SQUARING RECTANGLES: THE FINITE RIEMANN MAPPING THEOREM

J. W. Cannon, W. J. Floyd and W. R. Parry

Abstract. The classical Riemann mapping theorem asserts that any topological quadrilateral in the complex plane can be mapped conformally onto a rectangle. The finite Riemann mapping theorem asserts that any topological quadrilateral tiled by finitely many 2-cells can be mapped with minimal combinatorial distortion onto a rectangle tiled by squares. We prove the finite Riemann mapping theorem, discuss its connections with the classical theories of conformal mapping and electric circuits, and develop algorithms for calculating the finite Riemann mapping.

Table of contents.

1991 *Mathematics Subject Classification.* Primary 20F32, 30C62, 31A15; Secondary 30F10, 30F40, 57N10.

This research was supported in part by The Geometry Center, University of Minnesota, an STC funded by NSF, DOE, and Minnesota Technology, Inc. The first author was supported in part by NSF Research Grant No. DM-8902071.

0. Introduction

We dedicate this paper to Wilhelm Magnus with admiration and appreciation. This paper has mild ties with Magnus through the work of his thesis advisor, Max Dehn, who proved long ago that, for any rectangle which can be tiled by finitely many squares, the ratio of height to width is rational. This paper is directed in the long term to the recognition of discrete groups of isometries in hyperbolic 3-space, and the subject of discrete groups was one of Magnus's favorites.

Riemann, in formulating his famous Riemann mapping theorem, surely relied on the physics of electrical networks and conducting metal plates for motivation. In turn, Riemann's theorem and its many proofs can be used as starting points for a number of beautiful finite approximations which have combinatorial, geometric, and physical interpretations. These **finite** Riemann mapping theorems are the subject of very intense interest. One of them is the subject of Cannon's paper [8]. Another is of central importance to this paper. The paper is partly expository, especially in the beginning. However, in the later sections, we develop material important to our program of recognizing groups of hyperbolic motions combinatorially.

1. Squared rectangles

This section is purely expository. Our aim is to show a correspondence between rectangles tiled by rectangles and connected planar resistive circuits with one battery.

1.1. Scientific American. We begin our story with a puzzle whose solution is recounted in Chapter 17 of <u>The 2nd Scientific American Book of Mathematical Puzzles and Diversions</u> by Martin Gardner [13]: *Can a square be subdivided into smaller squares of which no two are alike?* Gardner's story teller is William T. Tutte, renowned geometer and combinatorialist at the University of Toronto. Tutte, as a student at Trinity College, Cambridge, during the years 1936-38 pursued an answer to this question with three other students, C. A. B. Smith, A. H. Stone, and R. L. Brooks. "Stone was intrigued by a statement in Dudeney's <u>Canterbury Puzzles</u> [12] which seemed to imply that it is impossible to cut up a **square** into unequal smaller squares." Soon the four friends "were spending much time constructing, and arguing about, dissections of rectangles into squares." They used a mixture of geometric and algebraic methods to create a large catalog of squared rectangles. "In the next stage of the research [they] abandoned experiment in favor of theory." The students "tried to represent squared rectangles by diagrams of different kinds." Smith suggested a diagram which showed that the problem could be reinterpreted as a part of the theory of electrical networks.

We shall now explain Smith's diagram. Consider as an example the rectangle of Figure 1.1.1 which is tiled by smaller rectangles. Each maximal horizontal line segment of the diagram is to be represented by a node of a simple resistive circuit. The top and the bottom of the rectangle represent nodes of the circuit which are to be connected to the terminals of a battery. Each small rectangle is to be represented by an edge of the circuit consisting of a single resistor. This resistor joins the nodes (maximal horizontal line segments) in which the top and bottom of the corresponding rectangle lie. We must assign values to the voltage of the battery and the resistance of the resistors. We interpret height as voltage drop and width as current flow. Hence the height of the large rectangle is assigned as the voltage of the battery. We apply Ohm's law which gives resistance as the quotient of voltage by current to determine the resistance of each resistor as the ratio of height to width in the corresponding small rectangle. We observe that a square in the tiling corresponds to a resistance of 1.

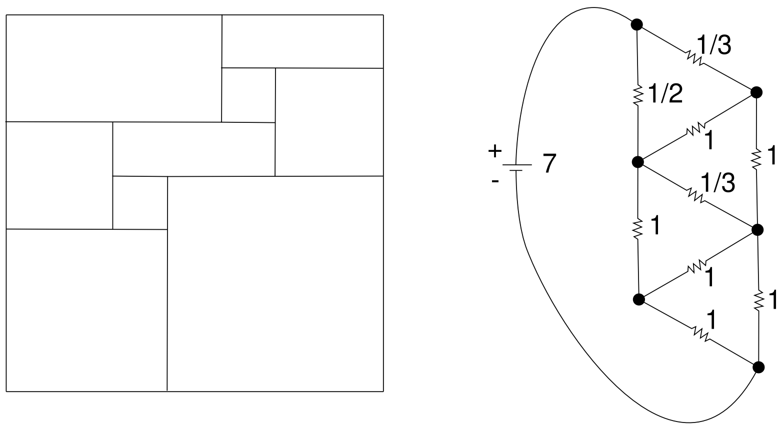


Figure 1.1.1.

One half of Smith's observation is that every rectangle tiled by rectangles gives rise to a circuit in which the sizes of both the large rectangle and its subrectangles are determined by the solution to the corresponding circuit laws. The second half of Smith's observation is that every connected planar resistive circuit with one battery gives rise to a rectangle tiled by rectangles. We shall outline proofs of these observations in subsequent sections.

Our four students used Smith's observation and classical circuit theory to find, after some considerable work, solutions to the problem with which this section began. We shall leave the reader to read their work in the literature. We conclude our discussion of their work by mentioning one related recreational puzzle. A rectangle tiled by rectangles can be cut into its underlying tiles. Many tiled rectangles can be reassembled in only one way into a rectangle. This reassembly constitutes an interesting jigsaw puzzle problem.

We express appreciation to Geoffrey Mess who, from his remarkable memory, was able to refer us to the original paper [4] by these four students in the Duke Mathematical Journal of 1940.

1.2. Resistive circuits and the laws of Ohm and Kirchhoff. Consider a finite directed and connected graph Γ, one edge of which, denoted e, represents a battery of voltage v, and each other edge of which, denoted generically by f, represents a resistor of resistance $r(f)$ and conductance $c(f) = 1/r(f) > 0$. We let $i(f)$ denote the current through the directed edge f and $v(f)$ denote the voltage or electrical potential across f. The flow of electricity in this circuit is governed by laws which we call **circuit laws**, laws attributed specifically to the names of Ohm and Kirchhoff.

Ohm's law. For every edge f in Γ with $f \neq e$,

$$i(f) = c(f)v(f).$$

Note that the current $i(f)$ and voltage $v(f)$ may be positive, negative, or zero, while the conductance $c(f)$, for f not the battery edge, is always positive.

Kirchhoff's voltage law. Let $P = E_1 \cdots E_k$ be a closed edge path in Γ for which the directions of the edges E_j do not necessarily agree with the orientation of P determined by the ordering of the edges $E_1, \ldots, E_k$. Then

$$(1)_P \qquad \pm\, v(E_1) \pm \cdots \pm v(E_k) = 0,$$

where the sign before $v(E_j)$ in the left side of $(1)_P$ is $+$ if the direction of E_j agrees with the orientation of P and $-$ otherwise. If one of these edges E_j is the directed battery edge e of Γ, then $v(E_j) = v$.

Kirchhoff's current law. If x is a vertex or node of the graph Γ distinct from the two endpoints of e, if $E_1, \cdots, E_k$ are the directed edges of Γ ending at x, and if $F_1, \cdots, F_l$ are the directed edges of Γ emanating from x, then

$$i(E_1) + \cdots + i(E_k) = i(F_1) + \cdots + i(F_l).$$

Equivalently, by Ohm's law,

$$(2)_x \qquad c(E_1)v(E_1) + \cdots + c(E_k)v(E_k) = c(F_1)v(F_1) + \cdots + c(F_l)v(F_l).$$

1.3. The solution of the resistive circuit problem.

Theorem 1.3.1. *There are unique values of the variables $v(f)$ and $i(f)$ satisfying the circuit laws.*

Proof. This marvelous and well-known result offers a perfect illustration of a fascinating fact of mathematical life described by Henri Poincaré in his wonderful book, The Value of Science [17], pp. 15-17.

> "It is impossible to study the works of the great mathematicians ... without noticing and distinguishing two ... entirely different kinds of minds. The one sort are above all preoccupied with logic .... The other sort are guided by intuition.
>
> "Look at [the intuition of] Professor Klein: ... He replaces his Riemann surface by a metallic surface whose electrical conductivity varies according to certain laws. He connects two of its points with the two poles of a battery. The current, says he, must pass, and the distribution of this current on the surface will define a function whose singularities will be precisely those called for ...."

Klein's argument applies almost verbatim to our situation and supplies an intuitive proof of Theorem 1.3.1.

But the analytical proof is almost as succinct and fascinating:

By means of Ohm's law we may remove the variables $i(f)$ from consideration and consider only the two sets of Kirchhoff's laws, linear equations in the variables $v(f)$. There are infinitely many closed loops P in any nontrivial circuit so that the equations $(1)_P$ are potentially infinite in number, but there are finite subsystems equivalent to the whole which we might obtain as follows. Pick a maximal tree T in Γ which contains as one of its edges the battery edge e. For each edge f of $\Gamma \setminus T$, consider the unique closed edge path $P = P(f, T)$ in $T \cup \{f\}$ which has f as one of its directed edges and is oriented in the direction of f. Retain the corresponding equation $(1)_P$.

Theorem 1.3.2. *The system (1)$\cup$(2), with (1) reduced to a basis as indicated in the previous paragraph, is a nonsingular linear system.*

Proof of Theorem 1.3.2. The first step is to note that the reduced system has the same number of equations as unknowns. We leave the proof to the reader: use the Euler characteristic of the connected graph Γ and remember that there is one voltage variable for each nonbattery edge, one vertex equation for each node not attached to the battery, and one loop equation for each edge omitted by a maximal tree.

The second and final step is to prove that the corresponding homogeneous linear system obtained by setting the battery voltage to 0 has a unique solution. It follows therefrom that the matrix of the system is nonsingular. Suppose given a solution to the homogeneous system. The Kirchhoff loop equations show that we may assign a well-defined potential to each node of the circuit as follows. Assign the two battery nodes the potential 0. For any other node x, choose an edge path Q from a battery node to x. Sum the values $\pm v(f)$ over the directed edges f of Q using the sign convention of $(1)_P$. Assign the result to x as its potential. The Kirchhoff vertex equations then show that this potential function cannot obtain a maximum or minimum other than 0. It follows that $v(f) = 0$ for every edge f of the circuit.
$\diamondsuit$ Theorems 1.3.1 and 1.3.2 $\diamondsuit$

We end our discussion of existence and uniqueness for solutions of the circuit equations by noting a result that is not as well-known among geometric topologists as it deserves to be. We have proved that the reduced linear system described above is nonsingular. Therefore the matrix has nonzero determinant. What is that determinant?

Theorem 1.3.3. *The determinant of the reduced linear system of circuit equations is nonzero. It has one nonzero term for each maximal tree in the circuit which contains the battery edge. Each nonzero term has the same sign. The absolute value of the term corresponding to a given maximal tree is the product of the conductances of the nonbattery edges of the tree. In particular, if each of the resistances is 1, then the absolute value of the determinant is the number of maximal trees in the circuit containing the battery edge.*

REMARK. The proof of Theorem 1.3.3 is a delightful exercise of which we would not deprive the reader.

1.4. Planar circuits and rectangles tiled by rectangles. We assume the same setting as that of the previous sections except that we assume in addition that the circuit has a planar diagram.

Theorem 1.4.1. *If the circuit Γ is planar, and if the voltages $v(f)$ and currents $i(f)$ are given by the unique solution to the circuit laws, then rectangular tiles of height $|v(f)|$ and width $|i(f)|$ can be assembled so as to tile a rectangle of height $|v|$, where v is the battery voltage, and width $|i|$, where i is the total current exiting or entering a battery node. The ratio $|v/i|$ may be interpreted as the total resistance of the circuit between the battery nodes. The Smith diagram of this tiled rectangle corresponds to the original circuit modified by removing those edges through which no current flows.*

Proof. We must first establish a number of simplifying assumptions, notations, and conventions.

We simplify matters by considering only the case where the voltage v and the conductances $c(f)$ are rational. The general case follows from a simple limiting argument. Alternatively, where we begin later on in the rational case to set out a finite grid, we could in the general case use transverse foliations in place of that grid.

In the rational case, Cramer's rule shows that all of the voltages $v(f)$ and currents $i(f)$ are rational as well. Multiplying by the least common multiple of the denominators, we find that we may assume all of the constants and variables involved are in fact integers.

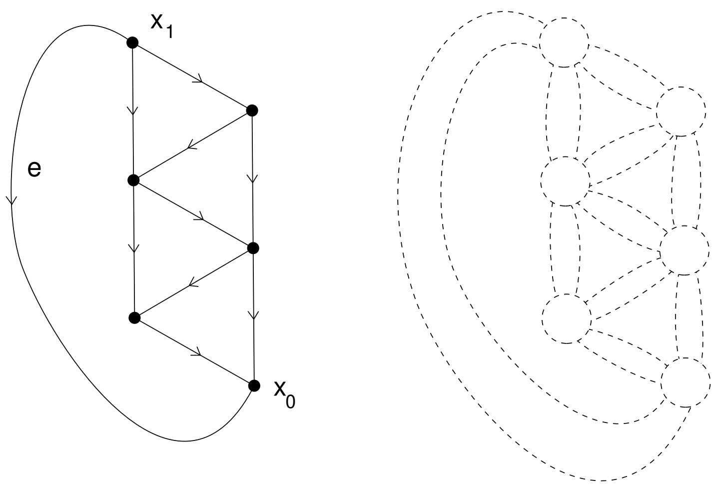


Figure 1.4.2.

We take a regular neighborhood N of the graph Γ in the plane and decompose N into small disks or 0-handles $D(x)$ surrounding the nodes x of the circuit and small strips or 1-handles $H(f)$ covering the uncovered portion of the edges f and joining the disks corresponding to the endpoints of f. See Figure 1.4.2.

We assign to one endpoint x_0 of the battery edge e the potential 0 and to the other endpoint x_1 the potential v. We may assume that $v < 0$ so that current flows from the vertex x_1 through the circuit to the vertex x_0. We proceed then to assign to each node x a potential $p(x)$ as in the proof of the existence and uniqueness theorem above. Actually it is customary to replace this potential function $p(x)$ by its negative $q(x) = -p(x)$ so that current flows along all nonbattery edges from vertices of higher potential to those of lower potential. We will follow this custom. The associated formula is the following. We assume that every nonbattery edge f is directed from its initial endpoint x of higher potential to its terminal endpoint y of lower potential. Then

$$v(f) = q(x) - q(y) = p(y) - p(x)$$

and

$$i(f) = c(f)v(f).$$

We obtain particular unity in our picture if we think of a current i flowing through the battery edge **from vertex x_0 to vertex x_1 from lower potential to higher potential.** Only along the battery edge does current flow from lower to higher potential. **The effect of this convention is that the Kirchhoff current node law is then satisfied at every node of the network and not just at the nonbattery nodes.**

With all of these preliminary matters out of the way, we are ready to prove the theorem. The idea is to impose a finite "rectangular" grid on the neighborhood N in the following way.

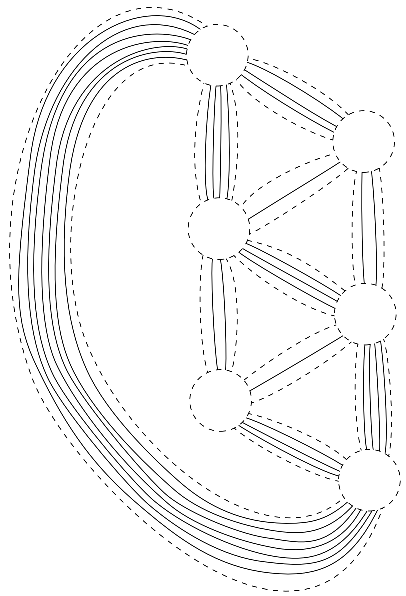


Figure 1.4.3.

