure 24.4. *Convergence of normals.*

Theorem 24.2. *Let C_1 and C_2 be two nonempty disjoint convex sets. Then there exists a unit vector v such that for all $c_1 \in C_1$ and $c_2 \in C_2$, we have*

$$c_1 \cdot v \leq c_2 \cdot v.$$

Moreover, if C_1 and C_2 are closed, we may assume that the inequality is strict: $c_1 v < c_2 v$.

Proof. $0 \notin X = \{c_1 - c_2 : c_1 \in C_1,\ c_2 \in C_2\}$. The closure $\bar{X}$ of X is a closed convex set. Let z be the closest point to 0 in $\bar{X}$ (simply set $z = 0$ if $0 \in \bar{X}$). Let v be the vector produced by the lemma for z.

$$(c_1 - c_2) \cdot v \leq z \cdot v \leq 0 \cdot v.$$

The result follows.

Moreover, if C_1 and C_2 are closed, $\bar{X} = X$ and $z = \hat{0} \neq 0$. Rather than appealing to the lemma, we can take $v = -z/\|z\|$. The inequality $z \cdot v < 0 \cdot v$ is strict. The inequality follows. □

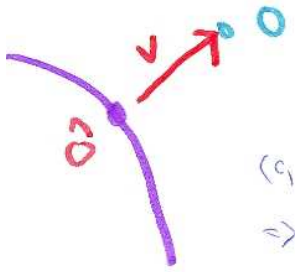


Figure 24.5. *$z = \hat{0}$ is the closest point to 0.*

Chapter 25

Linear Programming Duality

Consider the maximization problem

$$\begin{array}{lll} \max_x cx & \text{subject to constraints} & Ax = b \\ & & x \geq 0 \end{array}$$

Here c is a row vector, A is a matrix, x and b are column vectors. A, b, c are given. The maximization runs over all x whose coordinates are nonnegative real numbers.

The objective function cx is a linear function. Thus, the problem asks how far a hyperplane $cx = t$ can be shifted subject to the constraint that the hyperplane meets the convex polyhedron defined by the constraints.

This is called the *primal linear program*. It is said to be a *feasible* problem if there exists a vector x satisfying the given constraints. It is said to have a *bounded objective function* if there is a real number R such that $cx \leq R$ for all x satisfying the constraints.

We consider a second optimization problem, which is called the *dual linear program*.

$$\begin{array}{lll} \max_y yb & \text{subject to constraints} & \\ & & yA \geq c \end{array}$$

Similarly, it is said to be feasible if there exists a row vector y satisfying the constraints. it has a bounded objective function if yb is bounded below for all y satisfying the constraints.

Theorem 25.1. *Assume that the primal linear program is feasible with bounded objective function. Let $M = cx^*$ be the maximum value obtained. Then the dual linear program is feasible and has a bounded objective function. Let m be the minimum of the dual problem. Then $m = M$.*

Proof. The image of a convex set under a linear map is convex. Thus, the image K of the orthant $\{(x, s) \mid x \geq 0, s \geq 0\}$ (which is convex) under the linear map L:

$$\begin{pmatrix} s \\ x \end{pmatrix} \mapsto^L \begin{pmatrix} M & -c \\ b & -A \end{pmatrix} \begin{pmatrix} s \\ x \end{pmatrix} = \begin{pmatrix} Ms - cx \\ bs - Ax \end{pmatrix} = \begin{pmatrix} r \\ v \end{pmatrix} \in K$$

is a convex set.

We claim that $(-1, 0) \notin K$. For a contradiction assume it lies in K. This means that $-1 = sM - cx$ and $0 = sb - Ax$ for some non-negative s and x. These conditions imply that $x' = (1 - s)x^* + x$ is a feasible solution:

$$Ax' = (1 - s)b + sb = b.$$

Moreover, $cx' = (1-s)M+sM+1 > M$, which is contrary to the extremal property of M. Thus, $(-1, 0) \notin K$.

The image under L of $(s, x) = (1, x^*)$ is $(0, 0)$. So $(0, 0) \in K$.

Let $(-z, y)$ be the coefficients of a SEPARATING HYPERPLANE between the closed convex sets K and $\{(-1, 0)\}$. This means

$$-zr + yv < -z(-1) + y0 = z, \tag{25.1}$$

for all $(r, v) \in K$. In particular, substituting the point $(r, v) = (0, 0)$ into this inequality, we get $0 < z$. Divide both sides of the inequality by z and set $y^* = y/z$ to get

$$-r + y^*v < 1.$$

Writing $r = Ms - cx$ and $v = bs - Ax$, the inequality becomes

$$(y^*b - M)s + (c - y^*A)x < 1, \text{ for all } x \geq 0,\ s \geq 0.$$

If $\alpha s + \alpha' x' < 1$ for all $s \geq 0$ and $x' \geq 0$, then $\alpha, \alpha' \leq 0$. This implies that

$$y^*b \leq M \text{ and } y^*A \geq c.$$

This implies that y^* is a feasible solution of the dual.

If y is any feasible solution to the dual, then

$$yb = y(Ax^*) = (yA)x^* \geq cx^* = M \geq y^*b.$$

Thus, y^* achieves the minimum m for the dual linear program. Setting $y = y^*$, this chain of inequalities become sandwiched between y^*b on both ends, and a chain of equalities results. This gives $m = y^*b = M$. □

Linear programming problems admit many variations. Often, linear programming problems in nonstandard form can be converted to the standard primal form stated above.

For example, an inequality

$$ax \leq b$$

can be converted to an equality constraint by introducing a 'slack' variable y: $ax + y = b$, where $y \geq 0$.

If some variable x_i lacks a non-negativity constraint $x_i \geq 0$, then we can eliminate x_i from the system by writing it as a sum of non-negative variables $x_i = y_i - z_i$, for some $y_i, z_i \geq 0$.

Once a linear program has been converted to standard form, the linear programming duality theorem can be applied to get a dual problem with the same extreme value.

Chapter 26

Stress

A framework is a finite plane graph (with nodes p_i) whose edges are line segments $[p_i, p_j]$ in which every edge is labeled as either a 'bar' or a 'strut.'

In this piece, we will avoid further mention of the term *graph* in the sense of a collection of nodes and edges. This is to avoid confusion with the term *graph* in the sense of the graph of a function (graphical plot).

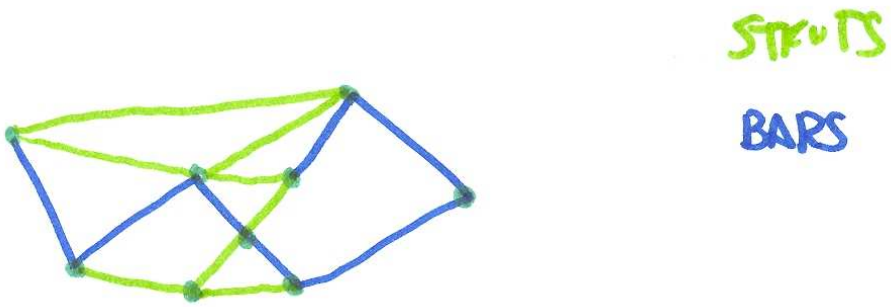


Figure 26.1. *A framework with its bars and struts. The bars in this example form a simple arc.*

An *equilibrium stress* on a framework is a function

$$\omega_{ij} : E \to \mathbb{R}$$

such that

$$\sum_{j:\{i,j\}\in E} \omega_{ij}(p_j - p_i) = 0, \text{ for all } i.$$

If each $\omega_{ij}(p_j - p_i)$ is viewed as a force, the equilibrium condition is that the vector sum of the forces is zero at each vertex.

An equilibrium stress is *proper* if $\omega_{ij} \leq 0$ for all struts $\{i, j\}$.

Theorem 26.1. *Suppose that the bars of a framework form a simple polygonal arc. Then the only proper equilibrium stress on the framework is zero on all struts:* $\omega_{ij} = 0$ *for all struts* $\{i, j\}$.

Proof. For a contradiction, let ω be a nonzero proper equilibrium stress.

By MAXWELL-CREMONA, there is a non-constant piecewise linear function $F : \mathbb{R}^2 \to \mathbb{R}$ that is identically zero on the unbounded face of the framework. The function F is linear on each face:

$$F|_{\text{face}}(x) = sx + t, \text{ some } s \in \mathbb{R}^2, t \in \mathbb{R}.$$

The slopes on adjacent faces along edge $\{i, j\}$ are related by

$$s = s' + \omega_{ij}(p_j - p_i)^{\perp}, \text{ with } (x, y)^{\perp} = (-y, x).$$

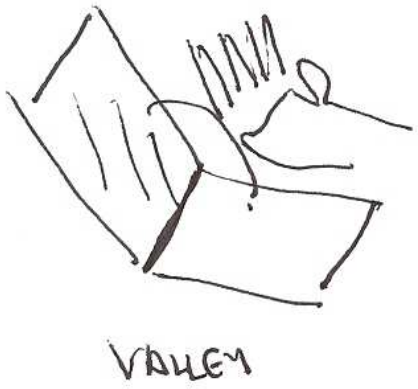


Figure 26.2. *The graph of a function at a valley.*

To get our contradiction, we study the polyhedron given by the graph of the function F. If $\omega_{ij} < 0$, we say the faces meet at a valley. (If the hand is place on top of the graph of F with the knuckles along the edge of the polyhedron between the faces, the fingers bend back along a valley.) The intersection of the graph of F with a plane $y = ax + b$ meeting transversally along the valley of the polyhedron has a greater fraction below the graph of F.

If $\omega_{ij} > 0$, we say the faces meet at a crest. (If the hand is place on top of the graph of F with the knuckles along the edge of the polyhedron between the faces, the fingers bend forward along a crest.) The intersection of the graph of F with a plane $y = ax + b$ meeting transversally along the crest of the polyhedron has a greater fraction above the graph of F.

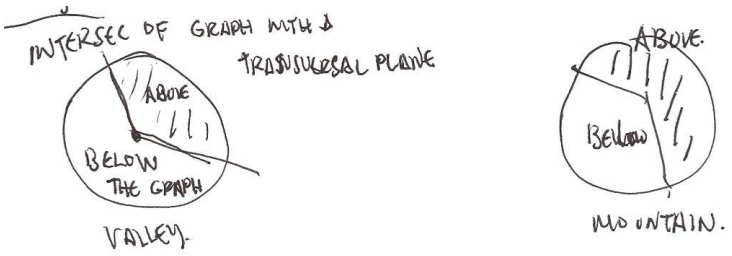


Figure 26.3. *The cross section of the graph at a valley or crest.*

We say that $x_0 \in \mathbb{R}^2$ is a tilted peak of F if there is a plane $y = ax + b$ such that

- $F(x_0) = ax_0 + b$
- $F(x) < ax + b$, for x in a small punctured disk around x_0.

In intuitive terms, we can tilt the graph by adding a linear function $-ax_0 - b$ to F, so that x becomes an isolated local maximum.

Lemma 26.2. *The function F does not have a tilted peak.*

Proof. Let x_0 be a tilted peak: $F(x_0) = ax_0 + b$. Then the level set $X_\epsilon = \{x | F(x) + \epsilon = ax + b\}$ for small $\epsilon > 0$ is a polygon (since F is piecewise linear). Let x_1, x_2, x_3 be extreme points of the convex hull of this polygon. Then the points x_i lie on distinct edges that are all crests (because the underside of the graph of F lies inside the convex hull and the picture looks like Fig **??**). These three crests give three bars that meet at the vertex x_0 of the framework. This is contrary to the assumption that the bars form a simple arc. □

Lemma 26.3. *For all x, we have $F(x) \leq 0$.*

Proof. Suppose that $\{x \mid F(x) > 0\}$ is nonempty. The function F is continuous and is identically zero on the unbounded region of the framework. Thus, F has compact support and attains its maximum value $M > 0$. Although the maximum of F may not be unique, we can fix this by tilting the graph of F: By adding a small linear function $-ax - b$ to F, we can arrange for F to have a tilted peak on the support of F. (In other words, we can tilt F to transform the global maximum of this piecewise linear function into a strict local max.) This is contrary to the previous lemma. □

Now we complete the proof of the main theorem. The bars form a simple arc, the complement of which is connected SIMPLE ARC. Thus, we can connect any face of the framework to any other by a path that does not cross any bars. The

function F is identically zero on the face at infinity. Starting from the unbounded face, follow a path from strut to strut avoiding bars. The first nonzero strut we encounter $\omega_{ij} < 0$ gives a valley, which imparts an upturn on the graph into a face on which $F(x) > 0$. This is contrary to the previous lema □

(Once the stress is zero on all struts, we see that the Maxwell-Cremona function has to be identically zero. It follows that the stress is identically zero on bars as well.)

Chapter 27

Carpenter's Rule

Theorem 27.1. *A polygonal arc in the plane with no unbent vertices admits a strict infinitesimal motion.*

Here is an elaboration of the meaning of the theorem. We have a polygonal arc with vertices $p_1, \ldots, p_N$ ordered consecutively along the arc formed by the segments $[p_i, p_{i+1}]$. To say that there are no unbent vertices means that we do not have three consecutive points p_i, p_{i+1}, p_{i+2} that are collinear. We can convert any polygonal arc into one with no unbent vertices by deleting the intermediate vertices along collinear segments.