Suppose that f is an edge and $H(f)$ the corresponding 1-handle. Construct $i(f)$ parallel directed segments through the 1-handle $H(f)$ joining one of the end 0-handles to the other. See Figure 1.4.3. The direction should coincide with the direction of the current. Call these segments **current curves**. Kirchhoff's current node law implies that, in each 0-handle, the endpoints of the current curves can be matched in pairs, incoming curves to outgoing curves. A simple induction shows that this matching can be realized in such a way that, if matched curves are joined across the 0-handle by directed arcs, the result is a family of disjoint oriented simple closed **current curves** in the plane, each curve actually lying in the neighborhood N. See Figure 1.4.4.

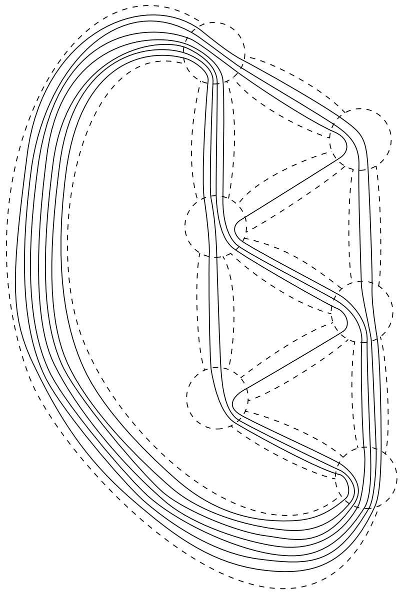


Figure 1.4.4.

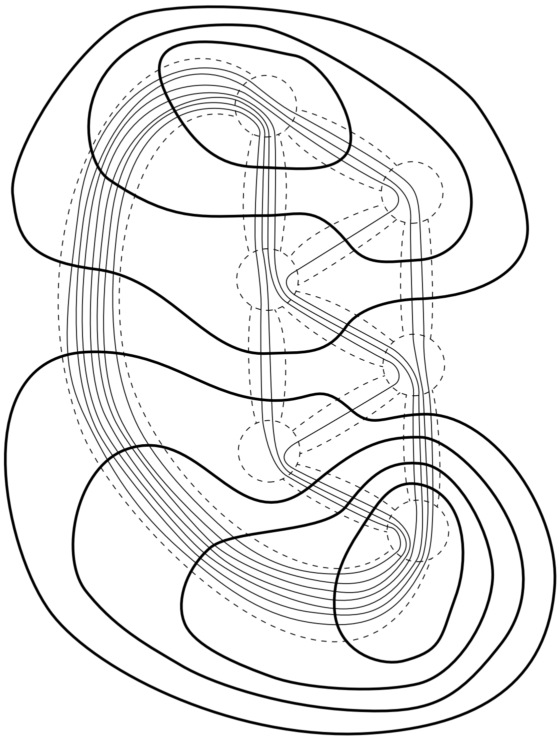


Figure 1.4.5.

We construct a second family of curves called **voltage** or **equipotential** curves as follows. Let f be as above. Assume f oriented from higher potential to lower potential. If f happens to be the battery edge, and in no other case, this orientation will oppose the current flow along the edge f. Construct $v(f)$ parallel segments across the 1-handle $H(f)$ "perpendicular" to the directed edge f, directed in such a way that any ordered pair of (intersecting) segments, (segment, f) forms a local right-handed coordinate system in the plane. Kirchhoff's voltage loop law implies that around the boundary curve of any complementary domain of the neighborhood N the endpoints of the voltage curves can be matched in pairs, incoming curves to outgoing curves. As before, this matching can be realized in such a way that, if matched curves are joined across the complementary domains by disjoint directed arcs, the result is a family of disjoint oriented simple closed **voltage curves** in the plane, each curve crossing N only in the 1-handles of N. See Figure 1.4.5.

It remains to analyze our two families of curves and their intersections. The key observations are these: (1) if we calculate the intersection number $sc(J, K)$ of an oriented voltage curve J with an oriented current curve K, then each intersection outside the battery handle contributes 1 to the intersection number and each intersection inside the battery handle contributes -1; (2) $sc(J, K) = 0$.

CLAIM 1. Each voltage curve and each current curve crosses the battery handle exactly once. For if one missed the battery handle, then all of its intersections with curves of the opposite kind would have the same orientation and the intersection numbers could not be 0. Likewise, it cannot cross the battery handle more than once without having a nonzero intersection number with an arc different from ± 1.

CLAIM 2. Each voltage curve hits each current curve exactly once outside of the battery handle. Indeed, all of the intersections outside the boundary handle have the same sign and, taken together, must exactly cancel algebraically the one intersection inside the battery handle.

The truth of the theorem now becomes apparent. The voltage curves are naturally ordered by the way they cross the voltage handle. The current curves are naturally ordered by the way they cross the voltage handle. Therefore the two families define rectilinear, orthogonal, integer coordinates which naturally turn the entire circuit into a geometric rectangle while turning each 1-handle into a subrectangle of size $i(f) \times v(f)$. The curves viewed in their entirety form a grid like graph paper which gives global integral coordinates.
$\diamondsuit$ Theorem 1.4.1 $\diamondsuit$

1.5. Max Dehn's contribution.

Theorem 1.5.1 (Dehn). *Let R denote a rectangle that can be tiled by squares. Then, if H is the height of the rectangle and W its width, the ratio H/W is rational.*

Proof. Scale R so that $H = 1$. Let Γ denote the Smith diagram associated with a tiling of R by squares. Then the voltage of the battery and all of the conductances are 1. Consequently all of the constants in the reduced circuit equations are integers. Since the voltages of the solution are given by Cramer's rule, they are all rational. Since the currents are expressed in terms of the voltages, the conductances, and Ohm's law, they too are rational. The width W of the rectangle is the sum of the currents entering or exiting a battery terminal; hence, it too is rational.
$\diamondsuit$ Dehn's Theorem $\diamondsuit$

2. Optimal weight functions

We assume now that we are given a combinatorial tiling of a topological quadrilateral by topological disks. We consider the problem of introducing a circuit whose solution would turn this tiling into a tiling of a rectangle by squares. We would consider the solution to such a circuit as giving a finite Riemann mapping defined on the tiled topological quadrilateral. There are problems in this task. First, a solution is incompatible with the given topology since, in a tiling of a rectangle by rectangles, at most four tiles can come together at a point while, in a combinatorial tiling, any number of tiles can come together at a point. Second, the only natural graph associated with the combinatorial tiling is the dual graph; and in that graph the tiles correspond to nodes and not to edges. Third, while we could mimic the idea of the Smith diagram so as to put the tiles on the edges of the diagram, in a combinatorial tiling there is no well-defined notion of "maximal horizontal edge" in the diagram so that any number of circuits might conceivably correspond to our combinatorial tiling. In view of these difficulties, it is amazing that this "impossible" problem essentially has a solution and that the solution is unique up to scaling. Historically, the variational formula giving its solution was recognized before it was understood that the formula had the geometric content necessary to solve the problem. Cannon [8] had been exploring various finite approximations to the classical Riemann mapping theorem with the aim of extracting from combinatorial tilings and their subdivisions inherent analytic information. Parry discovered that the formula Cannon chose as a working tool had the geometric properties required by the problem. Subsequently and independently, Schramm [18] and Robertson [unpublished], who were aware of Cannon's formula, discovered the same geometric interpretation.

2.1. The classical Riemann mapping theorem. We use the classical Riemann mapping theorem as a motivation for Cannon's formula. One version of the classical theorem is the following.

Theorem (Riemann). *Suppose given a quadrilateral, that is, a topological disk in the complex plane with four distinguished points on its boundary. Then there is a conformal mapping taking the interior of the quadrilateral onto the interior of a rectangle in such a way that the induced boundary map takes the four distinguished boundary points to the four corners of the rectangle.*

We have always been amazed, in view of the topological difficulties involved, that this theorem is true, let alone that Riemann could have conjectured and proved it. What may have happened is this. Riemann, like Klein in the passage quoted from Poincaré, may have considered the quadrilateral as a metallic conducting plate with battery terminals connected to its "top" and "bottom". "The current must pass," as Klein is supposed to have said. The current flow lines, connecting top to bottom, would have filled the quadrilateral from side to side, one line through each point of the quadrilateral. Equipotential lines, connecting side to side, would likewise have filled the quadrilateral from top to bottom. The pair of families would meet one another orthogonally and give rectilinear flat coordinates for the quadrilateral.

There are three obvious parameters associated with Riemann's theorem, namely the voltage or height H of the image rectangle, the current or width W of the image rectangle, and the total resistance H/W. It is a remarkable fact that this resistance can be realized as a conformal invariant of either the quadrilateral or its conformal equivalent, the image rectangle. In the terminology of complex variables, this total resistance is called the analytic conformal modulus of the quadrilateral. Note that, if $A = HW$ denotes the area of the image rectangle, then

$$H/W = \frac{H^2}{A} = \frac{A}{W^2}.$$

There is a wonderful trick for creating conformal invariants. (See Ahlfors [1].) For a fixed Riemannian surface, one simply assigns a number to each metric conformally equivalent to the given one and then takes either the supremum or the infimum of those numbers over all of the metrics.

The resistance or modulus H/W is precisely such an invariant. (See Lehto and Virtanen [16], Chapter 1.) It may be realized as follows. Begin with the standard Euclidean metric $|dz|$ on the quadrilateral with its accompanying area form $dA = dx \cdot dy$. With each positive metric multiplier $\rho = \rho(z) > 0$ associate the conformally equivalent metric $\rho|dz|$ with its associated area form $\rho^2 dA$. The new metric and area form define a new area $A(\rho)$, a new height $H(\rho)$, and a new width $W(\rho)$ for the quadrilateral which give respectively the area, the minimal distance between the ends, and the minimal distance between the sides of the quadrilateral with respect to the new Riemannian metric $\rho \cdot |dz|$ and the new area form $\rho^2 dx \cdot dy$. Then we have

$$H/W = \sup_\rho \frac{H(\rho)^2}{A(\rho)} = \inf_\rho \frac{A(\rho)}{W(\rho)^2}.$$

Furthermore, both the supremum and infimum are realized by that positive multiplier function ρ which turns the quadrilateral into a rectangle. That is, the optimal function ρ is the absolute value of the derivative of the Riemann mapping. This circumstance plays a central role in the finite theorem where the lack of local coordinates makes the definition of derivative, let alone its use, difficult.

2.2. The finite problem and the existence and uniqueness of its solution. We are dealing with a combinatorial situation in which no Riemannian structure is given. We wish to use a finite version of the Riemann mapping theorem as a step toward imposing a Riemannian structure. We attempt to copy the classical procedure presented above. As setting, we have a quadrilateral combinatorially tiled by topological disks. We use the tiling to define an approximate metric and approximate area for subsets of the quadrilateral. We simply define both the length and area of a subset to be the number of tiles that intersect it. That is, we assume that each tile has length and area equal to 1 so that it behaves as if it were a unit square. It is then analogous to the classical case if we make a "conformal" change of approximate metric by changing the length of the tile to ρ and the area of the tile to ρ^2. The number ρ may be an arbitrary nonnegative function of the tiles not identically equal to 0. The ρ-length and ρ-area of a subset are then simply the sums of the tile lengths and areas for tiles intersecting that subset. We obtain thereby heights, widths, and areas $H(\rho)$, $W(\rho)$, and $A(\rho)$ by the same formulas used in the classical case. Varying ρ over all possibilities, we obtain two approximate conformal moduli,

$$M_{\sup} = \sup_{\rho} \frac{H(\rho)^2}{A(\rho)} \quad \text{and} \quad m_{\inf} = \inf_{\rho} \frac{A(\rho)}{W(\rho)^2}.$$

In contrast to the classical case, it is in general not true that these two moduli are the same. See line 2.4.5.2 (where $m_f = m_{inf}$ and $M_f = M_{sup}$) and Example 6.2.2.

A metric multiplier ρ is said to be an **optimal weight function** if it realizes the supremum in the definition of $M_{\sup}$.

Theorem 2.2.1. *There is a unique optimal weight function ρ such that $A(\rho) = 1$. All other optimal weight functions are scalar multiples of this one.*

REMARK AND DEFINITIONS. In developing the properties of optimal weight functions, we will be well-served by a general setting in linear algebra. We shall prove a number of things in more generality than is required by the study of tiled quadrilaterals. This will lead to some slight changes in notation and terminology.

We therefore fix a positive integer n, which we view as the number of tiles in our tiling. We let P denote a nonempty finite subset of $\mathbf{N}^n \setminus \{0\}$, where $\mathbf{N}$ is the set of nonnegative integers. Elements of P will be called **path vectors**. A **weight vector** is a vector $w = (w_1, \dots, w_n) \in \mathbf{R}^n \setminus \{0\}$ with $w_i \geq 0$ for $i = 1, \dots, n$. Let $\langle , \rangle$ be the standard inner product on $\mathbf{R}^n$. The **length** of a path vector p relative to the weight vector w is by definition $\langle p, w \rangle$. Given a weight vector w, define **height** H_w relative to w to be the minimum length of a path vector in P relative to w. A path vector p such that $\langle p, w \rangle = H_w$, namely, a path vector whose length relative to w is minimal will be called a **w-minimal path vector**. Define **area** A_w by $A_w = \langle w, w \rangle$. Define **modulus** M by

$$M = \sup_{w} \frac{H_w^2}{A_w},$$

where the supremum varies over all weight vectors. Define an **optimal weight vector** to be a weight vector w with $M = H_w^2 / A_w$.

We are now ready for the proof of the existence and uniqueness of optimal weight functions. It is clear that to prove Theorem 2.2.1 it suffices to prove Theorem 2.2.2.

Theorem 2.2.2. *There is a unique optimal weight vector w with $A_w = 1$. All other optimal weight vectors are scalar multiples of this one.*

Proof. Scaling w does not change the value of H_w^2/A_w, and so in defining M the weight vector w may be restricted to the $(n-1)$-sphere S^{n-1}. Moreover, the function which maps a weight vector w in S^{n-1} to H_w^2 is the minimum of a finite number of continuous functions, and so it is also continuous. Thus it is easy to see that M is a positive real number attained for some weight vectors. This argument proves existence of optimal weight vectors.

The proof of uniqueness is a convexity argument. Let $w_0 \neq w_1$ be weight vectors in S^{n-1} with $H_{w_0} \leq H_{w_1}$. Let $w = (1-t)w_0 + tw_1$ for some real number t with $0 < t < 1$. If p is a path vector, then

$$\langle p, w\rangle = (1-t)\langle p, w_0\rangle + t\langle p, w_1\rangle \geq (1-t)H_{w_0} + tH_{w_1} \geq H_{w_0}.$$

Since $0 < \|w\| < 1$, $(1/\|w\|)w \in S^{n-1}$ and $\langle p, (1/\|w\|)w\rangle > H_{w_0}$. This argument easily establishes uniqueness.
$\diamondsuit$ Theorems 2.2.1 and 2.2.2 $\diamondsuit$

Denote by w_M the unique optimal weight vector in S^{n-1}.

2.3. The geometry of an optimal weight function: general results.

We next present a number of general results in the above vector space setting. Each result will appear in a numbered line followed by an explanation or proof.