Write $B = \{\{1,2\}, \ldots, \{N-1, N\}\}$ and let S be the complementary set of edges:

$$\{i,j\} \in S \text{ exactly when } \{i,j\} \notin B \text{ and } 1 \leq i \neq j \leq N.$$

A polygonal arc admits a *strict infinitesimal motion* if there are vectors v_i such that satisfy the inequalities.

$$\begin{array}{ll} (p_i - p_j) \cdot (v_i - v_j) & > 0 \text{ if } \{i,j\} \in S \\ (p_i - p_j) \cdot (v_i - v_j) & = 0 \text{ if } \{i,j\} \in B. \end{array}$$

A strict infinitesimal motion has the following interpretation. If p_i moves with constant velocity v_i, then its position at time t is $p_i + tv_i$. At time t, the points have distance

$$\|(p_i + tv_i) - (p_j + tv_j)\|^2 = \|p_i - p_j\|^2 + 2t(p_i - p_j) \cdot (v_i - v_j) + t^2\|v_i - v_j\|^2.$$

Under a strict infinitesimal motion, to first order, the segments of the polygonal arc maintain a fixed length and the other points move apart.

Proof. These conditions for a strict infinitesimal motion are linear and homogeneous in v_i. If a solution exists to this system of linear inequalities, we can rescale the velocities v_i so that there exists a solution where

$$(p_i - p_j) \cdot (v_i - v_j) \geq 1 \text{ if } \{i,j\} \in S$$

We can express the existence of a solution as saying that $m = 1$ is the maximum of the linear programming problem

$$\begin{array}{lrl} \max\ m & \text{subject to} & \\ & (p_i - p_j) \cdot (v_i - v_j) & \geq m \text{ if } \{i,j\} \in S \\ & (p_i - p_j) \cdot (v_i - v_j) & = 0 \text{ if } \{i,j\} \in B \\ & m & \leq 1. \end{array}$$

This system has a bounded objective function $m \leq 1$. It has a feasible solution $m = 0$ and $v_i = 0$ all i. It is not a linear program in the standard form described in DUALITY. However, it can be converted to standard form by the tricks at the end of DUALITY. The dual variables are indexed by constraints in the primal: w_{ij} for $\{i,j\} \in S \cup B$ and λ for the final primal equation $m \leq 1$. The dual is

$$\begin{array}{lrl} \min \lambda & \text{subject to} & \\ & \sum_{j:\{i,j\} \in S \cup B} w_{ij}(p_j - p_i) & = 0,\ \forall i \\ & w_{ij} & = w_{ji}, \\ & -\sum_{\{i,j\} \in S} w_{ij} + \lambda = 1. & \\ & w_{ij} & \geq 0, \{i,j\} \in S. \\ & \lambda \geq 0. & \end{array}$$

This is the dual in cleaned-up form, again using the tricks at the end of DUALITY. We will prove directly from the constraints that $w_{ij} = 0$ for all $\{i,j\} \in S$. It then follows from the dual constraints that $\lambda = 1$. By linear programming duality, it then follows that a strict infinitesimal motion exists.

Assume for a contradiction that a solution to these equations exists, with $w_{ij} > 0$ for some $\{i,j\} \in S$.

We convert the problem into a plane framework F as follows.

- Consider the set of segments $[p_i, p_j]$ in the plane, for $\{i,j\} \in B \cup S$. Define the set of vertices of F to be the set $\{q_k\}$ of points (including the original p_i) where two of these segments meet.
- The edges of the plane framework are $\{k, \ell\}$, where $[p_k, p_\ell]$ is a subset of one of the original segments $[p_i, p_j]$ that meets exactly two vertices of the framework (its endpoints).
- We call an edge $\{k, \ell\}$ a bar when it is contained in a segment $[p_i, p_j]$ of the original polygonal arc. Otherwise, it is a strut.

We can use w_{ij} to create a nonzero proper equilibrium stress ω_{kl} on the framework F. If $A\{k,\ell\}$ is the set of $\{i,j\} \in B \cup S$ such that $[p_k, p_\ell] \subset [p_i, p_j]$, then set

$$\omega_{k\ell} = -\sum_{\{i,j\} \in A\{k,\ell\}} w_{ij} \frac{\|q_k - q_\ell\|}{\|p_i - p_j\|}.$$

The weighting factor

$$\|q_k - q_\ell\| / \|p_i - p_j\|$$

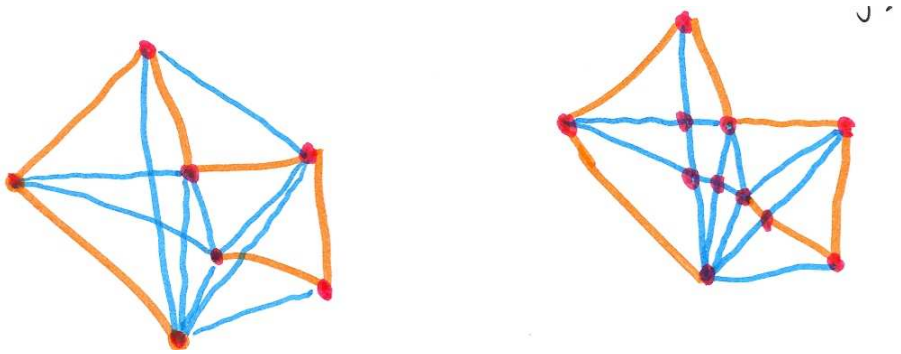


Figure 27.1. *Add vertices to create a planar framework.*

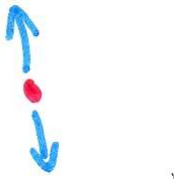


Figure 27.2. *The stresses on added vertices in opposite directions cancel, achieving equilibrium.*

is precisely the factor needed so that the constraints of the dual linear program imply the equilibrium stress conditions at each vertex q_k. The terms are all non-negative when $\{k, \ell\}$ is a stress (since $w_{ij} \geq 0$ for $\{i, j\} \in S$). Thus, the stress is proper.

By assumption, there exists $\{i, j\} \in S$ such that $w_{ij} > 0$. Since the original polygonal arc contains no unbent vertices, the segment $[p_i, p_j]$ is not entirely covered by bars of the framework. Thus, there exists a strut $\{k, \ell\}$ with $[p_k, p_\ell] \subset [p_i, p_j]$. For this index $\omega_{k\ell} < 0$. However, this is impossible by STRESS, which asserts that all struts carry zero stress: $\omega_{k\ell} = 0$.

This contradiction shows that the primal linear program has maximum $m = 1$, and hence that a feasible solution exists to the system of linear inequalities defining a strict infinitesimal motion. □

Chapter 28

Spherical Triangle Area

Theorem 28.1. *The area A of a spherical triangle T on a unit sphere, having angles α, β, γ is $A = \alpha + \beta + \gamma - \pi$.*

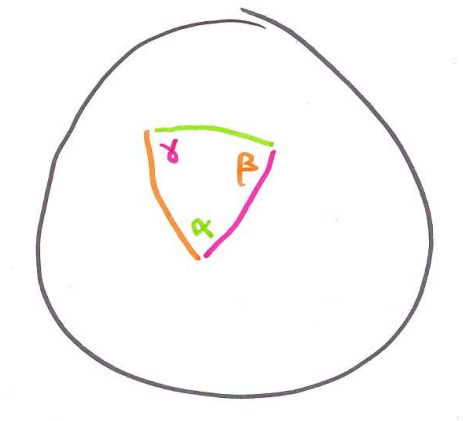


Figure 28.1. *Spherical Triangle*

Proof. Extend each edge to a great circle. The circles form an antipodal triangle T' with the same area.

The unit sphere has area 4π. A lune of angle α has area 2α, proportional to the angle. The three lunes of angles α, β, γ at the vertices of T and the three additional antipodal lunes of angles α, β, γ have total area

$$4\alpha + 4\beta + 4\gamma.$$

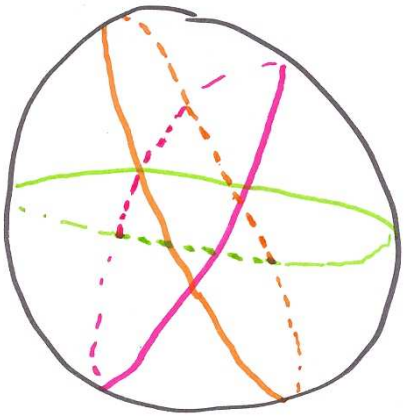


Figure 28.2. *Antipodal Triangle*

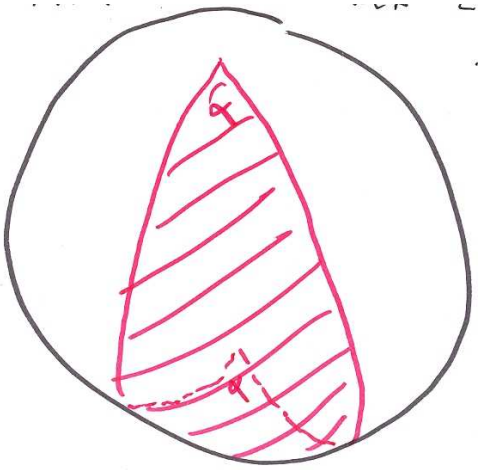


Figure 28.3. *A Lune*

They cover the sphere once except on the equal-area triangles T, T', which get covered three times:

$$4\alpha + 4\beta + 4\gamma = 4\pi + 4A.$$

Solve for A to get the conclusion. □

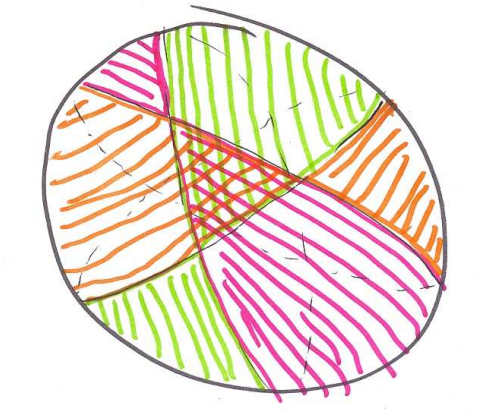


Figure 28.4. *Covering a sphere with lunes*

Chapter 29

Euler's Formula for Triangulations of the Sphere

Theorem 29.1. *Let T be a triangulation of the sphere into spherical triangles. Assume that in the triangulation there are V vertices, E edges, and F faces. Then*

$$V - E + F = 2.$$

Proof. Let A be the set of all angles of all the triangles in the triangulation. Let $\mathbf{V}$ be the set of vertices and $\mathbf{F}$ be the set of faces, so $V = \text{card}\mathbf{V}$, etc. The angles at a given vertex have sum 2π:

$$\sum_{\alpha\in A} \alpha = \sum_{v\in\mathbf{V}} \sum_{\alpha\sim v} \alpha = V(2\pi).$$

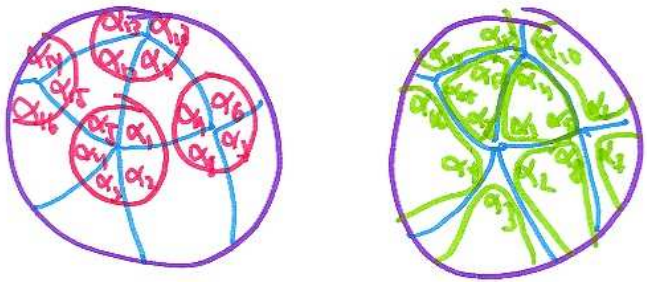


Figure 29.1. *Angles can be grouped by vertex or by face. The sum is the same either way.*

By HARRIOT's formula, the three angles α of a triangular face f satisfy

$$\text{area}(f) + \pi = \sum_{\alpha\sim f} \alpha.$$

The total area of all these triangles is full area 4π of the sphere, so

$$\sum_{\alpha\in A} \alpha = \sum_{f\in\mathbf{F}} \sum_{\alpha\sim f} \alpha = \sum_{f\in\mathbf{F}} (\text{area}(f) + \pi) = 4\pi + F\pi.$$

Solving for F, we get $F = 2V - 4$.

The total number of directed edges is twice the number of edges $2E$. The directed edges can be grouped in threes according to which triangle they appear in (with a counterclockwise orientation on each triangle). Thus,

$$2E = 3F.$$

We then have

$$V - E + F = \frac{F+4}{2} - \frac{3F}{2} + F = 2.$$

□

Chapter 30

Spherical Law of Cosines

The law of cosines in trigonometry states that if a triangles has sides a, b, c and if the angle opposite the side c is γ, then

$$c^2 = a^2 + b^2 - 2ab\cos\gamma.$$

This easy piece gives a proof of the spherical law of cosines, which is a formula for the angle of a spherical triangle in terms of the three sides.

[ENDNOTE] The law of cosines can be established by a coordinate system for which two sides of the triangle are given by non-zero vectors X and Y, of lengths a and b. We see that $X \cdot Y/(ab)$ is between -1 and 1 through the inequalities:

$$-1 \le -1 + \frac{1}{2}\left(\frac{X}{a} + \frac{Y}{b}\right)^2 = \frac{X \cdot Y}{ab} = 1 - \frac{1}{2}\left(\frac{X}{a} - \frac{Y}{b}\right)^2 \le 1.$$

Thus, $X \cdot Y/(ab)$ is the cosine of some γ:

$$X \cdot Y = ab\cos\gamma.$$

This serves as the definition of the angle γ. Then the law of cosines follows as a vector identity:

$$c^2 = (X - Y) \cdot (X - Y) = X \cdot X + Y \cdot Y - 2X \cdot Y = a^2 + b^2 - 2ab\cos\gamma.$$

We begin with some vector identities in $\mathbb{R}^3$.