2.3.1. Optimal weight vectors are in some sense generalized spherical circumcenters.

Following is an explanation of this statement. Let $x_1, \ldots, x_k$ be points in S^{n-1}. Set

$$r = \inf_{y \in S^{n-1}} \max_i \|x_i - y\|.$$

Then there exists at least one point x in S^{n-1} such that $\|x_i - x\| \leq r$ for $i = 1, \ldots, k$. Such a point x is called a spherical circumcenter of the set $\{x_1, \ldots, x_k\}$. Since

$$\|x_i - y\|^2 = \langle x_i - y, x_i - y\rangle = \langle x_i, x_i\rangle - 2\langle x_i, y\rangle + \langle y, y\rangle = 2 - 2\langle x_i, y\rangle,$$

maximizing $\|x_i - y\|$ is the same as minimizing $\langle x_i, y\rangle$. Thus

$$1 - \frac{1}{2}r^2 = \sup_{y \in S^{n-1}} \min_i \langle x_i, y\rangle,$$

and a spherical circumcenter of the set $\{x_1, \ldots, x_k\}$ is a point x in S^{n-1} such that $\langle x_i, x\rangle \geq 1 - \frac{1}{2}r^2$ for $i = 1, \ldots, k$. Now suppose that all of the path vectors $p_1, \ldots, p_k$ in P have the same length. In this special case the optimal weight vector in S^{n-1} is the spherical circumcenter of the set $\{x_1, \ldots, x_k\}$, where $x_i = \frac{1}{\|p_i\|}p_i$ for $i = 1, \ldots, k$. Thus in general it might be said that optimal weight vectors are generalized spherical circumcenters.

2.3.2. The optimal weight vector w_M lies in the cone in $\mathbf{R}^n$ spanned by the w_M-minimal path vectors.

Proof. Suppose line 2.3.2 is false. Then there exists a hyperplane V in $\mathbf{R}^n$ which separates w_M from the cone spanned by the w_M-minimal path vectors. Let v be the projection of w_M onto V, and let $u = w_M - v$, so that $u + v = w_M$ and $\langle u, v\rangle = 0$. Because w_M and the w_M-minimal path vectors lie on opposite sides of V, $\langle u, p\rangle < 0$ for every w_M-minimal path vector p. Thus $v - u \in S^{n-1}$ and $\langle v - u, p\rangle > \sqrt{M}$ for every w_M-minimal path vector p. By choosing u to be sufficiently small, equivalently, by choosing V to be sufficiently close to w_M, it follows that $\langle v - u, p\rangle > \sqrt{M}$ for every path vector p. This is a contradiction.
$\diamondsuit$ 2.3.2 $\diamondsuit$

2.3.3. Suppose $w_M = (w_1, \ldots, w_n)$. Then for every $i = 1, \ldots, n$ either $w_i = 0$ or there exists a w_M-minimal path vector $p = (p_1, \ldots, p_n)$ such that $p_i \neq 0$.

Proof. This follows immediately from line 2.3.2.
$\diamondsuit$ 2.3.3 $\diamondsuit$

REMARK. Line 2.3.3 corresponds to Proposition 4.1.2 in Cannon [8]. It is interesting to compare their proofs.

2.3.4. Let w be a weight vector, and let $p_1, \dots, p_k$ be path vectors whose lengths relative to w are minimal, namely, $\langle p_i, w\rangle = H_w$ for $i = 1, \dots, k$. If there exist real numbers $a_1, \dots, a_k$ such that

$$w = \sum_{i=1}^{k} a_i p_i,$$

then

$$A_w = H_w \sum_{i=1}^{k} a_i.$$

Proof. This can be seen from the following.

$$A_w = \langle w, w\rangle = \langle \sum_{i=1}^{k} a_i p_i, w\rangle = \sum_{i=1}^{k} a_i \langle p_i, w\rangle = H_w \sum_{i=1}^{k} a_i.$$

$\diamondsuit$ 2.3.4 $\diamondsuit$

2.3.5. Suppose that $w_M = \sum_{i=1}^{k} a_i p_i$, where $a_i, \dots, a_k$ are real numbers and $p_1, \dots, p_k$ are path vectors whose lengths relative to w_M are minimal. Then $\sum_{i=1}^{k} a_i = \frac{1}{\sqrt{M}}$.

Proof. This is a special case of line 2.3.4.
$\diamondsuit$ 2.3.5 $\diamondsuit$

Line 2.3.2 states that w_M lies in the cone spanned by its minimal path vectors. The next result shows that this property characterizes w_M.

2.3.6. Let w be a weight vector in S^{n-1} such that w lies in the cone spanned by the path vectors whose lengths relative to w are minimal. Then w is the optimal weight vector w_M.

Proof. Let $p_1, \dots, p_k$ be the path vectors whose lengths relative to w are minimal, and let $a_1, \dots, a_k$ be nonnegative real numbers such that $w = \sum_{i=1}^{k} a_i p_i$. Then

$$\langle w, w_M\rangle = \langle \sum_{i=1}^{k} a_i p_i, w_M\rangle = \sum_{i=1}^{k} a_i \langle p_i, w_M\rangle \geq \sum_{i=1}^{k} a_i \sqrt{M} = \frac{\sqrt{M}}{H_w},$$

the last equation coming from line 2.3.4. Since $\sqrt{M} \geq H_w$, $\frac{\sqrt{M}}{H_w} \geq 1$. Thus w and w_M are vectors in S^{n-1} whose inner product is at least 1, and so $w = w_M$.
$\diamondsuit$ 2.3.6 $\diamondsuit$

The next result is well known, but we include its proof for completeness.

2.3.7. Let Q be a finite set of vectors in $\mathbf{R}^n$, and for every $p \in Q$ let b_p be a nonnegative real number so that the vector $x = \sum_{p\in Q} b_p p$ is nonzero. Then there exists a linearly independent subset S of Q and for every $p \in S$ a nonnegative real number a_p such that $x = \sum_{p\in S} a_p p$.

Proof. The proof of line 2.3.7 will proceed by induction on the order $|Q|$ of Q. Because $x \neq 0$, $|Q| \neq 0$ and line 2.3.7 is clear if $|Q| = 1$. Suppose that $|Q| > 1$ and that Q is linearly dependent. Then $\sum_{p\in Q} c_p p = 0$ for some real numbers c_p which are not all 0. It may be assumed that $c_p < 0$ for some p. Then

$$x = \sum_{p\in Q} (b_p + tc_p)p$$

for every nonnegative real number t. Because $c_p < 0$ for some p, there exists a nonnegative value of t for which one of the coefficients $b_p + tc_p$ is 0. Let t_0 be the smallest value of t for which one of these coefficients is 0. Then

$$x = \sum_{p \in Q} (b_p + t_0 c_p) p,$$

which expresses x in the desired form with fewer than $|Q|$ nonzero coefficients.
$\diamondsuit$ 2.3.7 $\diamondsuit$

The next result sharpens line 2.3.2.

2.3.8. A scalar multiple of the optimal weight vector w_M is a linear combination of w_M-minimal path vectors in which the coefficients are nonnegative integers. In particular, some optimal weight vector is a vector of integers.

Proof. To prove line 2.3.8, apply line 2.3.2: w_M is a nonnegative linear combination of path vectors p_1, $\dots$, p_k with $\langle p_i, w_M\rangle = \sqrt{M}$ for $i = 1, \dots, k$. By line 2.3.7 it may furthermore be assumed that $p_1, \dots,$ p_k are linearly independent. Then $\frac{1}{\sqrt{M}} w_M$ is the unique vector v in the subspace generated by $p_1, \dots, p_k$ with $\langle p_i, v\rangle = 1$. In terms of matrices, this means that the invertible matrix, whose entries are the integers $\langle p_i, p_j\rangle$, multiplies the column vector v to the column vector whose entries are all 1. It follows that $\frac{1}{\sqrt{M}} w_M$ is a nonnegative rational linear combination of $p_1, \dots, p_k$. This easily proves line 2.3.8.
$\diamondsuit$ 2.3.8 $\diamondsuit$

We define a **fundamental family of paths** to be a k-tuple $(p_1, \dots, p_k)$ of w_M-minimal path vectors $p_1, \dots, p_k$ such that w_M is a scalar multiple of $\sum_{i=1}^{k} p_i$. Line 2.3.8 shows that a fundamental family of paths always exists.

We define the **reduced integral optimal weight vector** to be the optimal weight vector given by line 2.3.8 whose coordinates are integers with greatest common divisor 1.

2.4. The geometry of an optimal weight function: special results for quadrilaterals and rings. The results of Section 2.3, though introduced in the setting of tiled quadrilaterals, are completely valid in an abstract vector space setting. The results of this section make use of the plane geometry of quadrilaterals and rings. Furthermore, in developing a duality important for the study of the moduli of quadrilaterals and rings, we need to introduce an additional pair of moduli. Thus our treatment requires a substantial number of new definitions.

2.4.1. Further definitions. We define a **quadrilateral** to be a closed topological disk in the plane $\mathbf{C}$ with four distinguished points on its boundary, and we define a **ring** to be a closed topological annulus in the plane $\mathbf{C}$. Unless stated otherwise, X will denote a quadrilateral or ring in Section 2.4. A **shingling** of X is a finite set of compact connected subsets of X, called **shingles**, which cover X. Let S be a shingling of X. A **path** is a nonempty subset of S whose union is connected. A **fat path** is the set of all of the shingles in S which meet a given topological path in X. A **skinny path** is a nonempty subset of S consisting of distinct shingles which can be ordered as $s_1, \dots, s_m$ such that $s_{i-1} \cap s_i \neq \emptyset$ for $i = 2, \dots, m$. A fat path is **closed** if it has an underlying topological path which is closed, and a skinny path as above is **closed** if $s_1 \cap s_m \neq \emptyset$. It is easy to see that fat paths and skinny paths are paths.

If X is a quadrilateral, then one of its four distinguished boundary segments X_1 is called the **top** of X, and if X is a ring then one of its boundary components X_1 is called the **top** of X. The opposite boundary segment or component X_0 is called the **bottom** of X. A **flow** is a path which meets both X_1 and X_0. A **cut** is a path which separates X_1 from X_0 (there does not exist a topological path from X_1 to X_0 which misses all of the shingles in the given path). For brevity it will be said that a cut **separates the ends** of X. A **fat flow** is a fat path which has an underlying topological path joining the ends of X. A **fat cut** is a fat path which has an underlying topological path separating the ends of X. A **skinny flow** is a skinny path whose shingles can be ordered so that one of its extreme shingles meets X_1 and the other extreme shingle meets X_0. If X is a quadrilateral, then a **skinny cut** is analogous to a skinny flow. If X is a ring, then a **skinny cut** is a closed skinny path which separates the ends of X.

Before we continue with our definitions, it is useful to note the following topological fact which we leave to the reader as an exercise.

2.4.1.1. Every flow contains a subflow which is a skinny flow. Every cut contains a subcut which is a skinny cut.

A **weight function** on S is a function which assigns a nonnegative real number to every shingle in S such that some shingle in S has positive weight. Let w be a weight function on S. The **length** of a path relative to w is the sum of the w-weights of the shingles in that path. A **minimal skinny flow** relative to w is a skinny flow of minimal w-length which does not contain a proper skinny subflow. A **minimal skinny cut** relative to w is a skinny cut of minimal w-length which does not contain a proper skinny subcut. Line 2.4.1.1 implies that minimizing skinny flows and cuts is the same as minimizing flows and cuts. The **skinny height** of X relative to w is the length $H_{w,s}$ of a minimal skinny flow relative to w. The **skinny circumference** of X relative to w is the length $C_{w,s}$ of a minimal skinny cut relative to w. The **area** of X relative to w is the sum A_w of the squares of all w-weights. Define the **skinny cut modulus** and **skinny flow modulus** of X relative to S by

$$m_s = \inf_w \frac{A_w}{C^2_{w,s}}, \qquad \text{and} \qquad M_s = \sup_w \frac{H^2_{w,s}}{A_w},$$

where the infimum and supremum vary over all weight functions of the fixed shingling S. We have analogous **minimal fat flows** whose lengths are $H_{w,f}$ and **minimal fat cuts** whose lengths are $C_{w,f}$. Define the **fat cut modulus** and **fat flow modulus** of X relative to S by

$$m_f = \inf_w \frac{A_w}{C^2_{w,f}}, \qquad \text{and} \qquad M_f = \sup_w \frac{H^2_{w,f}}{A_w}.$$

2.4.2. The correspondence. We now point out the natural correspondence between the vector space setting introduced in Section 2.2 and the setting of this section. Suppose that S consists of distinct shingles $s_1, \ldots, s_n$. Then every path p corresponds to a path vector $\bar{p}$ in $\mathbf{N}^n \setminus \{0\}$ as follows. If the shingle s_i belongs to p, then the i-th component of $\bar{p}$ is 1. Otherwise the i-th component of $\bar{p}$ is 0. Likewise, every weight function w corresponds to a weight vector $\bar{w}$ so that the i-th component of $\bar{w}$ is $w(s_i)$. The problem of optimizing skinny flows in the present setting corresponds to the optimization problem in the vector space setting in which the set of path vectors P consists of the path vectors of skinny flows. The situation is the same for fat flows. The correspondence also holds for cuts although the modulus in the vector space setting is the reciprocal of the present modulus. The results of the vector space setting apply to each of the four cases under consideration. We thus have four types of **optimal weight functions**. To simplify notation we will identify p with $\bar{p}$ and w with $\bar{w}$.

2.4.3. Level curves and the relationship between skinny flows and fat cuts.

Consider the optimization problem for skinny flows. According to Theorem 2.2.2 there exists a unique optimal weight function w_{M_s} such that the area $A_{w_{M_s}}$ of X relative to w_{M_s} is 1 and

$$M_s = \frac{H^2_{w_{M_s},s}}{A_{w_{M_s}}} = H^2_{w_{M_s},s}.$$

Just as immediately after line 2.3.8, there is a **reduced integral optimal weight function** w, which is an optimal weight function whose values are integers with greatest common divisor 1. From here until line 2.4.3.4 the weight function w will be fixed. Let H denote the skinny height of X relative to w.

Let h be an integer with $1 \le h \le H$, and define a **level curve** L_h, which is a set of shingles of S, as follows. First, for a shingle s to belong to L_h, it is necessary that $w(s) \neq 0$. Now, given a shingle s in

S with $w(s) \neq 0$, line 2.3.3 implies that there exists a minimal skinny flow f containing s. Suppose that $f = \{s_1, \dots, s_m\}$, where s_1 meets X_1, s_m meets X_0 and $s_{i-1} \cap s_i \neq \emptyset$ for $i = 2, \dots, m$. Also suppose that $s = s_i$. Then s belongs to L_h if the length of the path $\{s_1, \dots, s_{i-1}\}$ is less than h and the length of the path $\{s_1, \dots, s_i\}$ is at least h. To see that this is well-defined, suppose that $g = \{t_1, \dots, t_k\}$ is also a minimal skinny flow from X_1 to X_0 with $s = t_j$ for some j. The minimality of f and g implies that the length of $\{s_1, \dots, s_{i-1}\}$ equals the length of $\{t_1, \dots, t_{j-1}\}$, for otherwise either the skinny flow $\{s_1, \dots, s_{i-1}, t_j, \dots, t_k\}$ or the skinny flow $\{t_1, \dots, t_{j-1}, s_i, \dots, s_m\}$ is shorter than f and g. Thus L_h is well-defined. Line 2.4.3.1 is clear from the definition.

2.4.3.1. Every shingle s in S belongs to exactly $w(s)$ consecutive level curves and every minimal skinny flow contains exactly one shingle in every level curve.

2.4.3.2. Every level curve is a fat cut. Moreover, there exist smooth arcs or simple closed curves $\alpha_1, \dots, \alpha_H$ (depending on whether X is a quadrilateral or a ring) in $X \setminus (X_0 \cup X_1)$ which separate the ends of X such that α_h is contained in the connected component of $X \setminus \alpha_{h-1}$ which contains X_0 for $h = 2, \dots, H$ and the set of shingles which meet α_h is L_h for $h = 1, \dots, H$.

Proof. The curves $\alpha_1, \dots, \alpha_H$ will be defined inductively. To aid this definition, another curve α_0 will be defined to begin the induction. Simply let α_0 be the top of X. No claims such as those in line 2.4.3.2 are being made about α_0.