Lemma 30.1. *Let $A, B, C \in \mathbb{R}^3$. Then*

- $A \cdot (B \times C) = C \cdot (A \times B)$.
- $(A \times B) \times C = B(A \cdot C) - A(B \cdot C)$.

Proof. Let $A = (a_1, a_2, a_3)$, $B = (b_1, b_2, b_3)$, and $C = (c_1, c_2, c_3)$ and compute both sides. □

Well, that is a proof, but there are certainly more elegant ways to prove these identities. For instance, we can remark that both sides of the identities are linear in A, B, and C, so that it is enough to check both sides when the vectors are elements of an orthonormal basis $\{e_1, e_2, e_1 \times e_2\}$.

Lemma 30.2. *Let $A, B \in \mathbb{R}^3$ be unit vectors. Suppose that $A \cdot B = \cos(c)$, for $0 \leq c \leq \pi$. Then*

$$\|A \times B\| = \sin(c).$$

Proof. Since both sides are nonnegative, it is enough to check that the squares of the two sides are equal. We use the previous lemma.

$$\begin{aligned} \|A \times B\|^2 &= (A \times B) \cdot (A \times B) \\ &= B \cdot ((A \times B) \times A) \\ &= B \cdot (B(A \cdot A) - A(B \cdot A)) \\ &= 1 - (A \cdot B)^2 = 1 - \cos^2(c) = \sin^2(c). \end{aligned}$$

□

Theorem 30.3 (Spherical Law of Cosines). *Let T be a spherical triangle with sides of length a, b, c. Then the angle γ between the sides of length a, b satisfies*

$$\cos\gamma = \frac{\cos c - \cos a \cos b}{\sin a \sin b}.$$

Proof. Take unit vectors A, B, C to the vertices of T. The angle γ equals the angle between the planes through the origin spanned by (C, A) and (C, B) respectively. Equivalently, it is the angle between the normal vectors $C \times A$ and $C \times B$. Hence

$$\cos\gamma = \frac{(C \times A) \cdot (C \times B)}{\|C \times A\| \, \|C \times B\|}.$$

We have $C \cdot A = \cos b$, and $\|C \times A\| = \sin b$. Similarly, $C \cdot B = \cos a$, and $\|C \times B\| = \sin a$. This gives the denominator of the law of cosines.

Consider the numerator. Use use the lemma again to finish the proof.

$$\begin{aligned} (C \times A) \cdot (C \times B) &= B \cdot ((C \times A) \times C) \\ &= B \cdot ((A(C \cdot C) - C(A \cdot C))) \\ &= (A \cdot B) - (B \cdot C)(A \cdot C) \\ &= \cos c - \cos a \cos b. \end{aligned}$$

□

Chapter 31

Polar Triangles

A unit vector v determines a point on the unit sphere S^2. It also determines a great circle

$$\{x : x \cdot v = 0\}.$$

The unit vector $-v$ determines the same great circle. This gives a bijection between great circles and the corresponding pair of poles $\{v, -v\}$ on the unit sphere. For example, the equator determines the north and south poles.

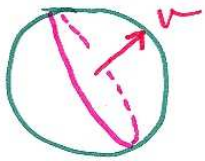


Figure 31.1. *A vector v determines a great circle and its poles $\pm v$.*

Let A, B, C be three points on the unit sphere, not all on the same great circle. The geodesic arc from A to B is the arc of a great circle from A to B of length less than π. Joining each pair AB, BC, CA by geodesic arcs, we obtain a spherical triangle ABC.

The great circle through A and B has poles $\pm C'$ on the unit sphere. We choose the sign $\pm$ so that C is closer to C' than to $-C'$. (The distances from C to C' and $-C'$ cannot be equal because of the non-collinearity constraint on ABC.) Pick A' and B' as the similarly chosen poles of the great circles BC and AC. By connecting A', B', and C' by geodesic arcs, we obtain a spherical triangle $A'B'C'$, called the polar triangle of ABC.

We claim that the polar triangle of $A'B'C'$ is ABC, again. In fact, the arc distance from A to B' and C' is $\pi/2$. Hence, $\pm A$ are the poles of the great circle

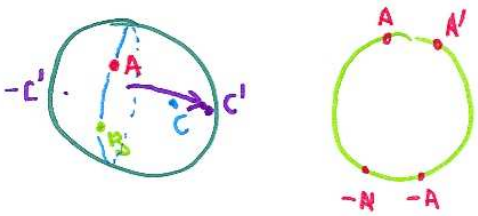


Figure 31.2. *The choice of pole C' makes C and C' closer to each other than $\pm C$ and $\mp C$.*

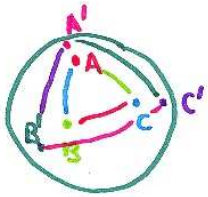


Figure 31.3. *Triangles ABC and $A'B'C'$ are polar triangles of one another.*

through B' and C'. Since, A is closer to A' than to $-A'$, we also have that A' is closer to A than $-A$. Thus, $A'' = A$ is the chosen pole determined by $B'C'$. Similarly, the great circles through $C'A'$ and $A'B'$ determine the poles B and C. Thus, we obtain a duality of spherical triangles, each triangle paired with its dual.

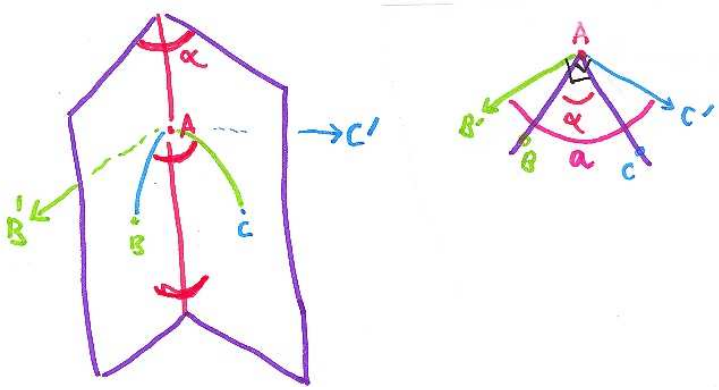


Figure 31.4. *The angle at α equals the angle formed by the planes through the origin containing AC and AB. The angle a is the angle between the normals to these planes. The angles α and a are related by $\alpha + a = \pi$.*

Name the lengths a, b, c and angles α, β, γ of triangle ABC, as well as the lengths a', b', c' and angles α', β', γ' of triangle $A'B'C'$. The diagram shows that

$\alpha + a' = \pi$. (The second frame shows the projection of the sphere to the plane perpendicular to the line through A and the origin. This projection preserves the magnitude of the angles α and a'.) Similarly, we obtain the following relations:

$$\begin{array}{llll} \alpha + a' & = \pi & \alpha' + a & = \pi \\ \beta + b' & = \pi & \beta' + b & = \pi \\ \gamma + c' & = \pi & \gamma' + c & = \pi \end{array}$$

As an example of what can be done with polar triangles, we derive the polar form of the spherical law of cosines.

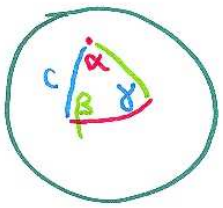


Figure 31.5. *Polar triangles give a dual form to the law of cosines, expressing edge lengths as a function of angles.*

Corollary 31.1. *Let a triangle have angles α, β, γ and side of arc length c opposite γ. Then*

$$\cos c = \frac{\cos\gamma + \cos\alpha\cos\beta}{\sin\alpha\sin\beta}.$$

Proof. The polar triangle satisfies the spherical law of cosines

$$\cos\gamma' = \frac{\cos c' - \cos a'\cos b'}{\sin a'\sin b'}.$$

Substitute $a' = \pi - \alpha$, $b' - \pi - \beta$, etc. for the primed quantities, and use the trig identities $\cos(\pi - x) = \cos(x)$ and $\sin(\pi - x) = \sin(x)$. The result follows immediately. □

Chapter 32

Lexell's Theorem

In the plane, we fix a point A and two rays R_B and R_C emanating from A. Fix a constant $p > 0$. We consider the locus of all triangles ABC with fixed perimeter p and vertex B on R_B and C on R_C.

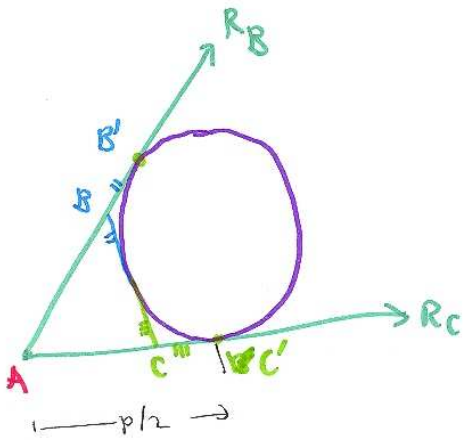


Figure 32.1. *Constant perimeter triangles have an edge BC tangent to the displayed circle.*

We show that such triangles ABC are in bijection with the arc of a circle in the plane tangent to both R_B and R_C. We construct the circle as follows. Let S be the circle that is tangent to R_B and R_C, meeting both R_B and R_C at points B', C' at distance $p/2$ from A.

The desired arc is the arc of S with endpoints at B', C', choosing the branch of the arc that passes closest to A. If B on R_B and C on R_C are now any two points joined by a segment tangent to this arc, the perimeter of the triangle is p.

In fact, since the tangents to a circle are equal in length, the segment BC is equal in length to the sum of the lengths BB' and CC'. Thus, the perimeter is equal to the sum of the lengths AB, BB', AC, CC', which is p.

If we move this construction from the plane to the unit sphere, and then dualize by replacing the triangle by its polar triangle, we obtain Lexell's theorem:

Theorem 32.1 (Lexell). *The locus of points A' that give the a spherical triangle $A'B'C'$ of a given area on a fixed base $B'C'$ is an arc of a circle whose endpoints are the antipodal points $-B'$ and $-C'$. The center of the circle lies on the perpendicular bisector of $B'C'$.*

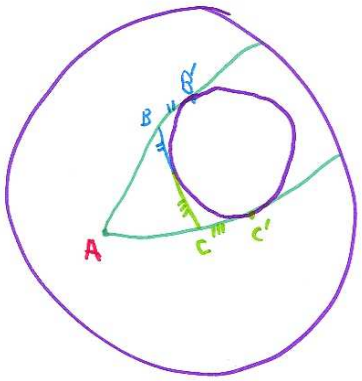


Figure 32.2. *Constant perimeter spherical triangles have an edge BC tangent to the displayed circle.*

Proof. Fix a lune formed by two great half circles R_B and R_C meeting at vertices $\pm A$. Fix a constant $p > 0$. Draw the circle S on the unit sphere, inscribed in the lune, tangent to both R_B and R_C, with each point of tangency at arc length $p/2$ from A. Select the arc of the circle, closest to A, with endpoints at the points of tangency. As in the planar case, if B on R_B and C on R_C are any two points joined by a segment tangent to this arc, then the triangle ABC has perimeter p.

Now consider the polar triangle $A'B'C'$. If the perimeter of ABC is $p = a + b + c$, then the area of $A'B'C'$ is given by Harriot's formula:

$$\alpha' + \beta' + \gamma' - \pi = (\pi - a) + (\pi - b) + (\pi - c) - \pi = 2\pi - p.$$

Thus, as ABC varies over triangles of equal perimeter, $A'B'C'$ varies over triangles of equal area. The triangles ABC have fixed angle α formed by the half circles R_B and R_C. Dually, the polar triangles $A'B'C'$ have fixed side $a' = \pi - \alpha$ on a fixed base $B'C'$.

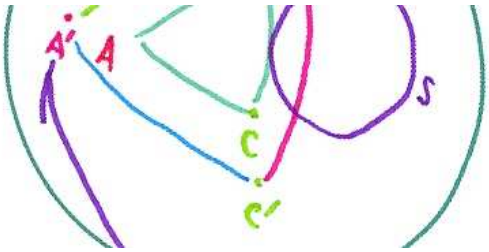


Figure 32.3. *The third vertex A' of the polar triangle sweeps over a circle at distance $\pi/2$ from the "constant perimeter" circle.*

For any great circle through BC tangent to S, its pole A' lies on a circle S' at distance $\pi/2$ from S. One endpoint of the given arc on S is the point of tangency of the great circle AB with S. Dually, one endpoint of the given arc through S' is the dual point $-C'$ on S'. Similarly, the other endpoint is $-B'$. The segment from A to the center of S bisects the rays R_B, R_C. Dually, the center of the circle S' lies on the perpendicular bisector of $B'C'$. □

Chapter 33

Tarski's Planks

The following result is more of a math puzzle than a theorem, but the solution is so slick, that I am unable to resist the temptation to include it.

Theorem 33.1. *Let strips of paper of various widths be given, all of constant width and infinite length. Suppose that they can be arranged to cover a disk of unit radius. Then the sum of their widths is at least* 2.

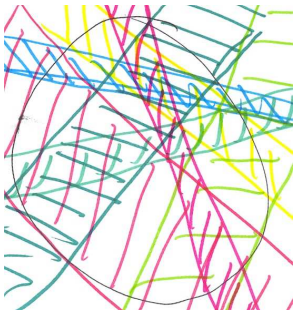


Figure 33.1. *Covering a disk with strips*

In other words, you cannot cover a disk more efficiently than with strips in parallel.

Proof. Set a hemisphere of the same diameter on the disk and view each strip as the shadow (vertical projection) of a curved strip bounded by upright circular arcs on the hemisphere.

If the flat strip on the disk has width d, the area of the curved strip on the sphere is $2\pi d$. (On the hemisphere it is half: πd.) This follows from a basic formula

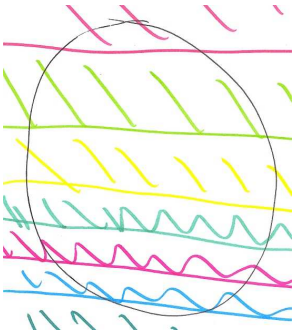


Figure 33.2. *A parallel cover by strips*

in calculus for the area of a surface of revolution:

$$\text{AREA} = \int_a^b 2\pi y\sqrt{1+y'^2}dx.$$

In our case $x^2 + y^2 = 1$ and $y' = -x/y$. A calculation shows that

$$y\sqrt{1+y'^2} = \sqrt{y^2+y^2y'^2} = \sqrt{y^2+x^2} = 1,$$

so that if $b - a = d$, we have

$$\text{AREA} = \int_a^b 2\pi dx = 2\pi d.$$

The key to the proof is that this value is independent of where on the sphere it lands. It depends only on the width of the plank.

The curved strips cover the hemisphere, hence have total area at least that of the hemisphere:

$$\begin{array}{ll} \sum \pi d_i & \geq 2\pi \\ \sum d_i & \geq 2. \end{array}$$

□

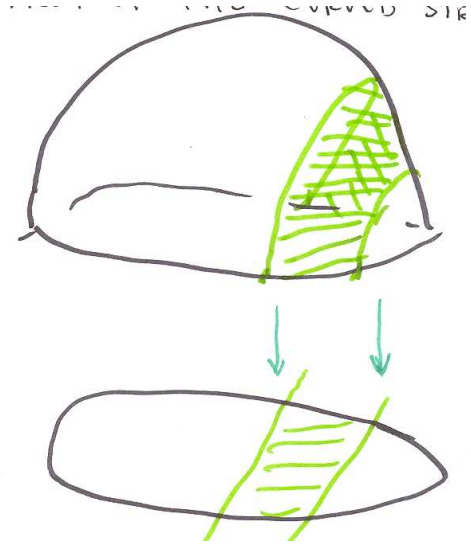


Figure 33.3. *Projecting from hemisphere to disk*

Figure 33.4. *Covering the hemisphere*

Chapter 34

Flat Torus Geometry

A flat torus is a quotient $\mathbb{R}^2/(\mathbb{Z}v_1 + \mathbb{Z}v_2)$ for two linearly independent vectors v_1 and v_2.

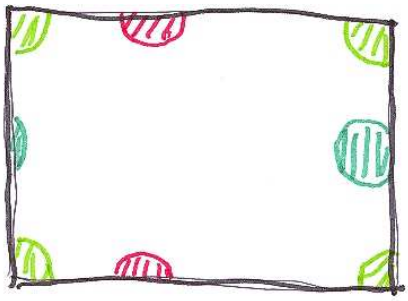


Figure 34.1. *An object dragged through the top of a flat torus screen reappears at the bottom of the screen.*

We can picture the flat torus in two ways. There first is to picture it as a computer screen, programmed so that any object that is dragged off the screen to the right reappears again on the left of the screen, and an object that is dragged off the top of the screen reappears at the bottom, and so forth. Thus, we can move objects indefinitely in all directions but we never leave the confines of the computer screen. When we represent disks on the computer screen, only finitely many will fit, with some of them possible extending beyond an edge of the screen to the opposite side of the screen. (To accommodate a general flat torus, we must allow the computer screen to be a non-rectangular parallelogram.)

The second way to picture a flat torus is as a plane partitioned into parallelograms. Any object placed in one parallelogram is replicated infinitely many times as translated copies in every parallelogram. As an object is dragged, all the

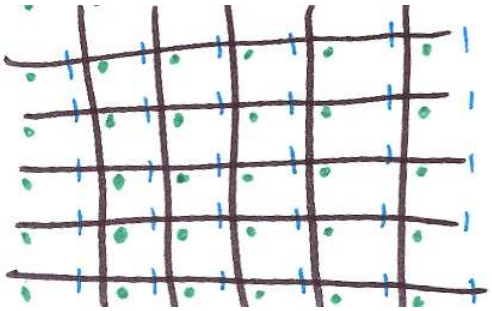


Figure 34.2. *Every object on the flat torus lattice appears infinitely often.*

translates of the object are dragged in unison. As an object is dragged from one parallelogram to the next one over, all the translates move to the corresponding next parallelogram over.

As a matter of general principle, the study of questions in geometry on a flat torus is equivalent to the study of questions in the plane constrained by a periodicity condition.

Chapter 35

Euler Formula in Flat Torus Geometry

Theorem 35.1. *Suppose that there is a graph with polygonal edges on a flat torus that partitions it into polygonal faces. Let V, E, F be the number of vertices, edges, and faces. Then*

$$V - E + F = 0.$$

Proof. We compute the sum of the angles $S = \sum \alpha$ of all of the polygonal faces in two ways. If we arrange angles according to the vertex they lie at, we get get a contribution of angle 2π around each vertex. Thus, $S = 2\pi V$.

If the face has n vertices, we show by induction the well-known result that the sum of the angles of that face is $(n-2)\pi$. For $n = 3$, the fact that the angles of a triangle sum to π is well known. This is the base case of the induction. For a general polygon, draw a diagonal (the existence of which is known from TRIANGULATION) breaking the polygon into a $(r+1)$-gon of angle sum $(r-1)\pi$ and an $(s+1)$-gon of angle sum $(s-1)\pi$, where $r + s = n$. Hence the angle sum for the original n-gon is $(s-1)\pi + (r-1)\pi = (n-2)\pi$, completing the induction.

If we arrange angles according to the polygonal face, we get a second expression for S, as

$$S = \sum_{i=1}^{F} (n_i - 2)\pi.$$

As we sum $\sum n_i$ each edge is counted twice. Hence

$$S = 2E\pi - 2F\pi.$$

Equating the two expressions for S, we get Euler's formula for a flat torus. □

In the proof, we use the fact that $\sum_{i=1}^{F} n_i = 2E$. This fact will be useful in other chapters. Since each polygon has at least three sides, we have $n_i \geq 3$. This gives as a consequence that $\sum n_i \geq 3F$, so that

$$2E \geq 3F$$

with equality holding exactly when the graph is a triangulation.

Chapter 36

Thue's Theorem

Theorem 36.1. *A (periodic) packing of disks of radius* 1 *in the plane cannot have density exceeding* $\pi/\sqrt{12}$.

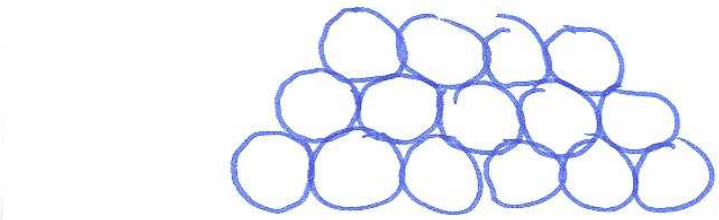


Figure 36.1. *The hexagonal packing has optimal density.*

The disks are allowed to touch, but they are not allowed to share an interior point. An example of a packing of disks that achieves the optimal bound $\pi/\sqrt{12}$ is the hexagonal packing.

Figure 36.2. *The density of the hexagonal packing is the ratio of the area of a disk to the area of its circumscribing hexagon.*

To avoid questions of what the density means, we restrict our attention to periodic packings; that is, to packings of on a flat torus. The density of a packing on a flat torus is just the ratio of the total area of the disks to the area of the flat

torus. (We mention as an aside, that the general result for non-periodic packings actually follows from the periodic case, since for any $\delta' < \delta$, a non-periodic packing of density δ achieves density δ' by the disks entirely contained in sufficiently large rectangles. Such rectangles can be repeated to a periodic structure of density δ'.)

The proof depends on the magical constant $\rho = 2/\sqrt{3}$ (which just happens to be the radius of a regular hexagon with inradius 1).

Lemma 36.2. *no point $q \in \mathbb{R}^2$ is within ρ of three disk centers p_1, p_2, p_3.*

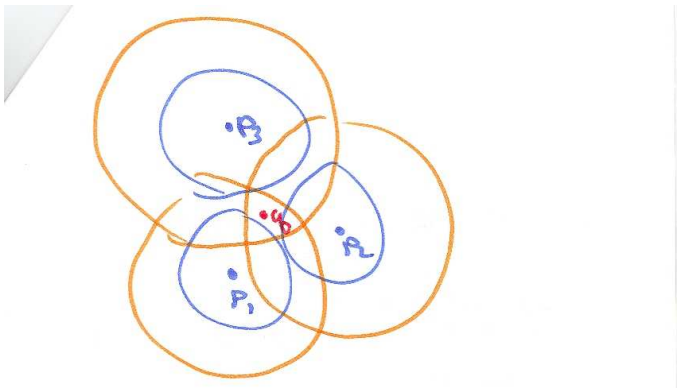


Figure 36.3. *An impossible situation.*

Proof. For a contradiction, assume $\|q - p_i\| < \rho$, for $i = 1, 2, 3$. One of the triangles (q, p_i, p_j) has angle $\alpha \leq 2\pi/3$ (and hence $\cos\alpha \geq -1/2$) at q (since the three angles at q sum to 2π).

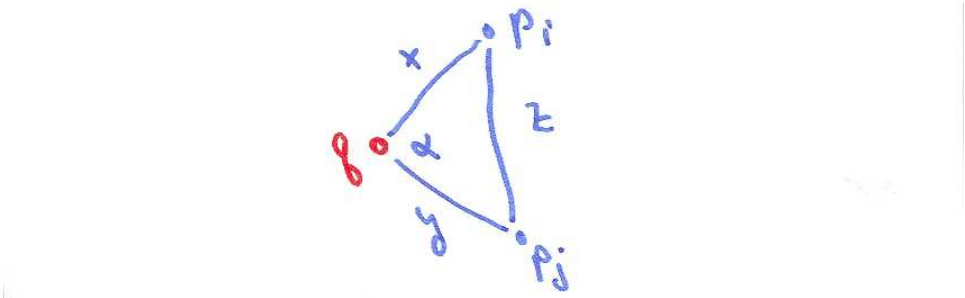


Figure 36.4. *The law of cosines.*

The law of cosines gives the desired contradiction.

$$4 \leq z^2 = x^2 + y^2 - 2xy\cos\alpha \leq x^2 + y^2 + xy < 4/3 + 4/3 + 4/3 = 4.$$

□

Now we prove Thue's theorem.

Figure 36.5. *The plane can be partitioned into three regions, the purple, orange, and green.*

Proof. We partition $\mathbb{R}^3$ into three types of regions.

- The (purple) region outside all disks of radius ρ.
- The (orange) circular sectors formed by the nonoverlapping parts of disks of radius ρ.
- The (green) rhombus regions formed by overlapping disks of radius ρ.

The purple region contains no disks in the packing, so its density is 0. Clearly $0 < \pi/\sqrt{12}$.

The orange regions have density equal to the ratio of the areas of the unit disk to the disk of radius ρ:

$$\frac{\pi}{\pi\rho^2} = \frac{3}{4} < \frac{\pi}{\sqrt{12}} \approx 0.9.$$

So the only place the density can be high is in a green rhombus. Its density depends on the angle θ. The angle is constrained $0 \leq \theta \leq \pi/3$ (for otherwise, $x = \rho\cos(\theta/2) < 1$ and the disks of the packing have illegal overlap). The density

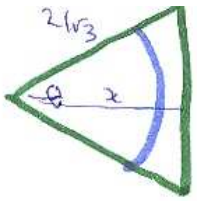


Figure 36.6. *The density within a green triangle is the ratio of the area of the disk sector to the area of the triangle.*

is the ratio of the area of the portion of the disk in the triangle to the area of the triangle:

$$\delta(\theta) = \frac{\theta}{\rho^2 \sin\theta}.$$

The derivative is

$$\delta'(\theta) = \rho^{-2} \csc^2\theta(\tan\theta - \theta).$$

The term $\tan\theta - \theta$ can be interpreted as an area, so the term and also the derivative is non-negative. So

$$\delta(\theta) \leq \delta(\frac{\pi}{3}) = \frac{\pi}{\sqrt{12}}.$$

□

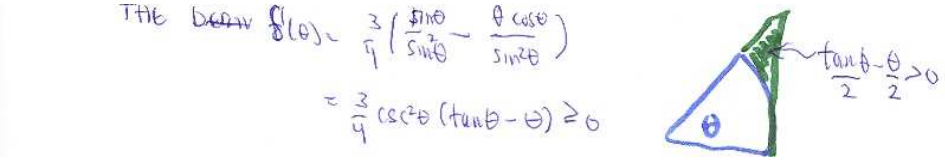


Figure 36.7. $(\tan\theta - \theta)/2$ *is an area, hence non-negative.*

[SIDE NOTE] The density, properly defined in the plane, is a limit of densities obtained by finite packings inside larger and larger regions of the plane. In keeping with the objective of this book to be as elementary as possible, we avoid the limit, by dealing with packings in a compact space, rather than packings in the entire plane. For this reason we have turned to flat torus geometry. [/SIDE NOTE]

Chapter 37

Optimal Lattice Packings

A lattice in space is a set of the form

$$\Lambda = \{n_1 v_1 + n_2 v_2 + n_3 v_3 : n_1, n_2, n_3 \in \mathbb{Z}\},$$

where v_1, v_2, and v_3 are linearly independent vectors in three dimensional Euclidean space. For example, the cubic lattice is obtained from the standard vectors $v_1 = (1, 0, 0)$, $v_2 = (0, 1, 0)$, $v_3 = (0, 0, 1)$.