Having defined α_0, suppose that curves $\alpha_1, \dots, \alpha_{h-1}$ as in line 2.4.3.2 are defined for some h from 1 to H, and define α_h as follows. Let the subset K_1 of X be the union of i) α_{h-1}, ii) the connected component of $X \setminus \alpha_{h-1}$ which contains X_1 (the empty set for $h = 1$) and iii) every shingle s in S for which there exists a skinny path $\{s_1, \dots, s_m\}$ of length less than h such that s_1 meets X_1 and $s_m = s$. Let the subset K_0 of X be the union of i) X_0 and ii) every shingle s in S such that $s \cap K_1 = \emptyset$. Then K_1 and K_0 are compact subsets of X, K_1 contains X_1, K_0 contains X_0 and it is easy to see that $K_1 \cap K_0 = \emptyset$. Now a standard theorem of plane topology implies that there exists a smooth arc or simple closed curve α_h in X which misses $K_1 \cup K_0$ and separates the ends of X.

In this paragraph it will be shown that the set of shingles of S which meet α_h is L_h. First of all it is easy to see that $K_1 \cup K_0$ contains every shingle with weight 0, so α_h misses every shingle with weight 0. Now suppose that s is a shingle in S with $w(s) \neq 0$. If the level curves to which s belongs have indices less than h, then there exists a skinny path $\{s_1, \dots, s_m\}$ of length less than h such that s_1 mets X_1 and $s_m = s$. Thus $s \subset K_1$, and α_h misses s. Now suppose that the level curves to which s belongs have indices greater than h. Let f be a minimal skinny flow containing s. If $h > 1$, then by induction α_{h-1} meets exactly one shingle t in f and $\{t\} = f \cap L_{h-1}$. If $h = 1$, then it is easy to see that α_0 meets exactly one shingle t in f and t is the first shingle in f. Hence s occurs after t in f, and s is contained in the connected component of $X \setminus \alpha_{h-1}$ which contains X_0. It easily follows that $s \cap K_1 = \emptyset$. Thus $s \subset K_0$, and α_h misses s. The argument above shows that every shingle which α_h meets is in L_h. For the opposite inclusion, let $s \in L_h$. Then there exists a minimal skinny flow f containing s. It is easy to see that α_h meets some shingle t in f. Hence $t \in L_h$, and so $t = s$ by line 2.4.3.1. This proves that L_h is precisely the set of shingles which meet α_h.
$\diamondsuit$ 2.4.3.2 $\diamondsuit$

The following results will be proven next.

2.4.3.3. Every level curve is a minimal fat cut.

2.4.3.4. $\sum_{h=1}^{H} L_h = w$

2.4.3.5. $w_{M_s} = w_{m_f}$

2.4.3.6. $M_s = m_f$

To begin these proofs, let c be any fat cut. Then there exists a topological path α that separates the ends of X such that c is the set of shingles in S which meet α. It is easy to see that α meets some shingle in every skinny flow f. Thus c and f have at least one shingle in common, and so $\langle c, f\rangle \geq 1$. According to line 2.3.2 there exist minimal skinny flows $f_1, \ldots, f_k$ and nonnegative real numbers $a_1, \ldots, a_k$ such that $w_{M_s} = \sum_{i=1}^{k} a_i f_i$. Thus

2.4.3.7.
$$\langle c, w_{M_s}\rangle = \langle c, \sum_{i=1}^{k} a_i f_i\rangle = \sum_{i=1}^{k} a_i \langle c, f_i\rangle \geq \sum_{i=1}^{k} a_i = \frac{1}{\sqrt{M_s}},$$

where the last equality comes from line 2.3.5. If c is replaced by a level curve, then line 2.4.3.1 implies that the inequality in line 2.4.3.7 is an equality. This proves line 2.4.3.3. Line 2.4.3.1 implies line 2.4.3.4. Lines 2.3.6 and 2.4.3.4 easily imply line 2.4.3.5. Now line 2.4.3.7 with c replaced by a level curve gives line 2.4.3.6.
$\diamondsuit$ 2.4.3.3–2.4.3.6 $\diamondsuit$

We say that two fat cuts are **parallel** if they have underlying topological paths which are disjoint. There is an obvious analogous notion for fat flows, and there are obvious generalizations to families of parallel fat cuts or flows. If $(p_1, \ldots, p_k)$ is a family of parallel fat cuts or flows, then a **family of parallel underlying topological paths** for $(p_1, \ldots, p_k)$ consists of disjoint topological paths $\alpha_1, \ldots, \alpha_k$ such that α_i is an underlying topological path for p_i for $i = 1, \ldots, k$.

Lines 2.4.3.2 and 2.4.3.4 show that every shingling of X has a fundamental family of parallel fat cuts. It follows that if X is a quadrilateral, then every shingling of X has a fundamental family of parallel fat flows.

2.4.4. The relationship between fat flows and skinny cuts for tilings.

A **tiling** of X is a shingling T of X whose shingles, called **tiles**, are the closed 2-cells in a finite cellular decomposition of X. A **vertex** or **edge** of T is a vertex or edge of the corresponding cellular decomposition of X. Fix a tiling T of X.

2.4.4.1. Every minimal fat flow and minimal fat cut consists of all of the tiles that meet a topological path α which either joins or separates the ends of X and misses every vertex of T.

Proof. Simply observe that if the path α meets a vertex, then it can be slightly deformed to miss that vertex so that the resulting fat path is a subpath of the original one.
$\diamondsuit$ 2.4.4.1 $\diamondsuit$

It easily follows from line 2.4.4.1 that every minimal fat flow and minimal fat cut for a tiling is a skinny flow or a skinny cut. However, such a skinny path need not be a minimal skinny flow or a minimal skinny cut.

Now consider the optimization problem for fat flows for the tiling T. There exists a reduced integral optimal weight function w for this case just as for the case of skinny flows in Section 2.4.3. Let H denote the fat height of X relative to w. Because every minimal fat flow is a skinny flow, it is possible to define level curves just as for the case of skinny flows in Section 2.4.3. The following analog of line 2.4.3.1 clearly holds.

2.4.4.2. Every tile t in T belongs to exactly $w(t)$ consecutive level curves, and every minimal fat flow contains exactly one tile in every level curve.

2.4.4.3. Every level curve is a skinny cut. Moreover, there exist piecewise smooth arcs or simple closed curves $\alpha_1, \ldots, \alpha_H$ (depending on whether X is a quadrilateral or a ring) in X which separate the ends of X such that α_h is contained in the closure of the connected component of $X \setminus \alpha_{h-1}$ which contains X_0 for $h = 2, \ldots, H$ and the tiles in L_h cover α_h for $h = 1, \ldots, H$. The curves α_h meet every tile t in either the empty set, a set of vertices of t or an arc which meets the boundary of t only at its endpoints, and they meet each other and the ends of X only at vertices.

Proof. Let h be an integer with $1 \leq h \leq H$. It will first be shown that L_h separates the ends of X, namely, there does not exist a topological path α joining the ends of X which misses all of the tiles in L_h. Suppose that such an α exists. Then as in the proof of line 2.4.4.1, α can be deformed so that the fat flow which it determines is a skinny flow $f = \{t_1, \dots, t_m\}$. By assumption, $f \cap L_h = \emptyset$. Note that either $w(t_1) = 0$ or $t_1 \in L_1$. Now let i be the largest index such that $t_i \in L_1 \cup \cdots \cup L_{h-1}$ if there is such an index, and if not let $i = 1$. Let j be the smallest index such that $j > i$ and $w(t_j) \neq 0$ if there is such an index, and if not let $j = m$. Then $w(t_k) = 0$ for every index k such that $i < k < j$. Now construct a fat flow g by concatenating i) a minimal fat path from X_1 to t_i, ii) $\{t_{i+1}, \dots, t_{j-1}\}$ and iii) a minimal fat path from t_j to X_0. It follows that the length of g is less than H, which is impossible. Thus L_h separates the ends of X.

Since L_h separates the ends of X, some connected component of the union of the tiles in L_h separates the ends of X. It will next be shown that the union of the tiles in L_h is connected. Choose a connected component of the union of the tiles in L_h which separates the ends of X, and suppose that L_h contains a tile t not in this connected component. Let f be a minimal fat flow containing t. Then f meets the chosen connected component which separates the ends of X. But then f contains at least two tiles in L_h, contrary to line 2.4.4.2. Thus the union of the tiles in L_h is connected, and so L_h is a cut.

It will be convenient to prove the following results next before completing the proof of line 2.4.4.3.

2.4.4.4. Every level curve is a minimal skinny cut.

2.4.4.5. $\sum_{h=1}^{H} L_h = w$.

2.4.4.6. $w_{M_f} = w_{m_s}$.

2.4.4.7. $M_f = m_s$.

These statements will be proven next by following the proofs of lines 2.4.3.3 through 2.4.3.6. Let c be any cut. It is easy to see that c meets every fat flow f. Thus $\langle c, f \rangle \geq 1$. According to line 2.3.2 there exist minimal fat flows $f_1, \dots, f_k$ and nonnegative real numbers $a_1, \dots, a_k$ such that $w_{M_f} = \sum_{i=1}^{k} a_i f_i$. Thus

2.4.4.8.
$$\langle c, w_{M_f} \rangle = \langle c, \sum_{i=1}^{k} a_i f_i \rangle = \sum_{i=1}^{k} a_i \langle c, f_i \rangle \geq \sum_{i=1}^{k} a_i = \frac{1}{\sqrt{M_f}}.$$
If c is replaced by a level curve, then line 2.4.4.2 implies that the inequality in line 2.4.4.8 is an equality. Thus every level curve is a minimal cut. Now line 2.4.1.1 shows that every level curve is a minimal skinny cut. This proves line 2.4.4.4 and the first statement in line 2.4.4.3. Line 2.4.4.2 implies line 2.4.4.5. Lines 2.3.6 and 2.4.4.5 imply line 2.4.4.6. Now line 2.4.4.8 with c replaced by a level curve gives line 2.4.4.7.
$\diamondsuit$ 2.4.4.4–2.4.4.7 $\diamondsuit$

It only remains to construct the curves $\alpha_1, \dots, \alpha_H$ in line 2.4.4.3. A somewhat brief description of such a construction will now be given. Construct α_1 as follows. Suppose that L_1 is given as a skinny cut by the ordered tiles $t_1, \dots, t_m$. If X is a quadrilateral, define points $x_0, \dots, x_m$ in X as follows. It may be assumed that t_1 meets the left side of X and t_m meets the right side of X. If the intersection of t_1 with the left side of X contains an edge, let x_0 be an interior point of that edge, and otherwise let x_0 be a vertex in the intersection of t_1 with the left side of X. In the same way define x_m in the intersection of t_m and the right side of X. Likewise for $i = 1, \dots, m-1$, if $t_i \cap t_{i+1}$ contains an edge, let x_i be an interior point in that edge, and otherwise let x_i be a vertex in $t_i \cap t_{i+1}$. Join x_{i-1} to x_i by a smooth arc whose interior lies in the interior of t_i for $i = 1, \dots, m$, and let α_1 be the concatenation of these arcs. If X is a ring, then define α_1 analogously.

Having defined α_1, suppose that $\alpha_1, \dots, \alpha_{h-1}$ are defined for $h \geq 2$. It is a straightforward matter to define α_h in the same way as α_1, taking care to stay strictly between X_0 and α_{h-1} except at vertices. It is possible to do this because no tile in L_h lies between α_{h-1} and X_1.
$\diamondsuit$ 2.4.4.3 $\diamondsuit$

It is a straightforward matter to define what it means for a family of skinny flows or cuts to be **parallel**. It is also straightforward to define the notion of a **family of parallel underlying arcs** or **simple closed curves** for a family of parallel skinny flows or cuts of a tiling.

2.4.4.9. Let T be a tiling of a quadrilateral or ring X. Then T has a fundamental family of parallel paths of all four types: fat cuts, skinny cuts, fat flows and skinny flows.

Proof. Lines 2.4.3.2 and 2.4.3.4 show that T has a fundamental family of parallel fat cuts. Lines 2.4.4.3 and 2.4.4.5 show that T has a fundamental family of parallel skinny cuts.

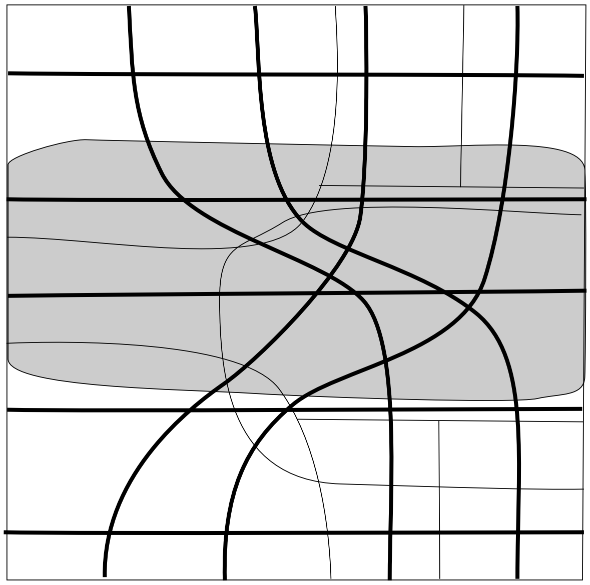
So let $(f_1, \ldots, f_k)$ be either a fundamental family of fat flows or a fundamental family of skinny flows of T. Let $(\alpha_1, \ldots, \alpha_k)$ be a family of underlying piecewise smooth arcs for $(f_1, \ldots, f_k)$. In the case of fat flows, we apply line 2.4.4.1 and assume that no arc α_i contains a vertex of T. Thus the flows $f_1, \ldots, f_k$ are skinny even if they are fat.

It is easy to see that we may make the following assumptions. The set of points in X which are contained in more than one of the arcs α_i is finite. Every arc α_i meets every tile t in T in either the empty set, a set of vertices of t or an arc which meets the boundary of t only at its endpoints.

The argument will now proceed by induction on the number n of triples (x, α_i, α_j) such that x is a point in the interior of X and the intersection number of the arcs α_i and α_j at x is not 0. If $n = 0$, then $f_1, \ldots, f_k$ are parallel, as desired. So suppose that $n > 0$ and that line 2.4.4.9 is true for smaller values of n.

Suppose that (x, α_i, α_j) is a triple such that x is a point in the interior of X and the intersection number of the arcs α_i and α_j at x is not 0. Define new arcs α_i' and α_j' so that α_i' consists of the segment of α_j preceding x followed by the segment of α_i following x and α_j' consists of the segment of α_i preceding x followed by the segment of α_j following x. It is possible to define flows f_i' and f_j' analogously because f_i and f_j are skinny. Just as when defining level curves, it follows that f_i' and f_j' have the same length as f_i and f_j. The result is a new fundamental family of flows and a new family of underlying piecewise smooth arcs, and it is not difficult to see that this operation reduces n. Thus $f_1, \ldots, f_k$ can be inductively transformed to a fundamental family of parallel flows, which proves line 2.4.4.9.
$\diamondsuit$ 2.4.4.9 $\diamondsuit$

EXAMPLE 2.4.4.10. Line 2.4.4.9 shows for every tiling of a quadrilateral or ring that there exists a fundamental family of parallel paths of all four types. The shingling S of a quadrilateral X below indicates that this result does not generalize to skinny flows of shinglings. This shingling has 8 shingles, one of which is shaded just to help identify it. The 5 heavy horizontal line segments joining the sides of X are underlying arcs for a fundamental family of fat cuts. They determine the reduced integral optimal weight function of S for fat cuts and skinny flows. The 4 heavy arcs joining the top and the bottom of X are what might be called underlying arcs for a fundamental family of skinny flows of S. These skinny flows are the only minimal skinny flows of S for this optimal weight function.