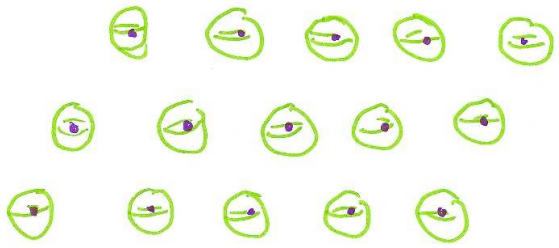


Figure 37.1. *A lattice packing.*

If every point of Λ has distance at least 2 from every other point of Λ, then we can center a ball of unit radius at each point of Λ. Some balls will be tangent to one another if the minimum distance separating two points of Λ is exactly 2. Such a packing of congruent balls is called a *lattice packing*. Such packings are atypical of a general packing of balls, because the entire lattice packing is entirely determined by the four balls centered at the origin, v_1, v_2, and v_3. We assume throughout this chapter that the points of Λ are separated by distance at least 2.

The density of the packing is the fraction of spaced filled by the balls. Since the structure is periodic, the density of the packing is the same as the density within the fundamental parallelopiped:

$$P = \{x_1 v_1 + x_2 v_2 + x_3 v_3 : x_1, x_2, x_3 \in [0, 1]\}$$

(The translates of this parallelopiped give a tiling of space.) The pieces of a ball at each of the eight corners of the parallelopiped can be reassembled into a single ball of radius 1. Thus, the density of the packing is simply

$$\frac{4\pi/3}{\mathrm{vol}(P)}.$$

Theorem 37.1 (Gauss). *The density of a lattice packing is never greater than that of the face-centered cubic lattice packing.*

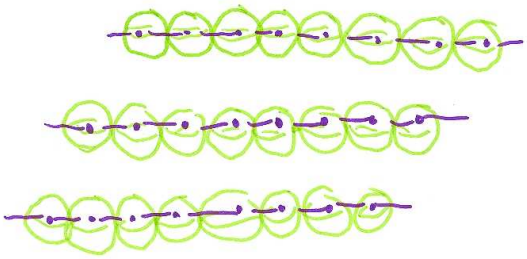
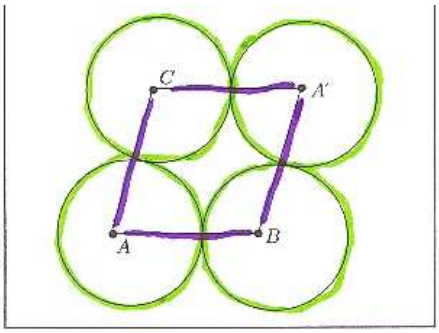


Figure 37.2. *In optimal lattice packings there must be contact between some pair of balls. The lattice property forces them to touch in long parallel strings.*

Proof. We start with an arbitrary lattice packing. The packing of highest density will certainly have the property that two of the balls touch one another. Once two balls touch, the lattice constraint forces the balls to touch along long parallel strings of balls, like a thick row of marshmallows on a roasting stick. In the best case it will also be certainly true that two of the long parallel beaded strings will touch. The lattice constraint forces the balls to be laid out in identical parallel plates.



touches all three
centers is equila
We now chan
of the balls as an
In each of those
peat the pattern
The balls of one l
ets of the layer l
three below it.
face-centered cu

Thue

The two-dimensi

Figure 37.3. *The centers of four balls form a parallelogram.*

The centers of four balls in a plate form a parallelogram. The parallel plates

should be set one on the other so that the plates are as close as possible. A ball D of the next layer is set in the pocket between three balls A, B, and C in the layer below so that it touches all three. The triangle ABD formed by the centers is equilateral.

We now change our point of view. We view all of the balls as arranged in planes parallel to ABD. In each of those layers the centers of the balls repeat the pattern of the equilateral triangles ABD. The balls of one layer should be nestled in the pockets of the layer below so that each ball rests on three below it. The lattice this describes is the face-centered cubic. □

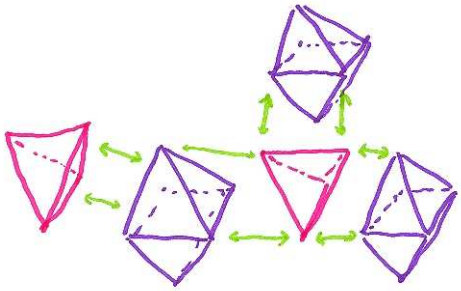


Figure 37.4. *The triangle based pyramid packing and the square based pyramid packing are the same lattice packing viewed from different orientations. In fact, oriented properly, the tetrahedra and octahedra tile space.*

The proofs indicates that the face-centered cubic lattice is the unique lattice packing with hexagonal layers nestled as closely as possible upon each other. Both the triangle based pyramid and the square based pyramid have this structure, thus they are isomorphic lattices, although they are oriented differently in space.

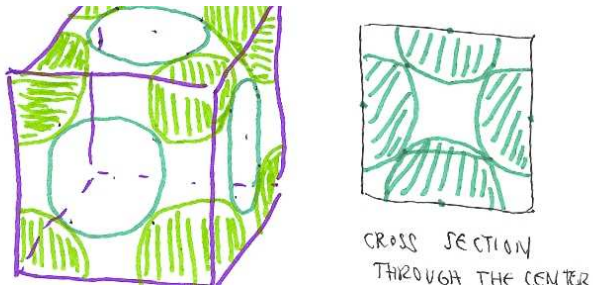


Figure 37.5. *The lattice packing is called the face-centered cubic lattice packing because it can be obtained from a cube by adding balls at each vertex and at the center of each face.*

The face-centered cubic packing can be obtained from the cubic lattice by placing an additional sphere at the center of each face. To accommodate the all these spheres without overlap, the edge of the cube must have length $2\sqrt{2}$. Its volume is $16\sqrt{2}$. Within each cube, there are 8 octants of balls at the vertices, which can be reassembled into a single ball, and 6 halves of balls at the faces, which give another 3 balls. The density of the face-centered cubic lattice packing is then

$$\frac{(4\pi/3)(4)}{16\sqrt{2}} = \frac{\pi}{\sqrt{18}} \approx 0.74.$$

Chapter 38

Delaunay Polygons

Consider a finite set of points X in the plane that are not all collinear. A Delaunay polygon is a convex polygon with vertices at some of the given points such that

1. the vertices consist of all points from the finite set on a circle, and

2. no point from the finite set lies interior to the circle.

Rather than state a theorem about Delaunay polygons, we will describe some of their properties. These properties will imply that Delaunay polygons give a partition the convex hull of X.

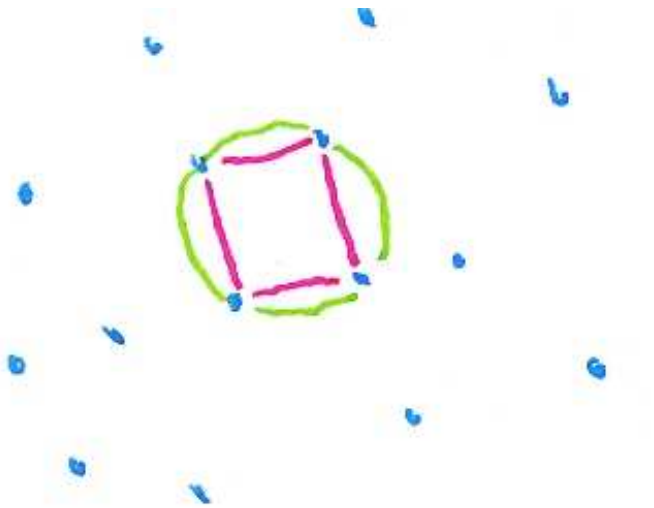


Figure 38.1. *A Delaunay polygon*

At least one Delaunay polygon exists. In fact, the circle whose diameter joins the two closest points of X has no points of X in its interior. The pencil of circles passing through these two points of X sweeps until a circle encounters a further point of X. The points of X on that circle form a Delaunay polygon.

Along a give edge of Delaunay polygon, there is a second Delaunay polygon, unless that edge is an edge of the convex hull of X. In fact, the pencil of circles passing through the two endpoints of the edge sweep through the half plane bounded by the line through the given edge. Eventually, the sweeping circle encounters a point and stops its sweep.

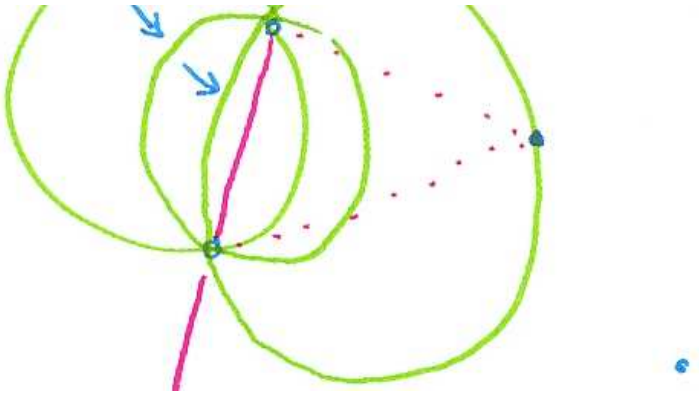


Figure 38.2. *Sweeping toward a new Delaunay polygons*

Two Delaunay polygons never intersect. In fact, the vertices for polygon A cannot lie interior to the circle B, hence they lie to one side of the dividing line. The vertices for B likewise lie to the other side of (or on) the line.

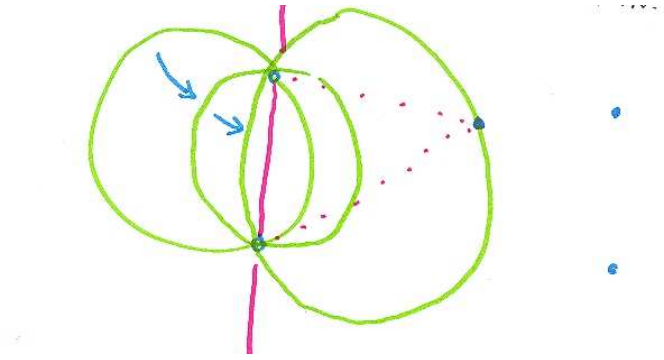


Figure 38.3. *Separation of polygons*

Figure 38.4. *Existence of a Delaunay polygons*

Every Delaunay polygon lies in the convex hull of X since the vertices of the polygon do.

Every point of the convex hull C of X lies in the interior or on the boundary of at least one Delaunay polygon. If not, pick $a \in C$ outside all Delaunay polygons. Draw a line segment from a to (a point other than an endpoint on) an edge e (not on the boundary of C) of a Delaunay polygon, such that the segment except its endpoints lies outside all Delaunay polygons. This is impossible, as there are Delaunay polygons along both sides of e.

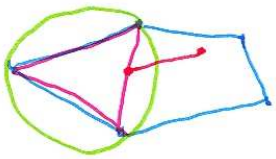


Figure 38.5. *Covering by Delaunay polygons*

Delaunay polygons thus partition C. If we triangulate each Delaunay polygon, we get what is called a Delaunay triangulation.

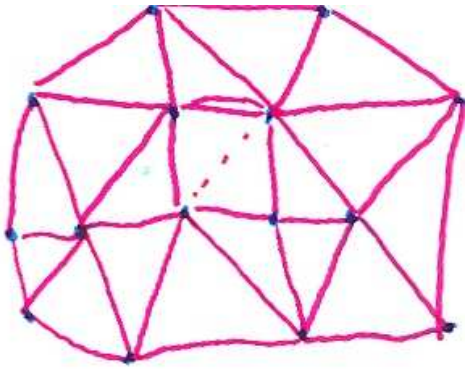


Figure 38.6. *Example*

Chapter 39

Spherical Delaunay Polygons

Take a finite set of points on a sphere not all lying in a hemisphere. The convex hull of the points in three dimensions is a polytope.

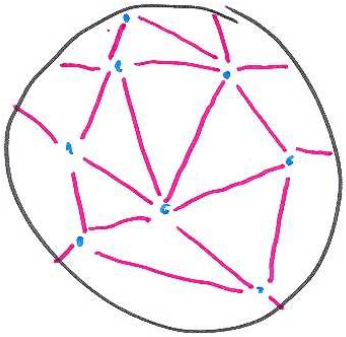


Figure 39.1. *Points on a sphere*

The faces of the polyhedron project to spherical polygons on the unit sphere that partition the sphere. These are the *spherical Delaunay polygons*. The plane through the face of the polyhedron cuts the sphere in a circle that contains the vertices of a spherical Delaunay polygon and none of the other points of the finite set in its interior.

The Delaunay decomposition of the plane can be viewed as taking the convex hull of points on a sphere in the limiting case of a sphere of curvature tending to zero.