2.4.5. The relationship between the four moduli.

2.4.5.1. For every shingling S of a quadrilateral or ring X,

$$M_s = m_f \leq m_s \leq M_f.$$

Proof. The equality is given by line 2.4.3.6. To prove the first inequality, let c be a minimal fat cut relative to the weight function w_{m_s}. By line 2.4.1.1 c contains a subcut b which is a skinny cut. Thus

$$m_f \leq \frac{1}{\langle c, w_{m_s}\rangle^2} \leq \frac{1}{\langle b, w_{m_s}\rangle^2} \leq \frac{1}{C^2_{w_{m_s},s}} = m_s,$$

which proves the first inequality. For the second inequality, let c be a minimal skinny cut relative to the weight function w_{M_f}. Although line 2.4.4.8 occurs in the context of tilings, it is also valid for general shinglings. Using line 2.4.4.8 it can easily be seen that

$$m_s \leq \frac{1}{\langle c, w_{M_f}\rangle^2} \leq M_f.$$

♢ 2.4.5.1 ♢

2.4.5.2. For every shingling S of a quadrilateral X,

$$M_s = m_f \leq m_s = M_f.$$

Proof. This follows easily from line 2.4.5.1.
♢ 2.4.5.2 ♢

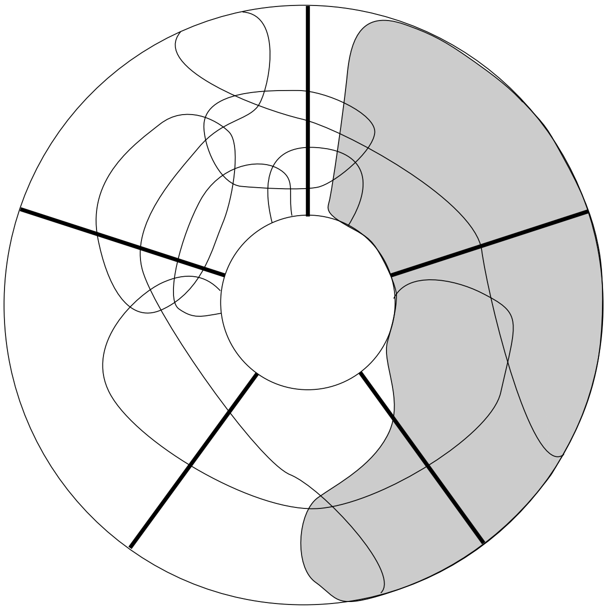
2.4.5.3. For every tiling T of a quadrilateral or ring X,

$$M_s = m_f \leq m_s = M_f.$$

Proof. This result follows from lines 2.4.5.1 and 2.4.4.7.
♢ 2.4.5.3 ♢

Lines 2.4.5.2 and 2.4.5.3 show that $m_s = M_f$ unless the shingling S is not a tiling and X is a ring. Following is a shingling of a ring for which $m_s < M_f$.

Example 2.4.5.4. Consider the shingling S of a ring X given below. This shingling has 8 shingles, one of which is shaded just to help identify it. The 5 heavy arcs joining the top and bottom of X are underlying arcs for a fundamental family of parallel fat flows of S. They determine the reduced integral optimal weight function w of S for fat flows. Hence $M_f = \frac{H_{w,f}^2}{A_w} = \frac{16}{20} = \frac{4}{5}$. However, there is just one skinny cut which is minimal for w, and it has w-length 5. Hence $\frac{A_w}{C_{w,s}^2} = \frac{20}{25} = \frac{4}{5}$. Since w is not in the cone spanned by its minimal skinny cuts, w is not an optimal weight function for skinny cuts of S by line 2.3.2. Thus $m_s < \frac{A_w}{C_{w,s}^2} = M_f$.



3. The finite Riemann mapping theorem.

We consider a quadrilateral or ring X and a tiling T of X. We optimize fat flows and skinny cuts so as to realize the equality $M_f = m_s$. The analysis of the previous sections shows that the reduced integral optimal weight function can be simultaneously realized as a sum of a fundamental family of parallel fat flows and a fundamental family of parallel skinny cuts. These minimal fat flows have underlying disjoint arcs joining the top and bottom of X, and these minimal skinny cuts have underlying arcs that are disjoint except where they pass through vertices of the tiling T, each arc separating the ends of X. Finally, each arc underlying a flow intersects each arc underlying a cut in precisely one point which is in the interior of some tile. As in Section 1.4, these arcs form a grid which can be pictured as graph paper that indicate how X can be realized as a rectangle or right circular cylinder tiled by squares (squares, since the number of flows passing through the interior of a tile in T is the same as the number of cuts passing through the same interior). Thus we have the following remarkable theorem.

Theorem 3.0.1. (Finite Riemann Mapping Theorem). *Every tiled quadrilateral corresponds uniquely under fat flow optimization to a squared rectangle. Under fat cut optimization it corresponds uniquely to another squared rectangle. Analogous results hold for tiled rings.*

Since squared rectangles solve the resistive circuit problems given by their Smith diagrams (Section 1), tiled quadrilaterals have a natural correspondence with planar circuits with edges representing resistors of unit resistance. There are two natural questions given by this correspondence.

Question 3.0.2. Of the many natural circuits that one could conceivably associate with a tiled quadrilateral, how do the ones associated with fat flow and fat cut optimization differ from the others?

Question 3.0.3. How can we find the optimal circuits associated with a tiled quadrilateral?

We shall answer Question 3.0.2 in Section 3.1. The answer is essentially this: our optimization problems give circuits whose squared rectangles are as like the original tiling combinatorially as possible. Question 3.0.3 is an important question whose answer relates intimately to the problem of calculating the solution to the finite Riemann mapping problem explicitly. We shall discuss algorithms designed to solve this mapping problem in Section 4. Obtaining the appropriate circuits will be discussed particularly in Section 4.4.

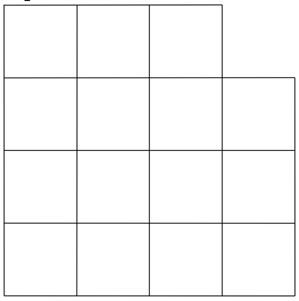


Figure 3.1.1.

3.1. Examples. In Figure 3.1.1 we give a sample tiled quadrilateral: a 4×4 square tiled by unit squares from which the upper right hand corner square has been removed. We consider the two upper right hand "indented" edges as the "top" of our quadrilateral and the left side and bottom of the square as the "bottom" of our quadrilateral. In Figure 3.1.2 we interpret certain collections of edges as potential "maximal horizontal segments" by means of which we form three associated Smith circuit diagrams. The Smith diagrams appear in Figure 3.1.3, and the associated squared rectangles appear in Figure 3.1.4.

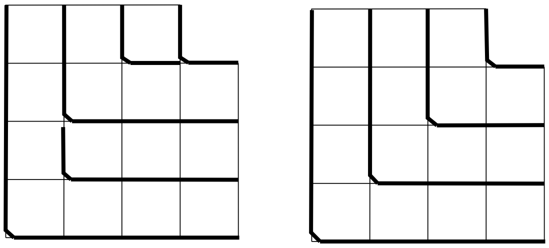


Figure 3.1.2.

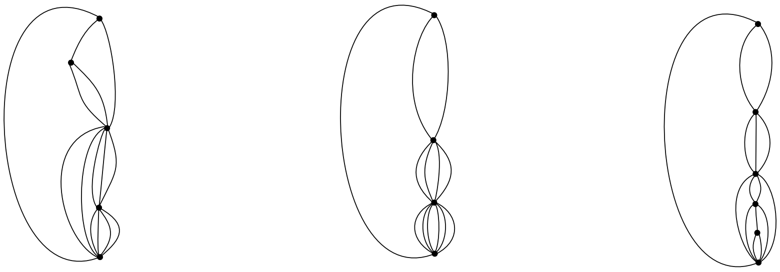


Figure 3.1.3.

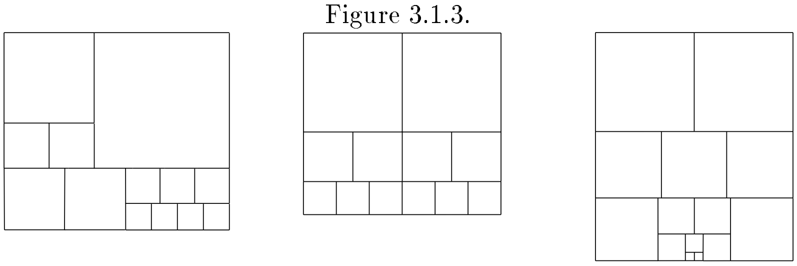


Figure 3.1.4.

In each of Figures 3.1.2, 3.1.3, and 3.1.4 the last of the three parallel figures represents the appropriate diagram for optimization of fat flows and skinny cuts. Note that the first two of the three parallel figures, while in some sense geometrically reasonable, do more geometric damage to the original combinatorics in the process of squaring than does the optimization squaring. The damage is easily quantifiable if we refer to the graph paper grid of the squaring developed in Section 1.4 given by the collections of current curves and voltage curves. If these curves are developed in the original tiling, then in the optimization pattern the curves can actually be realized as paths which travel from one tile into an adjacent tile. The fat flows cross edges from one tile to the next. The skinny cuts cross edges or vertices from one tile to the next. In the other diagrams the curves may jump from one tile to a nonadjacent tile. That is, **the optimization curves of the finite Riemann mapping theorem respect, in as far as is possible, the combinatorics of the original tiling.** There is a converse which will be studied in Section 6.1: **if the current and voltage curves, when developed in the original tiling, always pass from tile to adjacent tile, then the resulting modulus or resistance of the circuit (height/width) lies between m_f and M_f.**

In summary, the optimization conditions of the finite Riemann mapping theorem delineate a family of circuits related to an arbitrary tiled quadrilateral which, when solved, square a rectangle in a way which is combinatorially optimal or nearly optimal.

3.2. The Kirchhoff inequalities. We have demonstrated that the finite Riemann mapping theorem squares a rectangle, hence is associated with the solution of a finite resistive circuit with one battery. The difficulty already noted is that it is not at all clear from a tiled quadrilateral what the appropriate circuit should be. Nevertheless, we can formulate a family of equations and inequalities analogous to the Kirchhoff equations whose solution solves the finite Riemann mapping problem. We call these equations and inequalities the **Kirchhoff inequalities**. Unfortunately, there is a certain degree of circularity in the formulation which makes the inequalities suitable for *checking* a potential solution but not readily applicable for *finding* a potential solution.

The formulation views the dual graph of the tiling as a generalized electrical circuit in which resistances appear not on the edges but at the nodes. Such an interpretation was developed in discussions with Peter Doyle and his students, including Oded Schramm, at Princeton during the year 1988-89. Schramm, in his study of the finite Riemann mapping theorem, has made some effort to make the analogy precise.

We add to the tiling two new tiles t_0 and t_1 representing the bottom and top edge of the quadrilateral, respectively. We consider the undirected graph dual to the tiling. The tiles represent vertices of the same name. Pairs of tiles meeting along an edge represent edges of the dual graph.

The variables. If s and t are tiles meeting along an edge, then the pairs (s,t) and (t,s) represent an edge of the dual graph. We assign to this edge a variable $i(s,t) = -i(t,s)$ which by convention represents current flow from tile s to tile t along edge (s,t). If $i(s,t) > 0$, then we think of this current as flowing *out* of the tile s and *into* the tile t. The **weight** of a tile t is denoted by $w(t)$ and is the sum of the currents flowing *into* the tile t. Note that $w(t) \geq 0$. Our aim is to solve the finite Riemann mapping problem by using the nonnegative numbers $w(t)$, $t \neq t_0, t_1$, as our weight function. The Kirchhoff inequalities state necessary and sufficient conditions on these weights in order that they optimize fat flows from t_0 to t_1.

The vertex equations. For every vertex s other than t_0 and t_1, the sum of the currents flowing into s equals the sum of the currents flowing out of s.

The loop equations. Let $s_0, \ldots, s_k$ denote a sequence of tiles forming the vertices of a closed path or loop in our dual graph. Assume in addition that each edge (s_0,s_1), (s_1,s_2), $\ldots$, (s_k,s_0) carries nonzero current. We say that the loop is *extreme* at vertex s_i if the two currents $i(s_{i-1},s_i)$ and $i(s_i,s_{i+1})$ have opposite signs. The loop is *rising* at vertex s_i if the two currents just mentioned are positive, and *falling* at vertex s_i if the two currents are negative. The loop equation associated with our given loop requires that we add the weights of the vertices at which the loop is rising and subtract the weights of the vertices at which the loop is falling. The sum should then be 0.

The loop inequalities. We need to accommodate the possibility that some edges may carry no current ($i(s,t) = 0$). We define the length of a fat path to be the sum of the weights of its vertices. We require that at least one minimal path from t_0 to t_1 be strictly rising in the sense of the previous paragraph at each vertex other than t_0 and t_1. This requirement is equivalent to a whole family of inequalities that imply that a rising path cannot be shortened by inserting splices across edges which carry no current.

Theorem 3.2.1. *Suppose that there exist currents $i(s,t)$, not all 0, satisfying the Kirchhoff inequalities. Then the associated weights $w(s)$, $s \neq t_0, t_1$, form an optimal weight function w which optimizes fat flows from t_0 to t_1.*

Proof. Form $|i(s,t)|$ current lines parallel to the edge (s,t) of the dual graph in the direction of current flow. The vertex equations imply that, except at the extreme vertices, t_0 and t_1, the current lines may be extended through the vertices. By the definition of the weight function w, we see that w is the sum of the resulting paths given by the current lines. Each current line is a path which is rising at each of its vertices with the possible exception of t_0 and t_1.

The loop equations imply that no current path can be closed; for otherwise, since the path is rising at each vertex, the weight sum around it would be positive. Hence each current line must be an arc joining t_0 and t_1. The Kirchhoff inequalities guarantee the existence of a strictly rising minimal path from t_0 to t_1. If some current path rises from t_1 to t_0, then the concatenation of this and a path rising from t_0 to t_1 is a closed loop rising at every vertex other than t_0 and t_1, again contradicting the associated loop equation. Hence all current flows from t_0 to t_1. Furthermore, the loop equations imply that all have the same length.

Since the current paths are w-minimal flows from t_0 to t_1, w is a sum of its minimal flows. By line 2.3.6 w is an optimal weight function for fat flows.
◊ Theorem 3.2.1 ◊

4. Algorithms which calculate the finite Riemann mapping.

Although there does not seem to be a simple system of linear equations whose solution gives an optimal weight function for the finite Riemann mapping problem, nevertheless there are finite algorithms for calculating an optimal weight function. We give a beautiful algorithm which is slow and a hybrid algorithm that is substantially faster.

4.1. The minimal path algorithm. This algorithm involves the construction of a sequence of weight vectors. Line 4.1.1 is a description of this construction. We return to the vector space setting introduced in Section 2.2.

4.1.1. For every nonempty subset Q of the set P of paths fix a vector x_Q in $\mathbf{N}^n \setminus \{0\}$ which lies in the cone spanned by the path vectors in Q. Inductively define a sequence of weight vectors w_1, w_2, w_3, $\ldots$ in $\mathbf{N}^n$ as follows. Let w_1 be any weight vector in $\mathbf{N}^n$. Now suppose that w_ν is defined for some integer $\nu \geq 1$. Let Q_ν be the set of w_ν-minimal path vectors in P, and set $w_{\nu+1} = w_\nu + x_{Q_\nu}$.

The following three statements make successively stronger statements about the sequence w_ν. The first shows that the sequence, normalized to have norm 1, converges to the optimal weight vector of norm 1. The third is a key fact which will allow us to extract an exact solution to the mapping problem from the sequence. The remainder of Section 4.1 will be devoted to a description of the minimal path algorithm. We prove lines 4.1.2, 4.1.3 and 4.1.4 in Section 4.2.

4.1.2. The sequence of weight vectors $\frac{1}{||w_1||}w_1$, $\frac{1}{||w_2||}w_2$, $\frac{1}{||w_3||}w_3$, $\ldots$ in S^{n-1} converges to the optimal weight vector w_M.

4.1.3. The sequence of subsets Q_1, Q_2, Q_3, $\ldots$ of P is eventually periodic; that is, there exist positive integers N and μ such that if $\nu \geq N$, then $Q_{\mu+\nu} = Q_\nu$.

4.1.4. In the notation of line 4.1.3, if $\nu \geq N$, then $w_{\mu+\nu} - w_\nu$ is an optimal weight vector.

Here is how an exact algorithm can be extracted from line 4.1.4.