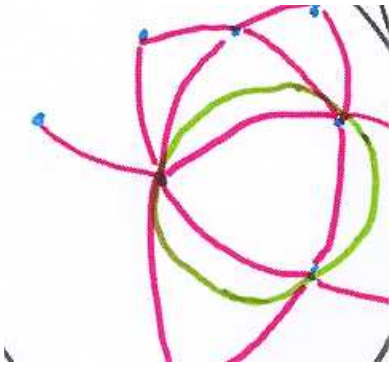


Figure 39.2. *The circumscribing circle of a polygon*

Chapter 40

Covering the Plane with Congruent Disks

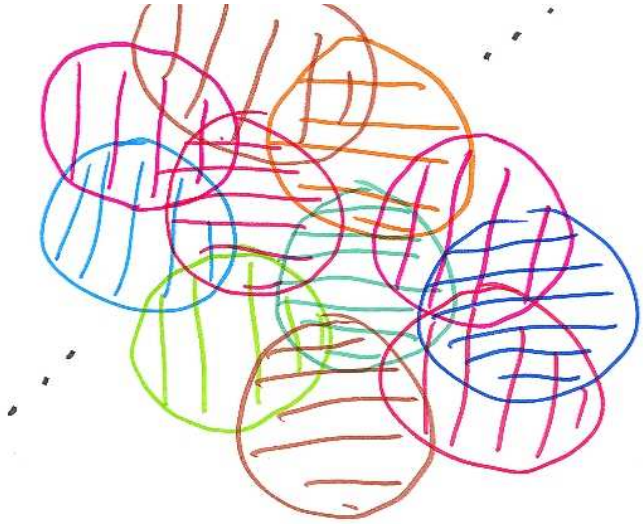


Figure 40.1. *Covering the plane with disks.*

Disks of unit radius *cover the plane* if each point of the plane lies in at least one disk. This section studies the question of determining the most efficient covering possible, in the sense of requiring the fewest disks per unit area. To avoid the complication of defining what the phrase *fewest disks per unit area* means, we restrict our attention in this chapter to coverings of the flat torus. On the flat torus we seek to minimize the ratio

$$\frac{\text{the number of disks}}{\text{the area of the flat torus}}$$

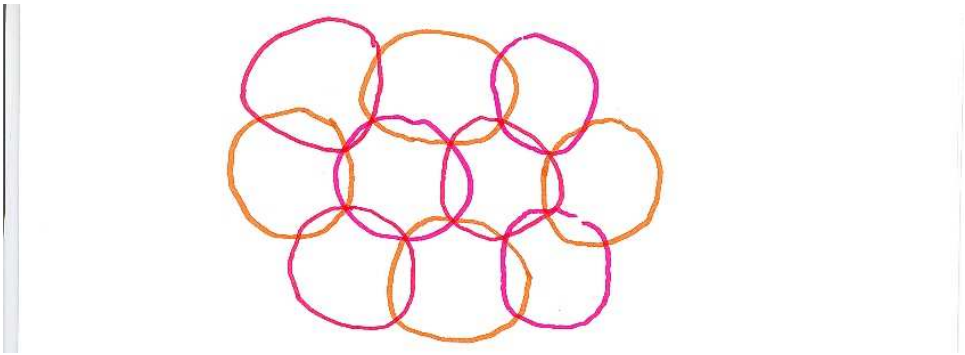


Figure 40.2. *The hexagonal covering.*

Theorem 40.1. *This covering ratio is never less than it is for the regular hexagonal covering.*

Proof. Take the Delaunay triangulation of the centers of the disks. (The results about Delaunay triangulations extend readily from finite subsets of the plane, to finite subsets of a flat torus.)

We claim that the total number of disks V is equal to half the number F of triangles. If the number of vertices, edges, and faces of a triangulation are V, E, F, then $V - E + F = 0$ and $2E = 3F$ [FLAT TORUS]. These relations imply $F = 2V$.

The area of the flat torus is the sum of the areas of the triangles. Thus, the covering ratio c satisfies

$$c = \frac{\text{the number of triangles}}{2\Sigma \text{ areas of triangles}} \geq \frac{1}{2\text{max. area of a triangle}}.$$

The proof then reduces to a calculation of the Delaunay triangle of maximum area that occurs in a covering. The circumcenter of each triangle is covered by a unit disk whose center lies outside the circumscribing circle of the triangle. So the circumradius of the triangle is at most 1. A triangle contained in a circle of radius 1 has area at most the area A of an equilateral triangle inscribed in the circle. We conclude that the covering ratio is at least

$$c \geq \frac{1}{2A}.$$

The right hand side is precisely the covering ratio of the regular hexagonal covering. (In fact, all its Delaunay triangles are equilateral with circumradius 1.) □

From an analysis of the case of equality in the proof, we see that the regular hexagonal covering is the only possible periodic structure that minimizes the covering ratio. (Even this does not exist on a flat torus unless the periods of the torus are commensurate with the periods of the regular hexagonal covering.)

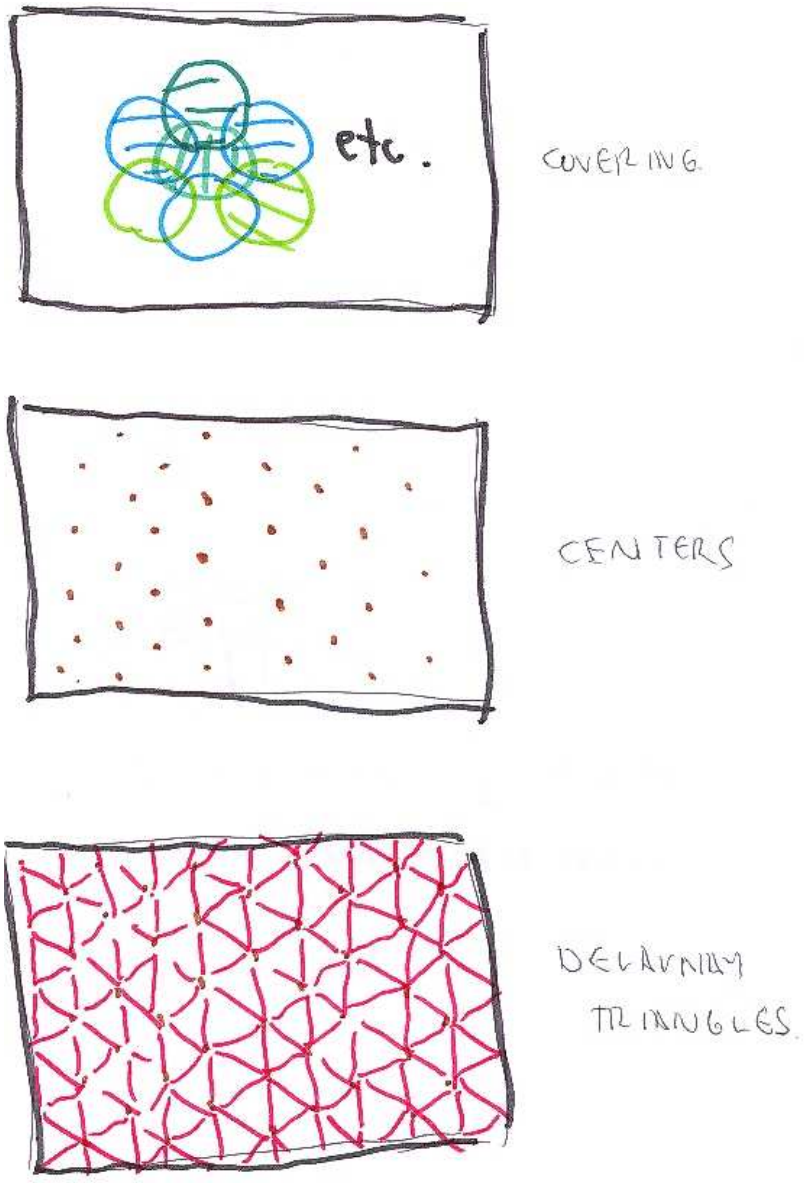


Figure 40.3. *The Delaunay triangulation of a covering.*

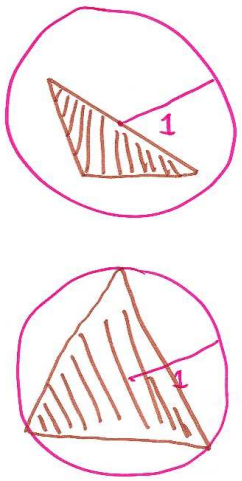


Figure 40.4. *The triangle of maximum area in a disk is an equilateral triangle.*

Chapter 41

Straight Lines Minimize Length

Hausdorff gave a definition of length that works for any subset of the plane. The length of a set is either a non-negative real number or infinity.

Theorem 41.1. *The Hausdorff length of any connected set in the plane containing x and y is at least the Hausdorff length of the line segment from x to y. Moreover, that distance is equal to the Euclidean distance from x to y.*

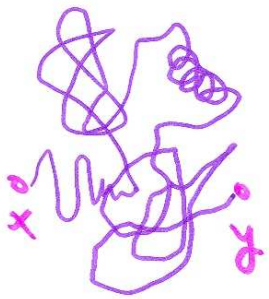


Figure 41.1. *A connected set containing x and y.*

Before diving into the proof, we recall the definition. The diameter of a bounded set $D \subset \mathbb{R}^2$ is

$$d(D) = \sup\{\|x - y\| : x, y \in D\}.$$

The Hausdorff length of a set X is

$$\mathcal{H}^1(X) = \lim_{\delta \mapsto 0} \inf_{X \subset \cup D_j,\ d(D_j) \leq \delta} \sum d(D_j).$$

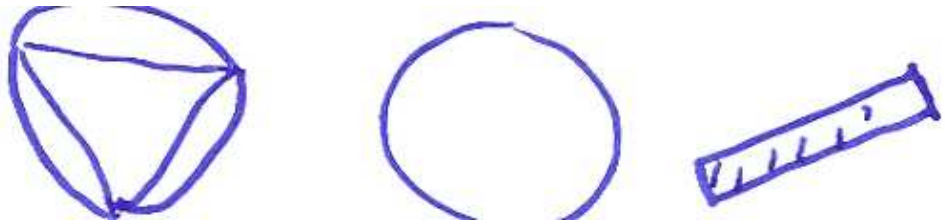


Figure 41.2. *Sets of diameter* 1.

That is, for each $\delta > 0$, we cover X by a countable collection of bounded sets individually having diameter at most δ, and sum the diameters. The limiting value as $\delta \to 0$, as we do this as efficiently as possible, is the Hausdorff length.

Proof. Let X be a connected set containing the given points x and y. The definition of Hausdorff length does not depend on a coordinate system, so we may assume without loss of generality that $x = 0$ and $y = (a, 0)$.

If D is a bounded subset of the x-axis, we have $d(D) = \sup D - \inf D$, its width. The projection $\pi(X)$ of X to the x-axis is a connected set containing x and y. As it is connected, $[0, a] \subset \pi(X)$. For any $D \subset \mathbb{R}^2$, let

$$\pi_1(D) = \pi(D) \cap [0, a].$$

In particular $\pi_1(X) = [0, a]$. If we cover X with sets D_j, then we also have a cover of $\pi_1(X)$ with sets $\pi_1(D_j)$, satisfying $d(\pi_1(D_j)) \le d(D_j)$. It follows from this inequality that we can cover $[0, a]$ at least as efficiently as X, and

$$\mathcal{H}^1([0, a]) \le \mathcal{H}^1(X).$$

This completes the first statement of the theorem.

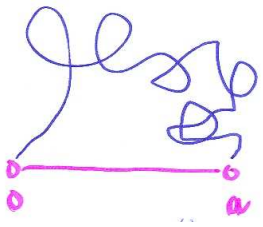


Figure 41.3. *The projection to the x-axis.*

Part 2: We claim $\mathcal{H}^1([0, a]) = a$. The proof is really just a matter of unwinding the definition of Hausdorff length in the special context of line segments. Let $h = \mathcal{H}^1([0, a])$. The proof is trivial in the direction $a \ge h$: it is enough to cover $[0, a]$ with increasingly small intervals $D_j = [aj/N, a(j + 1)/N]$ (with $1/N < \delta$) of total length a.

To show that $a \leq h$, it is enough to show that $a \leq h + \epsilon$ for all $\epsilon > 0$. Pick δ, and a cover D_j $(j = 1, 2, \ldots)$ of $[0, a]$ by sets (in $\mathbb{R}$) such that in reference to the definition of Hausdorff length, we have

$$\sum d(D_j) \leq h + \epsilon/2.$$

Replace D_j by an open interval D'_j, whose width is not greater than that of D_j by more than $\epsilon 2^{-j}$. Then

$$\sum d(D'_j) \leq h + \epsilon.$$

The sets D'_j form an open cover of the compact set $[0, a]$. Pass to a finite subcover, reindexing so that the elements of the subcover have indices $j = 1, \ldots, N$. Take Boolean combinations

$$D''_j = D'_j \setminus (D'_1 \cup \cdots \cup D'_{j-1})$$

to make the sets disjoint. Suppose that D''_j is a finite disjoint union of intervals D''_{js} (open,closed, or half-open). Then

$$a = \sum_{js} d(D''_{js}) \leq \sum_j d(D''_j) \leq \sum d(D'_j) \leq h + \epsilon,$$

as desired. □

Chapter 42

Isoperimetric Inequality for Polygons

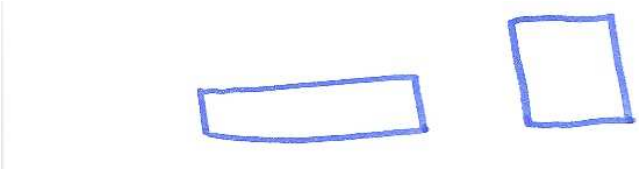


Figure 42.1. *A rectangle has longer perimeter than a square enclosing the same area.*

Theorem 42.1. *Let $L_n(a)$ be the perimeter of a regular polygon of area a. Let P be any polygon with n sides, and area at least a. Then the perimeter L of P is at least $L_n(a)$. Moreover, equality holds only for regular n-gons of area a.*

Proof. The set of polygons n of perimeter at most $L_n(a)$ and area at least a, with one vertex fixed at the origin, is a compact space. (For example, the coordinates (x_i, y_i) of other $(n-1)$ vertices give an embedding of this set of polygons as a subset of the closed product of $(n-1)$ balls of radius $L_N(a)$ centered at the origin. The condition that the area is at least a imposes a closed condition on the set.) The length function L is a continuous function on this compact set. Thus, there exists a polygon P' that achieves the minimum length L' among all polygons of area at least a. We study the properties of an extremal polygon P'.

Its area must be exactly a. Otherwise, we may contract the polygon by a homothety $(x_i, y_i) \mapsto \lambda(x_i, y_i)$ to decrease the area to a, which will decrease the perimeter further.

The polygon P' must be convex. Otherwise the convex hull of P' has smaller perimeter (and greater area). (The convex hull may have fewer sides than the original polygon. If so, insert additional vertices along an edge of the convex hull

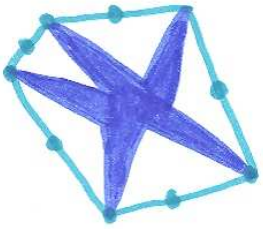


Figure 42.2. *The convex hull has smaller perimeter and larger area.*

to bring the number of vertices back up to n.)

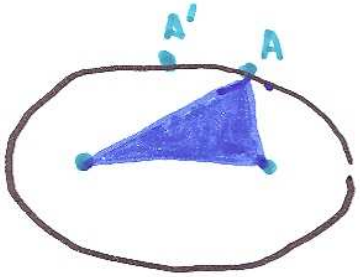


Figure 42.3. *On an ellipse the sum of the distances to the foci is constant. The top of the ellipse gives the triangle of greatest area.*

The polygon P' must have sides of equal length. In fact, consider any two consecutive edges AB and BC. Let p be the sum of the lengths of AB and BC. We may move B, keeping the length sum p constant. As we do so, B sweeps through the points of an ellipse. The point on the ellipse giving greatest area to the triangle ABC occurs when AB and BC have equal length.

Now we finish the proof in two cases depending on whether n is even or odd.

Case 1: $n = 2k$. Let A and B be opposite vertices the polygon P'. Since all sides have equal length, the perimeter is cut into two equal lengths by the opposite vertices A and B.

The diameter from A to B cuts the polygon P' into two equal areas $a/2$ as well. Otherwise, we could take the side of greater area, and reflect it across the diameter to form a polygon of the same length perimeter and area greater than a. This is contrary to the established fact that the extremal polygons have area exactly a.

Let C be any vertex of the polygon other than A, B. The triangle ABC must have a right angle at C. Otherwise, considering the half of the polygon along the

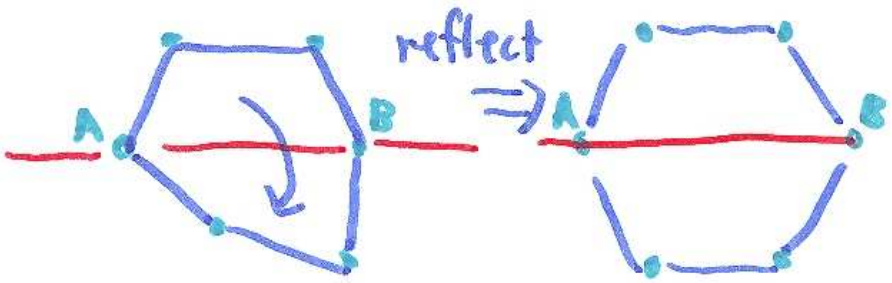


Figure 42.4. *Reflecting the polygon through the perimeter bisector can increase the area, if the areas on the two sides are unequal.*

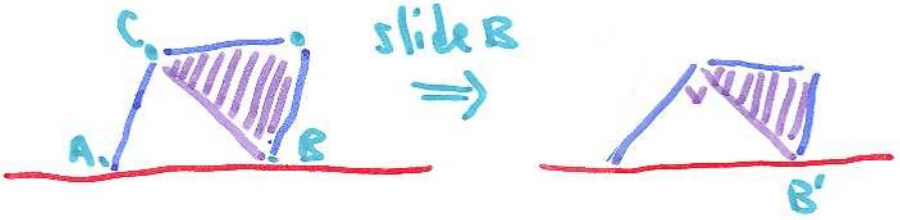


Figure 42.5. *The triangle ABC can be increased in area by making the angle at C a right angle.*

diagonal AB, we may expand or contracting the diagonal and hinging at C we can increase the area of the given half of polygon, keeping the perimeter fixed. Reflecting the half along the diagonal AB, we obtain a polygon of the same perimeter and area greater than a. This is contrary to the established fact that extremal polygons have area exactly a.

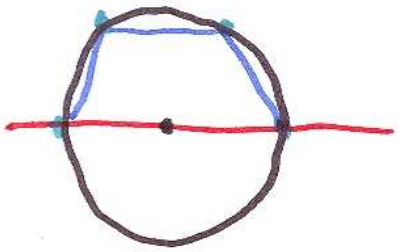


Figure 42.6. *If the triangles ABC are right triangles, then the polygon is inscribed in a circle.*

Thales' theorem asserts that triangle ABC with right angle at C is inscribed

in a circle with diameter AB. Since the vertex C was arbitrary, the polygon P' is inscribed the circle with diameter AB. Thus, it is regular.

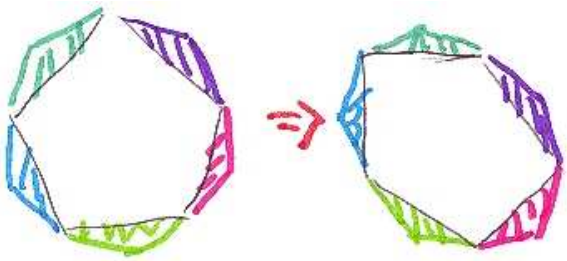


Figure 42.7. *If there is irregular odd-sided polygon with greater area than the regular one, it can be used to construct an irregular even sided-one as well enclosing greater area than the regular one.*

Case 2: $n = 2k + 1$. Assume we have a counterexample: P' is an n-gon with equal sides that is not regular. Connect every second vertex of a regular $2n$-gon to form a regular n-gon. Suppose P' is a polygon with edge lengths the same at this regular n-gon. We can change every second angle of the $2n$-gon so P' is polygon formed by taking every second vertex. If P' is a counterexample in the odd case, the resulting polygon would be a counterexample to the even case. However, the even case has already been fully established, so no such counterexample P' can exist. □

[SIDE NOTE] Thales' theorem can be established as follows. Take a triangle with right angle at Q and other two endpoints P and $-P$. The right angle condition can be expressed as a dot product:

$$0 = (Q - P) \cdot (Q + P) = \|Q\|^2 - \|P\|^2.$$

Thus, $\|Q\| = \|P\|$ and Q lies on the circle at the origin of radius $\|P\|$. [/SIDE NOTE]

Chapter 43

Isoperimetric Inequality for Jordan Curves

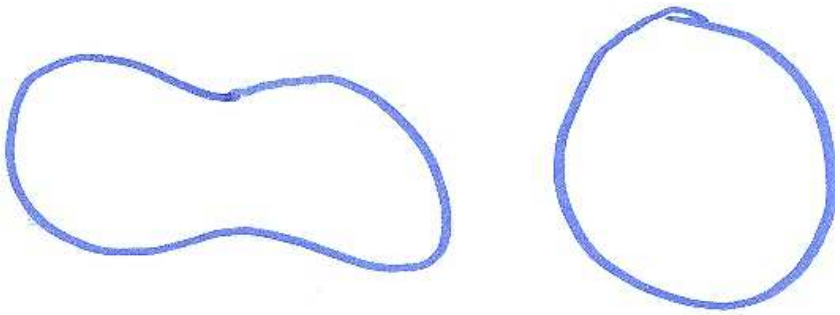


Figure 43.1. *Circles have the shortest perimeter of all Jordan curves enclosing a given area.*

Theorem 43.1. *Let J be a Jordan curve in the plane. Let L be its Hausdorff length and let a be the area of the bounded region of its complement. Then*

$$L^2 \geq 4\pi a.$$

Equality holds exactly when J is a circle.

The Hausdorff length of a Jordan curve may be infinite. In this case, we understand the statement as the assertion that infinity dominates any finite constant. We are really only concerned with the case that J has finite Hausdorff length.

Proof. We begin with a proof of the inequality. It is enough to show that

$$L^2 \geq 4\pi(a - \epsilon),$$

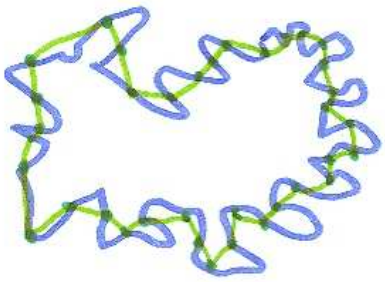


Figure 43.2. *Jordan curves can be approximated by polygons that enclose roughly the same area and have shorter perimeter.*

for all $\epsilon > 0$. By DETOURS AND TRAPS, there is a polygon P with the property that its perimeter is no longer than that of J and its area is at least $(a - \epsilon)$. If it has n sides then by ISOPERIMETRIC POLYGON, its perimeter L' is at least that of a regular polygon of area $a' = (a - \epsilon)$. We have

$$\begin{aligned} \theta &= \pi/n \\ a' &= n \sin\theta \cos\theta, \\ L' &= 2n \sin\theta. \end{aligned}$$

The inequality $\tan\theta \geq \theta$ (which was proved in XXX) gives us:

$$\begin{aligned} L^2 \geq L'^2 &= 4n^2 \tan\theta \sin\theta \cos\theta \\ &\geq 4n^2\theta \sin\theta \cos\theta \\ &= 4\pi a' = 4\pi(a - \epsilon). \end{aligned}$$

This completes the proof of the inequality.

Now we examine the case of equality in detail. Let J be a Jordan curve of finite Hausdorff length whose length and area satisfy the extremal condition

$$L^2 = 4\pi a.$$

First of all the region D bounded by J must be convex. In fact, if the segment between $x, y \in D$ is not contained in D, then there are $x', y' \in J \cap [x, y]$ such that the open segment (x', y') lies exterior to J. Then the modified curve J' that replaces the arc of J from x' to y' with the line segment $[x', y']$ has no greater perimeter LINES MINIMIZE and greater area. Thus, J if not convex is not extremal.

Let A and B be points on J that bisect the perimeter into two equal lengths. Repeating the reflection argument applied to polygons in ISOPERIMETRIC POLYGON], we see that the diameter from A to B must cut the area in half.

Repeating the argument from the last chapter based on Thales' theorem, we find that every point C on J (distinct from A and B) forms a triangle ABC with right angle at C, hence lies on the circle with diameter AB. (To justify the hinging

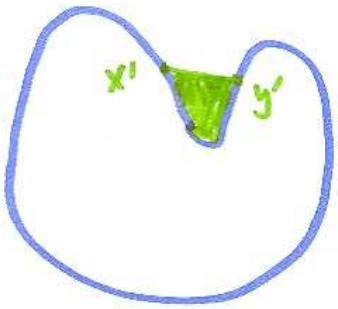


Figure 43.3. *A non-convex region can be improved by adding a segment to shorten the perimeter and increase the area.*

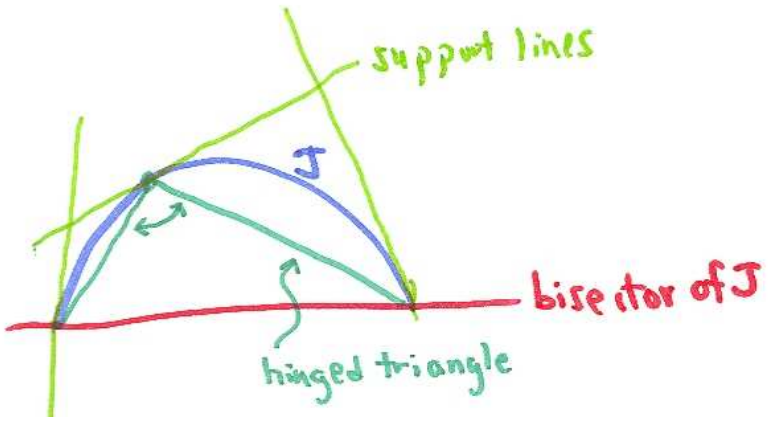


Figure 43.4. *A triangular region can be increased in area without changing the perimeter if it is not a right triangle.*

argument in the case of an arbitrary convex J, we need to check that the hinging argument does not lead to self-intersections of the hinged curve J'. We may use HYPERPLANE SUPPORT theorem to show that J and its interior are contained in a triangle formed by supporting lines to J at A, B, and C. Thus, we may flex the curve J at C a small amount without creating self-intersections on the resulting curve.) All points of J thus lie on the circle with diameter AB. Thus, J itself is a circle. □

Chapter 44

The Honeycomb Theorem

Theorem 44.1. *Partition a flat torus of area F into convex polygons, each having unit area. The total perimeter of the network of polygons is at least $Fh/2$, where h is the area of a regular hexagon of unit area. We have equality exactly when each of the polygons is a regular hexagon of unit area.*

Proof. Let $Q_1, \ldots, Q_F$ be the convex polygons. Assume that Q_i has n_i edges. By the isoperimetric inequality for polygons, the perimeter p_i of Q_i is at least $\ell(n_i)$, where

$$\ell(n) = 2\sqrt{n \tan(\pi/n)}.$$

Note that $h = \ell(6)$. (We note that $\ell(n)$ was called $L_n(1)$ in earlier chapters.)