4.1.5. For every subset Q of P choose x_Q to be the sum of all of the path vectors in Q. Choose $w_1 = (1,1,1,\dots)$.

Lines 4.1.1 and 4.1.5 determine a sequence w_1, w_2, w_3, ... of weight vectors in $\mathbf{N}^n$.

At this point one encounters the problem of determining when one has an optimal weight vector as in line 4.1.4. Here is one approach to this problem.

4.1.6. At step number ν compute w_ν, record w_ν and record $\sum_{p\in Q_\nu} a_p$, where $x_{Q_\nu} = \sum_{p\in Q_\nu} a_p p$. Then let the integer ν' vary from 1 to $\nu - 1$. Choose such a ν', and set $w = w_\nu - w_{\nu'}$.

If x_Q is chosen as in line 4.1.5, then every a_p in line 4.1.6 equals 1.

This paragraph describes a test, which if satisfied, implies that w is an optimal wieght vector. For this, compute H_w and A_w. There exist positive real numbers $b_1,\dots,b_k$ and path vectors $p_1,\dots,p_k$ such that $w = \sum_{i=1}^k b_i p_i$. The b_i's and the p_i's are easily gotten from the a_p's and the p's which occur in line 4.1.6. Hence

4.1.7.

$$A_w = \langle w,w\rangle = \langle \sum_{i=1}^k b_i p_i, w\rangle = \sum_{i=1}^k b_i\langle p_i,w\rangle \geq H_w \sum_{i=1}^k b_i.$$

Furthermore, equality holds throughout line 4.1.7 if and only if p_i is a w-minimal path vector for every i. Thus if $A_w = H_w \sum_{i=1}^k b_i$, then w lies in the cone spanned by its minimal path vectors, and so $\frac{1}{||w||}w = w_M$ by line 2.3.6. So, to test whether or not w is an optimal weight vector,

4.1.8. compute A_w, H_w and $\sum_{i=1}^k b_i$; if $A_w = H_w \sum_{i=1}^k b_i$, then w is an optimal weight vector.

It remains to show that there exist positive integers $\nu' < \nu$ such that equality holds throughout line 4.1.7. For this it is easy to see that there exists a neighborhood U of w_M in S^{n-1} such that if p is a path vector whose length is minimal relative to some weight vector in U, then the length of p is minimal relative to w_M. Combining this with line 4.1.2 shows that if ν' is sufficiently large, then the length of every path vector in $Q_{\nu'}, Q_{\nu'+1}, Q_{\nu'+2},\dots$ is minimal relative to w_M. This and line 4.1.4 show that there exist ν' and ν such that w is a scalar multiple of w_M and $Q_{\nu'}, Q_{\nu'+1}, Q_{\nu'+2},\dots$ consist of path vectors whose lengths are minimal relative to w_M. For such integers ν' and ν, equality holds throughout line 4.1.7. Thus the algorithm eventually stops.

In conclusion, to compute w_M, compute weight vectors $w_1, w_2, w_3, \dots$ and their differences as described in lines 4.1.1, 4.1.5 and 4.1.6. Line 4.1.8 describes a test which eventually identifies a difference of the form $w_\nu - w_{\nu'}$ as an optimal weight vector. At this point the algorithm stops.

This paragraph deals with a slight modification of the algorithm described in the previous paragraph which improves computing efficiency. Instead of allowing ν' to be any integer from 1 to $\nu - 1$, take ν' to be the largest power of 2 less than ν. In other words, first take $\nu' = 1$ and $\nu = 2$, then $\nu' = 2$ and $\nu = 3, 4$, then $\nu' = 4$ and $\nu = 5, 6, 7, 8$ and so on. In the case where ν' is any integer from 1 to $\nu - 1$, the number of times that the computation in line 4.1.8 must be performed grows quadratically in $\mu + N$, where N and μ are as in line 4.1.3. In the case where ν' is taken to be a power of 2, the number of times that the computation in line 4.1.8 must be performed grows linearly in $\mu + N$.

4.2. Proofs of the claims about the minimal path algorithm. This section is concerned with proving lines 4.1.2, 4.1.3, and 4.1.4. The first thing to be done here is to present a construction of sequences of weight vectors which is more general than the construction given in line 4.1.1.

Fix two positive real numbers $K < L$. For every nonempty subset Q of P, let $C_Q(K,L)$ denote the set of vectors x in $\mathbf{R}^n$ which lie in the cone spanned by the path vectors in Q such that $K \leq ||x|| \leq L$. Inductively define a sequence of weight vectors w_1, w_2, w_3, $\ldots$ in $\mathbf{R}^n$ as follows. Let w_1 be any weight vector in $\mathbf{R}^n$. Now suppose that w_ν is defined for some integer $\nu \geq 1$. Let Q be the set of w_ν-minimal path vectors in P. Choose a vector x_ν in $C_Q(K, L)$, and set $w_{\nu+1} = w_\nu + x_\nu$.

This construction is clearly more general than the one given in line 4.1.1. Differences between this construction and the one in line 4.1.1 are that the previous weight vectors w_ν are weight vectors of integers and the previous w_ν uniquely determines the previous $w_{\nu+1}$. Thus to prove line 4.1.2, it suffices to prove line 4.2.1 for the present construction.

4.2.1. The sequence of weight vectors $\frac{1}{||w_1||}w_1$, $\frac{1}{||w_2||}w_2$, $\frac{1}{||w_3||}w_3$, $\ldots$ in S^{n-1} converges to the optimal weight vector w_M.

The proof of line 4.2.1 will now begin. Let V denote the linear hyperplane in $\mathbf{R}^n$ which is orthogonal to w_M. Let $\pi : \mathbf{R}^n \to V$ denote the canonical projection. Thus for every x in $\mathbf{R}^n$, $x = \pi(x) + \langle x, w_M \rangle w_M$.

The assertion in line 4.2.1 is equivalent to saying that if C is a cone in $\mathbf{R}^n$ which is a neighborhood of w_M, then C contains all but finitely many of the vectors w_1, w_2, w_3, $\ldots$. Because the coordinates of the vectors x_1, x_2, x_3, $\ldots$ are nonnegative and $||x_\nu|| \geq K$ for every ν, the lengths of w_1, w_2, w_3, $\ldots$ increase monotonically to ∞. Combining the last two sentences easily shows the following.

4.2.2. To prove line 4.2.1 it suffices to prove that the sequence of vectors $\pi(w_1)$, $\pi(w_2)$, $\pi(w_3)$, $\ldots$ is bounded.

In other words, if the vectors w_1, w_2, w_3, $\ldots$ are contained in a tubular neighborhood of the ray spanned by w_M, then every cone which is a neighborhood of w_M contains all but finitely many of them.

This leads to an investigation of the projection π. The first result in this direction is the following.

4.2.3. If w is a weight vector and p is a w-minimal path vector, then

$$\langle \pi(p), \pi(w) \rangle \leq 0.$$

To begin the proof of line 4.2.3, let p_1, $\ldots$, p_k be the w_M-minimal path vectors. By line 2.3.2 there exist nonnegative real numbers a_1, $\ldots$, a_k such that $w_M = \sum_{i=1}^k a_i p_i$. Hence

$$0 = \pi(w_M) = \pi(\sum_{i=1}^k a_i p_i) = \sum_{i=1}^k a_i \pi(p_i),$$

that is, 0 is in the convex hull of the vectors $\pi(p_1)$, $\ldots$, $\pi(p_k)$. This easily implies that if w is any weight vector, then $\langle \pi(p_i), \pi(w) \rangle \leq 0$ for some i. Thus for this value of i

4.2.4.

$$\begin{aligned} H_w &\leq \langle p_i, w \rangle \\ &= \langle \pi(p_i) + \langle p_i, w_M \rangle w_M, \pi(w) + \langle w, w_M \rangle w_M \rangle \\ &= \langle \pi(p_i), \pi(w) \rangle + \langle p_i, w_M \rangle \langle w, w_M \rangle \\ &\leq H_{w_M} \langle w, w_M \rangle. \end{aligned}$$

On the other hand, if p is a w-minimal path vector, then

4.2.5.

$$\begin{aligned} H_w &= \langle p, w \rangle \\ &= \langle \pi(p), \pi(w) \rangle + \langle p, w_M \rangle \langle w, w_M \rangle \\ &\geq \langle \pi(p), \pi(w) \rangle + H_{w_M} \langle w, w_M \rangle. \end{aligned}$$

Combining lines 4.2.4 and 4.2.5 gives

$$\langle \pi(p), \pi(w) \rangle + H_{w_M}\langle w, w_M \rangle \leq H_{w_M}\langle w, w_M \rangle,$$

and so

$$\langle \pi(p), \pi(w) \rangle \leq 0.$$

This proves line 4.2.3.
◊ 4.2.3 ◊

Now let ν be a positive integer, and consider how $\pi(w_{\nu+1})$ is gotten from $\pi(w_\nu)$. Let Q be the set of w_ν-minimal path vectors. By definition $w_{\nu+1} = w_\nu + x_\nu$, where x_ν is an element of $C_Q(K, L)$. In particular, $\pi(w_{\nu+1}) = \pi(w_\nu) + \pi(x_\nu)$. Furthermore, since $||x_\nu|| \leq L$, it is easy to see that $||\pi(x_\nu)|| \leq L$, and line 4.2.3 shows that $\pi(x_\nu)$ lies in the cone spanned by the vectors in $\pi(P)$ whose inner products with $\pi(w_\nu)$ are nonpositive.

The discussion from line 4.2.2 to here shows that proving line 4.2.1 reduces to proving line 4.2.6, which follows shortly.

Notation will change from here to line 4.2.18, which is the end of the proof of line 4.2.6. The essential difference is that the projection π is omitted. The integer $n-1$ becomes m.

To prepare for the statement of line 4.2.6, first fix some Euclidean space V of dimension $m \geq 1$. Let P be a finite set of vectors in V with 0 in its convex hull. Fix a positive real number L, and for every nonempty subset Q of P, let $C_Q(L)$ denote the set of vectors x in V which lie in the cone spanned by the vectors in Q such that $||x|| \leq L$. Inductively define a sequence of vectors $w_1, w_2, w_3, \ldots$ in V as follows. Let w_1 be any vector in V. Now suppose that w_ν is defined for some integer $\nu \geq 1$. Let Q be the set of vectors p in P such that $\langle p, w_\nu \rangle \leq 0$. Choose a vector x_ν in $C_Q(L)$, and set $w_{\nu+1} = w_\nu + x_\nu$.

4.2.6. The sequence of vectors $w_1, w_2, w_3, \ldots$ in V is bounded.

It is easy to see that proving line 4.2.6 in turn reduces to proving line 4.2.7. The following definitions are needed to state line 4.2.7.

Let $\lambda_0, \ldots, \lambda_m$ be real numbers such that

$$0 = \lambda_m < \cdots < \lambda_2 < \lambda_1 < \lambda_0 = 1.$$

For every subset Q of P let V_Q be the subspace of V spanned by Q, let $d_Q = \dim(V_Q)$ and for every positive real number r let

$$N_Q(r) = \{x + y \in V | x \in V_Q, y \in V_Q^\perp, ||y|| \leq \lambda_{d_Q} r\}.$$

The set $N_Q(r)$ consists of all vectors in V having distance at most $\lambda_{d_Q} r$ from V_Q. Let

$$B(r) = \cap_{Q \subset P} N_Q(r).$$

It is possible for Q to be the empty set, and $N_\emptyset(r)$ is the closed ball of radius r in V centered at 0. Thus $B(r)$ is contained in this closed ball of radius r.

4.2.7. There exists a choice of $\lambda_1, \ldots, \lambda_{m-1}$ and a positive real number R all of which depend only on P and L such that the following holds. If $w \in B(r)$ for some real number $r \geq R$ and Q is the set of vectors p in P such that $\langle p, w \rangle \leq 0$, then $w + x \in B(r)$ for every vector x in $C_Q(L)$.

To begin the proof of line 4.2.7 observe that $N_Q(r)$ is convex for every Q and r. Thus $B(r)$ is convex for every r. Moreover, since $N_Q(r) = rN_Q(1)$ for every Q and r, $B(r) = rB(1)$ for every r.

Now suppose that $w \in B(r)$ and $x \in C_Q(L)$ as in line 4.2.7. Then $x = \sum_{p \in Q} a_p p$ for some nonnegative real numbers a_p, and $\|x\| \leq L$. Since line 4.2.7 is obvious if $x = 0$, it may be assumed that $x \neq 0$. Line 2.3.7 shows that we may assume that the set of vectors $S = \{p \in Q : a_p \neq 0\}$ is linearly independent. Hence there exists a linear functional $f : V \to \mathbf{R}$ such that $f(p) = 1$ for every p in S. Clearly, $f(x) = \sum_{p \in Q} a_p$. Because $C_Q(L)$ is compact, $f(x)$ is bounded by a bound which depends only on Q and L. This and the fact that P is a finite set imply the following.

4.2.8. There exists a positive real number Λ which depends only on P and L such that it is possible to express x as $x = \sum_{p \in Q} a_p p$, where the a_p's are nonnegative real numbers and $(\sum_{p\in Q} a_p)\|q\| \leq \Lambda$ for every q in Q.

Now suppose that x is expressed as in line 4.2.8. Let $a = \sum_{p\in Q} a_p$, and let $b_p = a_p/a$ for every p in Q. Then $x = \sum_{p\in Q} b_p(ap)$, $\sum_{p\in Q} b_p = 1$ and $\|ap\| \leq \Lambda$ for every p in Q. In particular, this expresses x as a convex combination of the vectors ap.

Suppose that it can be shown that $w + ap \in B(r)$ for every p in Q. Then because $B(r)$ is convex, it also contains

$$\sum_{p\in Q} b_p(w + ap) = \sum_{p\in Q} b_p w + \sum_{p\in Q} b_p(ap) = w + x.$$

Thus to prove line 4.2.7 it suffices to prove the following.

4.2.9. There exists a choice of $\lambda_1, \ldots, \lambda_{m-1}$ and a positive real number R all of which depend only on P and L such that the following holds. If $w \in B(r)$ for some real number $r \geq R$ and p is a vector in P such that $\langle p, w\rangle \leq 0$, then $w + ap \in B(r)$ for every nonnegative real number a such that $\|ap\| \leq \Lambda$.

The next thing to do is to choose the numbers $\lambda_1, \ldots, \lambda_{m-1}$. To prepare for this the following definition will be made. Given a subset Q of P and a positive real number r, let

$$D_Q(r) = V_Q^\perp \cap (\cap_{Q\subset V_{Q'}} N_{Q'}(r)).$$

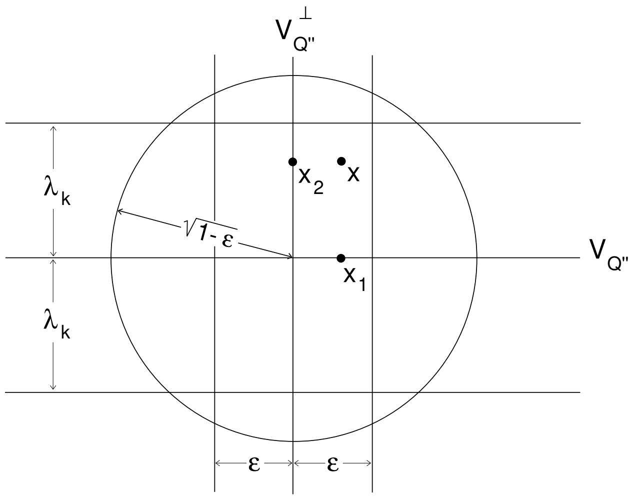
The choice of $\lambda_1, \ldots, \lambda_{m-1}$ will be made so that the following holds.

4.2.10. If Q is a subset of P consisting of exactly one nonzero vector and Q' is a subset of P for which $V_{Q'}$ does not contain Q, then $N_{Q'}(1)$ is a neighborhood of $D_Q(1)$.

The numbers $\lambda_1, \ldots, \lambda_{m-1}$ will be chosen so that line 4.2.10 holds for all such subsets Q' by means of a backward induction on $d_{Q'}$. The case $d_{Q'} = m$ is vacuously true with no restriction on $\lambda_1, \ldots, \lambda_{m-1}$. Now suppose that $d_{Q'} = k-1$ for some integer k with $1 \leq k \leq m$ and that $\lambda_k, \ldots, \lambda_{m-1}$ have been chosen to satisfy line 4.2.10 for subsets of the form Q' which generate subspaces of V having dimensions larger than $d_{Q'}$.

Set $Q'' = Q' \cup Q$. Since $V_{Q'}$ does not contain Q and Q contains exactly one nonzero vector, $d_{Q''} = k$. Let ϵ be a positive real number such that $\epsilon^2 + \epsilon < 1 - \lambda_k^2$. Lines 4.2.11 and 4.2.12 will be needed to choose λ_{k-1}.

4.2.11. If $x \in D_Q(1)$ and the distance from x to $V_{Q''}^\perp$ is at most ϵ, then $\|x\| < \sqrt{1-\epsilon}$.



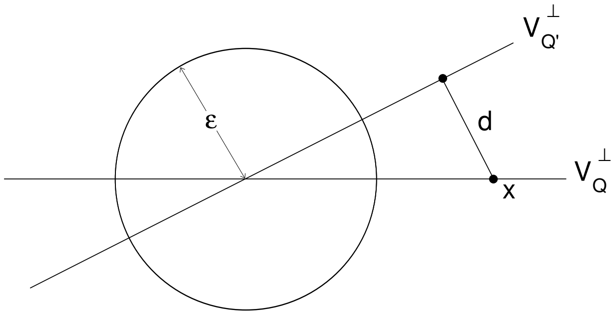
The proof of line 4.2.11 follows easily from the fact that if $x \in D_Q(1)$, then $x \in N_{Q''}(1)$, and so $x = x_1 + x_2$ with $x_1 \in V_{Q''}$, $x_2 \in V_{Q''}^{\perp}$, $\|x_1\| \leq \epsilon$ and $\|x_2\| \leq \lambda_k$. Thus

$$\|x\|^2 \leq \epsilon^2 + \lambda_k^2 < (1 - \epsilon - \lambda_k^2) + \lambda_k^2 = 1 - \epsilon.$$

This proves line 4.2.11.
◊ 4.2.11 ◊

The next thing to be proven is line 4.2.12.

4.2.12. There exists a positive real number $\delta < 1$ for which the following holds. For every choice of Q, Q', and Q'' as immediately before line 4.2.11, if x is a vector in $V_Q^{\perp}$ whose distance from $V_{Q''}^{\perp}$ is greater than ϵ, then the distance from x to $V_{Q'}^{\perp}$ is greater than δ.



To prove line 4.2.12 first choose Q, Q', and Q'' as immediately before line 4.2.11. Choose sets of vectors S_1, S_2, and S_3 in V such that S_1 is a basis of $V_{Q''}^{\perp}$, $S_1 \cup S_2$ is a basis of $V_Q^{\perp}$ and $S_1 \cup S_3$ is a basis of $V_{Q'}^{\perp}$. It is easy to see that S_3 contains just one vector, and $S_3 \not\subset V_Q^{\perp}$ because $Q \not\subset V_{Q'}$. Thus $S_1 \cup S_2 \cup S_3$ is a basis of V.

Now choose a linear isomorphism $T : V \to \mathbf{R}^m$ such that the sets $T(S_1)$, $T(S_2)$, $T(S_3)$ are mutually orthogonal. Then $T(x) = x_1 + x_2$, where x_i is in the subspace of $\mathbf{R}^m$ generated by $T(S_i)$ for $i = 1, 2$. Because T is a linear isomorphism, there exists a positive real number ϵ' independent of x such that the distance from $T(x)$ to $T(V_{Q''}^{\perp})$ is greater than ϵ'. In other words, $\|x_2\| > \epsilon'$. But the distance from $T(x)$ to $T(V_{Q'}^{\perp})$ is also $\|x_2\|$. Thus for the same reason that ϵ' exists there exists a positive real number δ' independent of x such that the distance from x to $V_{Q'}^{\perp}$ is greater than δ'. This proves line 4.2.12 because there are only finitely many possibilities for Q, Q', and Q''.
◊ 4.2.12 ◊

Now that lines 4.2.11 and 4.2.12 are established, choose λ_{k-1} so that $\sqrt{1-\epsilon} \leq \lambda_{k-1}$ and $\sqrt{1-\delta^2} \leq \lambda_{k-1}$. To prove that $N_{Q'}(1)$ is a neighborhood of $D_Q(1)$, choose $x \in D_Q(1)$. If the distance from x to $V_{Q''}^{\perp}$ is at most ϵ, then $\|x\| < \sqrt{1-\epsilon} \leq \lambda_{k-1}$ by line 4.2.11. Thus $N_{Q'}(1)$ is a neighborhood of x because $N_{Q'}(1)$ contains the closed ball in V of radius λ_{k-1} centered at 0. On the other hand, if the distance from x to $V_{Q''}^{\perp}$ is greater than ϵ, then the distance from x to $V_{Q'}^{\perp}$ is greater than δ by line 4.2.12. Since $\|x\| \leq 1$, this implies that the length of the projection of x to $V_{Q'}^{\perp}$ is less than $\sqrt{1-\delta^2}$. Since $\sqrt{1-\delta^2} \leq \lambda_{k-1}$, it follows that $N_{Q'}(1)$ is a neighborhood of x. The proof of line 4.2.10 is now complete.
◊ 4.2.10 ◊

Let Q continue to denote a subset of P which consists of exactly one nonzero vector. Line 4.2.10 implies that

4.2.13. $\cap_{Q \not\subset V_{Q'}} N_{Q'}(1)$ is a neighborhood of $D_Q(1)$.

Thus

4.2.14. $D_Q(1) = V_Q^{\perp} \cap B(1)$.

On the other hand it is easy to see that

4.2.15.
$$\cap_{Q\subset V_{Q'}} N_{Q'}(1) = V_Q \perp D_Q(1).$$
(The notation $A = B \perp C$ means that A is the direct product of B and C and that B and C are orthogonal to one another.)

Now let $\pi : V \to V_Q^{\perp}$ denote the canonical projection. Suppose that $x \in B(1)$. Then $x \in \cap_{Q\subset V_{Q'}} N_{Q'}(1)$. Hence $\pi(x) \in D_Q(1)$ by line 4.2.15. Now line 4.2.14 implies that $\pi(B(1)) \subset B(1)$, and so scaling by r yields

4.2.16. $\pi(B(r)) \subset B(r)$ for every positive real number r.

Since $\cap_{Q\not\subset V_{Q'}} N_{Q'}(1)$ is a neighborhood of $D_Q(1)$ by line 4.2.13, lines 4.2.14 and 4.2.15 show that $B(1)$ is locally a product near $V_Q^{\perp}$. More precisely, there exists a positive real number ϵ such that $\{x \in V_Q | \|x\| \leq \epsilon\} \perp D_Q(1)$ is the closed ϵ-neighborhood of $V_Q^{\perp} \cap B(1)$ in $B(1)$. This proves the following.

4.2.17. There exists a positive real number R such that $\{x \in V_Q | \|x\| \leq \Lambda\} \perp D_Q(r)$ is the closed Λ-neighborhood of $V_Q^{\perp} \cap B(r)$ in $B(r)$ for every real number $r \geq R$.

The preparations have finally been made to prove line 4.2.9. The numbers $\lambda_1, \ldots, \lambda_{m-1}$ and R have been chosen to depend only on P and L. Let w, p, and a be given as in line 4.2.9. Let $Q = \{p\}$.

First suppose that the distance from w to $V_Q^{\perp}$ is at least Λ. Because $\langle p, w\rangle \leq 0$, the vectors w and p are on opposite sides of $V_Q^{\perp}$, so that $w + ap$ is closer to $V_Q^{\perp}$ than w. By line 4.2.16, $\pi(w) \in B(r)$. Hence the line segment from w to $\pi(w)$ is contained in $B(r)$ because $B(r)$ is convex. Thus $w + ap \in B(r)$ if the distance from w to $V_Q^{\perp}$ is at least Λ.

Now suppose that the distance from w to $V_Q^{\perp}$ is at most Λ. As before, either w and p are on opposite sides of $V_Q^{\perp}$ or $w \in V_Q^{\perp}$. In this case the local product structure of $B(r)$ described in line 4.2.17 easily shows that $w + ap \in B(r)$.

4.2.18. This completes the proof of line 4.2.9 and therefore the proofs of lines 4.2.7, 4.2.6, 4.2.1 and 4.1.2. $\diamond$ $\diamond$

Line 4.1.3 will be proven now. As after line 4.2.1 let V denote the linear hyperplane in $\mathbf{R}^n$ which is orthogonal to w_M. Let $\pi : \mathbf{R}^n \to V$ denote the canonical projection. Then $x = \pi(x) + \langle x, w_M\rangle w_M$ for every x in $\mathbf{R}^n$, and so

4.2.19.
$$\pi(x) = x - \frac{\langle x, aw_M\rangle}{a^2}(aw_M)$$
for every vector x in $\mathbf{R}^n$, where a is a positive real number such that aw_M is a vector of integers. Such an a exists by line 2.3.8. Since a is the length of a vector of integers, $a^2 \in \mathbf{Z}$.

According to line 4.1.2 the sequence of weight vectors $\frac{1}{\|w_1\|}w_1, \frac{1}{\|w_2\|}w_2, \frac{1}{\|w_3\|}w_3, \ldots$ in S^{n-1} converges to w_M. It is easy to see that given any weight vector w in S^{n-1} there exists a neighborhood U of w in S^{n-1} such that if p is a path vector whose length is minimal relative to some weight vector in U, then the length of p is minimal relative to w. Thus taking $w = w_M$ shows that there exists a positive integer N_0 such that if ν is an integer with $\nu \geq N_0$, then every w_ν-minimal path vector is a w_M-minimal path vector. In other words, if $\nu \geq N_0$, then Q_ν consists of w_M-minimal path vectors.

Now let p be a w_M-minimal path vector. Then
$$p = \pi(p) + \langle p, w_M\rangle w_M = \pi(p) + H_{w_M} w_M.$$
Thus
$$\begin{aligned}\langle p, w_\nu\rangle &= \langle \pi(p) + H_{w_M} w_M, \pi(w_\nu) + \langle w_\nu, w_M\rangle w_M\rangle \\ &= \langle \pi(p), \pi(w_\nu)\rangle + H_{w_M}\langle w_\nu, w_M\rangle.\end{aligned}$$

This shows that $\langle p, w_\nu\rangle$ is minimal if and only if $\langle \pi(p), \pi(w_\nu)\rangle$ is minimal.

The argument above shows for $\nu \geq N_0$ that $\pi(w_{\nu+1})$ is obtained from $\pi(w_\nu)$ as follows. Replace the sets P and $\pi(P)$ of all path vectors and their projections to V by the sets of all w_M-minimal path vectors and their projections to V. Let $\pi(Q_\nu)$ be the projection to V of the set Q_ν of w_M-minimal weight vectors p for which $\langle \pi(p), \pi(w_\nu)\rangle$ is minimal. Then $\pi(w_{\nu+1}) = \pi(w_\nu) + \pi(x_{Q_\nu})$, where, of course, $\pi(x_{Q_\nu})$ is a fixed vector in the cone spanned by $\pi(Q_\nu)$.

This construction of $\pi(w_{N_0})$, $\pi(w_{N_0+1})$, $\pi(w_{N_0+2})$, $\ldots$ is a special case of the construction immediately preceding line 4.2.6. Thus not only is the sequence $\pi(w_{N_0})$, $\pi(w_{N_0+1})$, $\pi(w_{N_0+2})$, $\ldots$ bounded as given by line 4.2.6, but Q_ν and hence $\pi(w_{\nu+1})$ are uniquely determined by $\pi(w_\nu)$ for $\nu \geq N_0$.

Since the coordinates of w_1, w_2, w_3, $\ldots$ are integers, line 4.2.19 shows that the coordinates of $\pi(w_1)$, $\pi(w_2)$, $\pi(w_3)$, $\ldots$ are rational numbers whose denominators are bounded. Combining the last statement with the previous paragraph shows that there exist positive integers N and μ such that $N \geq N_0$ and $\pi(w_{\mu+N}) = \pi(w_N)$. Because $\pi(w_N)$ uniquely determines $\pi(w_{N+1})$, $\pi(w_{\mu+N+1}) = \pi(w_{N+1})$. A straightforward induction argument now gives line 4.1.3.
$\diamondsuit$ 4.1.3 $\diamondsuit$

To prove line 4.1.4 let $w = w_{\mu+\nu} - w_\nu$. Then $w_{m\mu+\nu} = w_\nu + mw$ for every positive integer m. This implies that the sequence $\frac{1}{\|w_1\|}w_1$, $\frac{1}{\|w_2\|}w_2$, $\frac{1}{\|w_3\|}w_3$, $\ldots$ converges to $\frac{1}{\|w\|}w$. Thus line 4.1.2 implies that w is a scalar multiple of w_M.
$\diamondsuit$ 4.1.4 $\diamondsuit$

4.3. The efficiency of the minimal path algorithm. The minimal path algorithm depends upon finding a cycle in the path additions of the algorithm. We next construct tilings which are juxtapositions of tilings such that the cycle length of the juxtaposition is the least common multiple of the cycle lengths of the component tilings. For this construction we first consider quadrilaterals which are actually rectangles. Such a rectangle R is divided into two halves by a vertical line segment. We tile R so that the left half of R is a single tile and the right half of R is tiled arbitrarily. It is easy to see that when the minimal path algorithm for either skinny flows or fat flows is applied to such a tiling of a rectangle R that at step i the weight of the left half of R is i and the height of R is i. Now consider a side-to-side juxtaposition of several congruent rectangles tiled in this way. It is easy to see that the i-th step of the minimal path algorithm for such a juxtaposition is gotten by combining the i-th steps of the minimal path algorithm for the component tilings. It follows that the cycle length of the juxtaposition is the least common multiple of the cycle lengths of the component tilings. As a result, the minimal path algorithm tends to run very slowly for such juxtapositions unless component cycle lengths have small least common multiple. As a consequence, we will develop a hybrid algorithm in the next section which does not suffer from this defect. We shall give an example in Section 4.5.

4.4. Flow diagrams and a hybrid algorithm for the finite Riemann mapping problem. In this section we present a hybrid algorithm for computing optimal weight functions for fat flows of tilings of quadrilaterals. Our hybrid algorithm depends on using the main construction of the minimal path algorithm to form a guess as to the appropriate circuit diagram to associate with a tiled quadrilateral. We then solve the circuit problem and check whether the answer is indeed a solution to the finite Riemann mapping problem. If not, we iterate the process to make a better guess for the circuit. In general, the hybrid algorithm works many times faster than the pure minimal path algorithm and avoids the least common multiple problem of the minimal path algorithm. It is convenient to work not with the circuit of the appropriate Smith diagram but rather directly with the generalized circuit described in Section 3.2 on the Kirchhoff inequalities. The guessing process is designed to find the edges along which positive current will flow in the solution of the finite Riemann mapping problem as described in Section 3.2. We deal with a subgraph Γ' of the dual graph Γ of the tiling as in Section 3.2. We did not direct the edges of Γ, but we will direct the edges of Γ'. Here is the hybrid algorithm.

(1) Choose strictly increasing sequences u_1, u_2, u_3, $\ldots$ and v_1, v_2, v_3, $\ldots$ of positive integers such that the differences $u_j - v_j$ are positive and approach ∞ as j approaches ∞. The integer i will index iterations

of the hybrid algorithm. First set $i = 1$.