We claim that

$$\ell(n) \geq \ell(6) + 0.08(6 - n), \text{ for } n = 3, 4, 5, \ldots \tag{44.1}$$

and that equality holds exactly when $n = 6$. This inequality holds for $n = 3, 4, 5, 6, 7, 8$ by computing the constants on both sides. When $n \geq 9$, we use the ISOPERIMETRIC INEQUALITY as follows. The perimeter $\ell(n)$ of a regular n-gon of unit area is at least $2\sqrt{\pi}$, which is the perimeter $(2\pi r)$ of a disk of unit area (and radius $r = 1/\sqrt{pi}$). Thus

$$\ell(n) > 2\sqrt{\pi} \geq \ell(6) + 0.08(6 - 9) \geq \ell(6) + 0.08(6 - n).$$

By EULER FOR A FLAT TORUS, we have $6F \geq 2E$ and $\sum_{i=1}^{F} n_i = 2E$.

Twice the perimeter of the network is

$$\begin{aligned} \sum_{i=1}^{F} p_i &\geq \sum_{i=1}^{F} \ell(n_i) \\ &\geq \sum_{i=1}^{F} [\ell(6) + 0.08(6 - n_i)] \\ &= F\ell(6) + 0.08(6F - 2E) \\ &\geq F\ell(6). \end{aligned}$$

This is the desired result. To analyze the case of equality, we note that $p_i = \ell(n_i)$ requires P_i to be a regular polygon. Also, the Inequality 44.1 holds exactly when $n = 6$. □

Chapter 45

Kissing in Eight Dimensions

Theorem 45.1. *There exists an arrangement of* 240 *unit balls in eight dimensions that do not overlap one another, and such that every one is tangent to a unit ball at the origin.*

Proof. Let $e_1 = (1,0,0,0,0,0,0,0), e_2 = (0,1,0,\ldots,0), \ldots, e_8 = (0,0,0,0,0,0,0,1)$.
Let $C = C_1 \cup C_2$, where

$$\begin{aligned} C_1 &= \{\pm e_i \pm e_j \mid i \neq j\} \\ C_2 &= \{v = \pm\tfrac{e_1}{2} \pm \tfrac{e_2}{2} \cdots \pm \tfrac{e_8}{2} \mid v \cdot (1,1,1,1,1,1,1,1) \text{ is even }\}. \end{aligned}$$

The cardinality of C_1 is $4\binom{8}{2} = 112$. (The factor of four comes from the four possibilities for the signs, and the factor $\binom{8}{2}$ comes from choosing the two indices i and j from a set of 8.) In the set C_2 if exactly one sign is changed, the parity of the given dot product switches between even and odd. Thus, we can set 7 of the signs, and let the final sign be determined by a parity condition. Thus, the cardinality of C_2 is $2^7 = 128$. The cardinality of C is $112 + 128 = 240$.

We claim that if $v \in C$, then $v \cdot v = 2$. In fact, if $v \in C_1$, the dot product is $e_i \cdot e_i + e_j \cdot e_j = 2$. If $v \in C_2$, the dot product is $e_1 \cdot e_1/4 + \cdots + e_8 \cdot e_8/4 = 2$.

We claim that if $v \neq w \in C$, then $v \cdot w \leq 1$. If w and v are both in C_1, the largest dot product from unequal vectors comes when two components are equal: say $v = e_i \pm e_j$ and $w = e_i \pm e_k$, so $v \cdot w = e_i \cdot e_i = 1$. If w and v are both in C_2, the largest dot product from unequal vectors comes when all but two components are equal (by the parity condition, we cannot change the sign of a single component). This gives dot product

$$1/4 + 1/4 + \cdots + 1/4 - 1/4 - 1/4 = 6/4 - 2/4 = 1.$$

Finally, if $v \in C_1$ and $w \in C_2$, the largest dot product is $v \cdot w = (e_i \cdot e_i)/2 + (e_j \cdot e_j)/2 = 1$.

It follows that if we the place centers of the 240 balls at $\sqrt{2}v$, with $v \in C$, then the centers all have distance 2 from the origin, making the balls all tangent to the one at the center. Also, the distances between centers will be

$$\|\sqrt{2}v - \sqrt{2}w\| = \sqrt{2}\|v - w\| = \sqrt{2(v \cdot v - 2v \cdot w + w \cdot w)} \geq \sqrt{2(2 - 2 + 2)} = 2.$$

So the unit balls with these centers do not overlap. □

Chapter 46

Generation of Surfaces

Theorem 46.1. *Every compact surface can be generated from the sphere by adding handles and cross-caps.*

Note: This chapter is not complete. The proof will follow C. Thomassen's slick proof. It is based on the triangulation theorem for surfaces.

There will also be an side note sketching how the full classification theorem for surfaces follows without much additional work.

Chapter 47

Bellows Conjecture

Theorem 47.1. *If a polyhedral body in three dimensional space is deformed in such a way that every face remains rigid, then the volume of the body is preserved under that deformation.*

Note: This chapter is not complete. The proof will follow the proof by Sabatier and Connelly. The chapter appears here because the induction hypothesis that gets used in this proof is precisely the same induction hypothesis that appears in Thomassen's proof of surface generation.

Making this an easy piece will be a bit tricky because of their use of valuations! There will also be an side note giving an example of such a deformable body.

Chapter 48

Penrose Tilings

If we start tiling the plane with regular pentagons, we get scraps of empty space between pentagons.

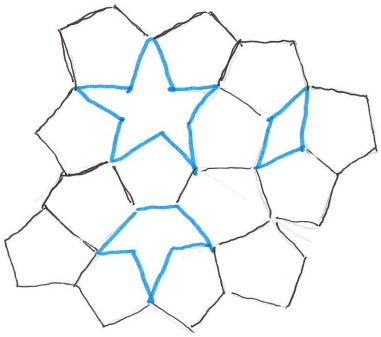


Figure 48.1. *Regular pentagons leave scraps of the plane uncovered.*

Theorem 48.1. *There exists a tiling of the plane with the following six Penrose tiles (in Figure 48.2) in which the color scheme along each edge is limited to those indicated in Figure 48.3.*

(In other words, we can place pentagons in such a way that we don't get stuck after placing finitely many. The only scrap tiles we need are the star, cut star, and rhombus.)

As an example of the color scheme, an analysis of the colors of the rhombus leads to the conclusion that the tiles around it are as shown in Figure 48.4. (This analysis assumes that we do not place a tile that will lead immediately to an

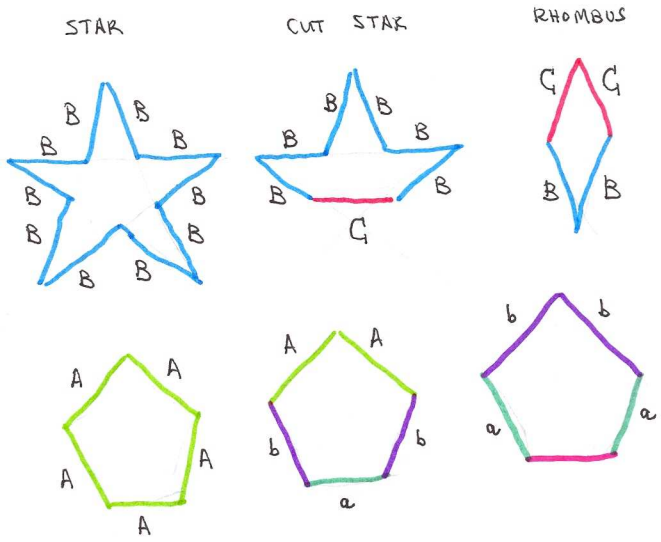


Figure 48.2. *The Penrose tiles.*

Figure 48.3. *The color scheme on edges.*

inextensible arrangement.)

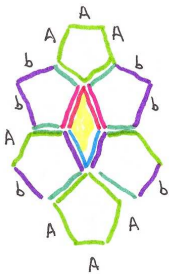


Figure 48.4. *The color scheme forces the surrounding tiles of a rhombus.*

Proof. For any partial tiling of the plane, we can substitute the pieces of Figure 48.5 to get another partial tiling at a smaller scale. This process is called *decomposition*. For example, by decomposing, the partial tiling P_0 of Figure 48.4, we obtain the partial tiling P_1 of Figure 48.6. We can 'grow' a Penrose tiling by repeating the

process and rescaling to standard size. For example, if we repeat twice more and enlarge the tiles back to standard size, we get P_2 and P_3 of Figures 48.7 and 48.8.

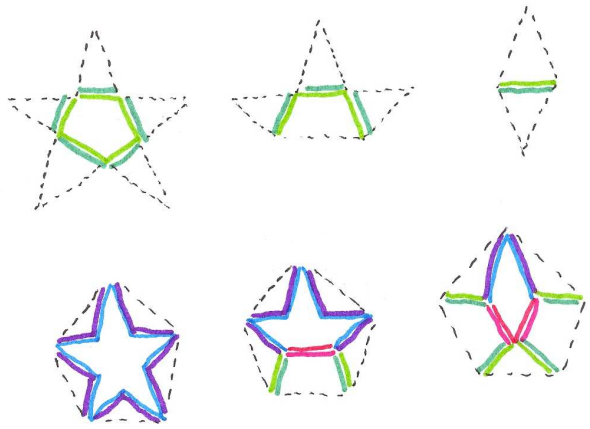


Figure 48.5. *The tiles used for decomposition.*

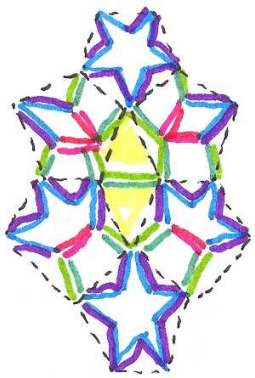


Figure 48.6. *The decomposition P_1 of P_0.*

Note that the original rhombus and partial tiling of P_0 has reappeared at the center of P_3. Using this repeated pattern, we can iterate to fill the plane as follows. Make an increasing sequence of nested rhombi. (The sizes are in geometric progression in the ratio of the size of the rhombus in P_0 to that of P_3.)

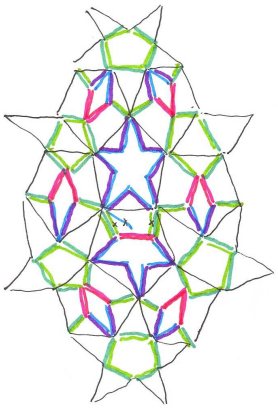


Figure 48.7. *The decomposition P_2 of P_1.*

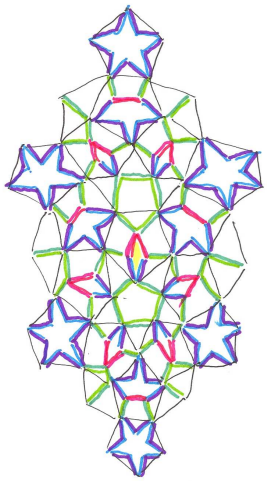


Figure 48.8. *The decomposition P_3 of P_2.*

Starting with the rhombus at level n, surround it by neighboring pentagons to get a partial tiling $P_0(n)$ similar to P_0. Decompose 3 times to get $P_3(n)$. As we have observed, this contains a smaller rhombus: $P_0(n-1) \subset P_3(n)$. As we continue to decompose, they stay in sync:

$$P_i(n-1) \subset P_{i+3}(n), \text{ for all } n, i.$$

Thus,

$$P_0(0) \subset P_3(1) \subset P_6(2) \subset \cdots \subset P_{3n}(n) \subset \cdots$$

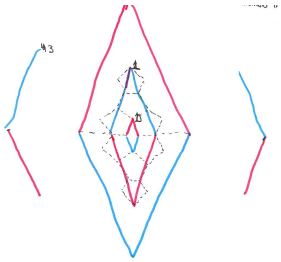


Figure 48.9. *Nested rhombi in geometric series.*

The union

$$\bigcup_i P_{3i}(i)$$

is a tiling of the plane. □

Note: I plan to insert a side note sketching why any such tiling is aperiodic.

Chapter 49

Glossary

Bibliography

[AH83] G. Alefeld and J. Herzeberger, Introduction to Interval Computations, Academic Press, New York, 1983.

[arXiv] http://xxx.lanl.gov.

[Wie05] F. Wiedijk, Formalizing 100 Theorems
http://www.cs.ru.nl/~freek/100/,