(2) Construct the sequence of weight functions $w_1, w_2, w_3, \ldots$ given by the minimal path algorithm and keep track of all the edges appearing in the minimal paths added during the steps which form w_{j+1} from w_j for $j = v_i, \ldots, u_i - 1$. In order to direct these edges, we assume that each of the minimal paths is oriented from the vertex t_0 corresponding to the bottom edge of the tiled quadrilateral to the vertex t_1 representing the top edge of the tiled quadrilateral. The orientations of these minimal paths unambiguously determine directions for all such edges. The union of these directed edges forms a subgraph Γ' of the graph Γ whose vertices correspond to some subset of the tiles of the tiling and whose edges correspond to some subset of ordered pairs of tiles intersecting along a common edge. This graph Γ' is necessarily connected, connecting t_0 to t_1.

(3) Associate a *current* variable $i(s,t)$ with each directed edge (s,t) of the graph Γ'. Require that this variable only be allowed to take on positive values: $i(s,t) > 0$ if (s,t) is a directed edge associated with the directed graph Γ'. Associate with each vertex tile s of Γ' a vertex equation as in Section 3.2. Choose a basis for the first homology of Γ' and associate with each of the basis loops a loop equation as in Section 3.2. Ignore for the time being the third condition raised by the Kirchhoff inequalities, the only one of the three conditions in which actual inequalities arise. The result is a system of equations in the current variables.

(4) Solve the system of equations derived in step (3). The expected solution space is one dimensional. If that is the case, and if there is a positive solution in this one-dimensional space, proceed to step 5. Otherwise return to step (2) with i incremented by 1.

(5) With the positive variables determined, it is possible to define the weight $w(s)$ of each vertex tile s of Γ': add up the value of the currents flowing into that tile. Assign the weight 0 to each tile not in the subgraph Γ'. Find the length of the w-minimal path from t_0 to t_1 in the full dual graph Γ. If this length agrees with the (equal) lengths of the paths used in forming the directed graph Γ', then w is the desired optimal weight function by Theorem 3.2.1. Otherwise return to step (2) with i incremented by 1.

A combination of line 4.1.4 and Theorem 3.2.1 guarantees that the algorithm finds an optimal weight function. Line 4.1.4 guarantees that step (2) will eventually supply the directed edges through which the solution to the finite Riemann mapping problem will send positive current. A comparison with the appropriate Smith diagram shows that we are one equation short of the number needed to determine the currents uniquely. What is missing is the voltage of the battery edge which changes the solution by a single scale factor. Finally, if there is a positive solution, then Theorem 3.2.1 tells how to check to see whether the solution is correct as in step (5).

4.5. An example of the effectiveness of the hybrid algorithm. We first tested the effectiveness of the hybrid algorithm on what we call a **corner**. As with the tiling in Section 3.1 we removed one tile from the upper right corner of a large square to form a small "top", with left side and bottom serving as "bottom". On a 9×9 large square the cyclic algorithm ran overnight without producing the optimal weight function. We made a hand implementation of the hybrid algorithm by running the cyclic algorithm a thousand cycles, observing the edges crossed by the minimal paths, and sending the resulting equations to a linear algebra package. The total computer time involved was something under four seconds. From the numbers alone it is easy to see that the minimal path algorithm must run through at least 6,476,565 cycles in order to find the optimal weight function, and with our actual implementation had to run at least 19,439,490 cycles. Here is the tiling with optimal weight function.

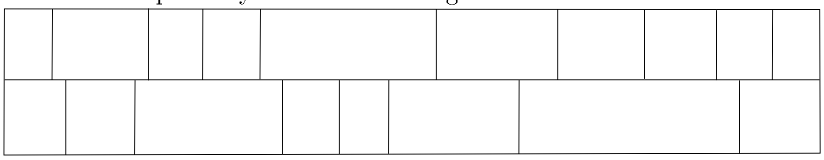
6476565	6476565	6476565	6479830	7775796	9719745	12959660	19439490	
4748175	4748175	4748175	5185170	6479830	7775796	9719745	12959660	19439490
3379050	3379050	3379050	4107375	5185170	6479830	7775796	9719745	12959660
2247750	2247750	2417850	3223800	4107375	5185170	6479830	7775796	9719745
1385100	1385100	1725300	2417850	3223800	4107375	5185170	6479830	7775796
765450	801900	1202850	1725300	2417850	3223800	4107375	5185170	6479830
328050	437400	801900	1202850	1725300	2417850	3379050	4748175	6476565
109350	218700	437400	801900	1385100	2247750	3379050	4748175	6476565
0	109350	328050	765450	1385100	2247750	3379050	4748175	6476565

5. Optimal weight functions for 2-layer valence 3 tilings of quadrilaterals.

Optimal weight functions for tiled quadrilaterals are in general very complicated. Even simple tilings supply tantalizingly difficult problems. We consider here a special case which can be used as a student puzzle.

Consider congruent rectangles R_i, $i = 0, 1$, and divide R_i into rectangular tiles by vertical line segments joining the top and bottom edges of R_i. Now stack R_1 on top of R_0 and assume that none of the vertical segments subdividing R_0 meets a vertical segment subdividing R_1. The result is a 2-**layer valence** 3 **tiling** T of the quadrilateral $R = R_0 \cup R_1$.

Example 5.0.1. Here is a sample 2-layer valence 3 tiling.



Our goal is to understand optimal weight functions for T relative to fat flows joining the top and bottom of R. We shall see that the tiling T has a natural and unique factorization into **prime** 2-layer valence 3 tilings, that the prime tilings can be completely understood in simple terms, that prime tilings can be multiplied in a simple way, and that the prime decomposition of a 2-layer valence 3 tiling can be found by efficient algorithms. We challenge the reader to carry out the analysis without hints.

We found that the minimal path algorithm was enough to get students started on the project. In just such a way, an undergraduate student at BYU by the name of Geo Meyer helped our research by discovering some of the first properties of the solution.

5.1. Basic definitions and properties.

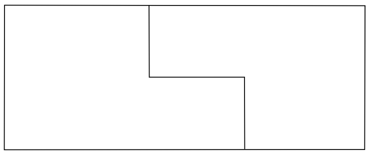
Combinatorially equivalent tilings are said to be **equivalent**. The equivalence class of a tiling T is denoted by $[T]$.

The **weight** of a tile in the tiling T will be taken here to be the value of the reduced integral optimal weight function of T.

The intersection of R_0 with R_1 is called the **midline** of R. The tiles in R_0 form the **lower layer** of T, and the tiles in R_1 form the **upper layer**. We refer as needed to the left side, right side, top and bottom of R as the **left side**, **right side**, **top**, and **bottom** of T.

If u is a union of tiles in a layer of T and v is a union of tiles in the other layer of T, then u **dominates** v if $u \cap m \supset v \cap m$, where m is the midline of R.

Let T be a 2-layer valence 3 tiling of a quadrilateral R. A **right joint** of T is a union $j = e_1 \cup e_2 \cup e_3$ of three edges e_1, e_2, e_3 of T (viewing T as giving a cellular decomposition of R) such that i) e_2 is contained in the interior of the midline of T, ii) e_1 joins the top of T to the left vertex of e_2, and iii) e_3 joins the bottom of T to the right vertex of e_2. The right joint j appears in R as follows.

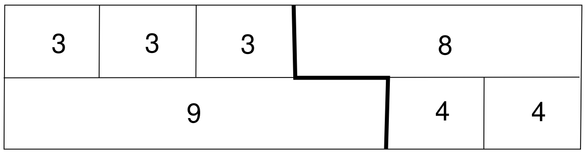


The right joint j separates R into two quadrilaterals Q_1 and Q_2, where j is the right side of Q_1 and the left side of Q_2. Similarly, j determines two subtilings T_1 and T_2 of T, where T_1 and T_2 are 2-layer valence 3 tilings of Q_1 and Q_2, respectively. The tiling T is called the **join** of T_1 and T_2. In general the expression "join of T_1 and T_2" means that T_1 is left of the joint and T_2 is right of the joint. If the optimal weight function for T induces on T_1 and T_2 scalar multiples of their optimal weight functions, then T is called the **(right) product** of T_1 and T_2. In this case j is called a **right break**.

There are analogous notions of **left joint, left join, (left) product** and **left break**.

If a tiling T is the product of tilings T_1 and T_2 as above, then T_1 and T_2 are called **factors** of T. It will be convenient to say that every tiling is a factor of itself. A **prime tiling** is a 2-layer valence 3 tiling of a quadrilateral which has no proper factors.

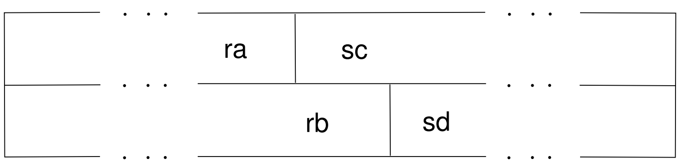
Example 5.1.1. Here is an example of a 2-layer valence 3 tiling of a quadrilateral which is the (right) product of two prime tilings. The right break between the factors is highlighted, and the weight of every tile appears in that tile.



The next proposition provides a simple criterion to determine when a join is a product.

Proposition 5.1.2. *Let T be a 2-layer valence 3 tiling of a quadrilateral which is the join of two subtilings T_1 and T_2. Using the reduced integral optimal weight function for T_1, let the rightmost upper, respectively lower, tile of T_1 have weight a, respectively b. Using the reduced integral optimal weight function for T_2, let the leftmost upper, respectively lower, tile of T_2 have weight c, respectively d. If the join is a right join, then it is a product if and only if $\frac{a}{b} \leq \frac{c}{d}$. If the join is a left join, then it is a product if and only if $\frac{a}{b} \geq \frac{c}{d}$.*

Proof. Suppose that T is the right product of T_1 and T_2. Furthermore suppose that the reduced integral optimal weight functions for T_1 and T_2 are scaled by factors r and s to give the reduced integral optimal weight function for T.



This situation leads to the following.

$$ra + rb = sc + sd \leq sc + rb$$

if and only if

$$\frac{s}{r} = \frac{a+b}{c+d} \text{ and } a + b \leq \frac{s}{r}c + b$$

if and only if

$$\frac{s}{r} = \frac{a+b}{c+d} \text{ and } a + b \leq \frac{a+b}{c+d}c + b$$

if and only if

$$\frac{s}{r} = \frac{a+b}{c+d} \text{ and } \frac{a}{b} \leq \frac{c}{d}.$$

Thus $\frac{a}{b} \leq \frac{c}{d}$.

Conversely, if $\frac{a}{b} \leq \frac{c}{d}$, and r, s are chosen so that $\frac{s}{r} = \frac{a+b}{c+d}$, then the calculations just completed easily show that minimal fat flows for T_1 and T_2 are minimal fat flows for T. Hence the weight function on T induced from the weight functions on T_1 and T_2 lies in the cone spanned by its minimal fat flows by line 2.3.2, and so it is an optimal weight function by line 2.3.6. Thus the right join of T_1 and T_2 is a product.

Analogous arguments prove the corresponding assertions for left joins.
$\diamondsuit$ Proposition 5.1.2 $\diamondsuit$

It is clear that the notions of left join, right join, left product, and right product extend to the set of equivalence classes of 2-layer valence 3 tilings of quadrilaterals.

5.2. Unique prime factorization theorem.

Following is a unique prime factorization theorem for 2-layer valence 3 tilings of quadrilaterals.

Theorem 5.2.1. *Every 2-layer valence 3 tiling of a quadrilateral can be uniquely expressed as a product of prime tilings.*

Proof. Let T be a 2-layer valence 3 tiling of a quadrilateral. Let w be the reduced integral optimal weight function of T. Let $F = (f_1, \dots, f_k)$ be a fundamental family of parallel fat flows whose sum is w in order from left to right. Let $s_1, \dots, s_m$ be the upper tiles of T in order from left to right. Let $t_1, \dots, t_n$ be the lower tiles of T in order from left to right. Let the weight of s_1 be ra, and let the weight of t_1 be rb, where a, b, r are positive integers with $GCD(a, b) = 1$.

Line 5.2.2 will be proved next.

5.2.2. If i is an integer with $1 \leq i < b$ and $i \leq m$, then $i + 1 \leq m$ and there exists an integer h with $1 \leq h < k$ such that f_h contains s_i, f_{h+1} contains s_{i+1}, and the lower tiles in f_h and f_{h+1} are equal.

It suffices to prove line 5.2.2 under the assumption that it is true for smaller values of i, using the fact that it is vacuously true for $i = 0$. Let $f_1, \dots, f_h$ be the flows in F which contain one of $s_1, \dots, s_i$, and let $t_1, \dots, t_j$ be the lower tiles which are dominated by $s_1 \cup \cdots \cup s_i$. Because line 5.2.2 is true for smaller values of i and every flow in F has length $ra + rb$, it is not difficult to see that $s_1, \dots, s_i$ all have weight ra and $t_1, \dots, t_j$ all have weight rb.

In this paragraph we prove that $s_1, \ldots, s_i$ are not the upper tiles of a factor of T. For this first note that $h = rai$. Because $i < b$ and $GCD(a, b) = 1$, it is impossible for rb to divide h. On the other hand, it is easy to see that $rbj \leq h \leq rb(j+1)$. Hence $rbj < h < rb(j+1)$. This proves that $s_1, \ldots, s_i$ are not the upper tiles of a factor of T.

Since $s_1, \ldots, s_i$ are not the upper tiles of a factor of T, it follows that t_{j+1} exists and f_h contains s_i and t_{j+1}. Moreover, not only do f_{h+1} and s_{i+1} exist, but f_{h+1} contains s_{i+1} and t_{j+1}. This proves line 5.2.2.
$\diamondsuit$ 5.2.2 $\diamondsuit$

Since line 5.2.2 is proved, $b \leq m$. As was observed above, line 5.2.2 implies that $s_1, \ldots, s_b$ all have weight ra. Now let $f_1, \ldots, f_h$ be the flows in F which contain one of $s_1, \ldots, s_b$. It follows that $h = rab$. From this it is easy to see that $t_1, \ldots, t_a$ exist, they all have weight rb, and $f_1, \ldots, f_h$ are exactly the flows in F which contain $t_1, \ldots, t_a$. Let P be the subtiling of T which consists of $s_1, \ldots, s_b$, $t_1, \ldots, t_a$. Then P is a factor of T. It is proved in the next-to-last paragraph that P is prime. The prime tiling P is uniquely determined by T because the optimal weight function is unique.

Here is what the argument above proves. Given a 2-layer valence 3 tiling of a quadrilateral there exists one and only one prime tiling which is a left factor of T. An obvious induction argument completes the proof of Theorem 5.2.1.
$\diamondsuit$ Theorem 5.2.1 $\diamondsuit$

5.3. Prime tilings.

The proof of Theorem 5.2.1 yields information about prime tilings. In particular, there is a map from the set of prime tilings to the set of positive rational numbers. The prime tiling P maps to the rational number $\frac{a}{b}$, where P has b upper tiles each with weight a and a lower tiles each with weight b. The number $\frac{a}{b}$ will be called the **weight ratio** of the prime tiling P. In general a fraction $\frac{a}{b}$ will be called the **reduced weight ratio** of a prime tiling P if a, b are relatively prime positive integers and $\frac{a}{b}$ is the weight ratio of P.

It is clear that prime tilings are equivalent if and only if they have equal weight ratios. Thus the weight ratio map induces an injective map from the set of equivalence classes of prime tilings to the set of positive rational numbers.

Prime tilings with arbitrarily prescribed weight ratios will be constructed in the following paragraphs. In particular, this construction shows that the weight ratio map from the set of equivalence classes of prime tilings to the set of positive rational numbers is bijective.

Let a and b be positive integers with $GCD(a, b) = 1$. A prime tiling with weight ratio $\frac{a}{b}$ will now be constructed assuming that $a \geq b$. Let $x_1, \ldots, x_b$ be integers such that $x_i \equiv ia \mod b$ and $0 \leq x_i < b$ for $i = 1, \ldots, b$. Let $a = bq + r$, where q, r are integers such that $0 \leq r < b$. In particular, $r = x_1$. Construct a 2-layer valence 3 tiling P of a rectangle R as follows. The midline of the tiling is the line segment joining the midpoints of the sides of R. Let the upper layer of P consist of b rectangular tiles $s_1, \ldots, s_b$. Construct the lower layer of P to consist of rectangular tiles none of which dominates an upper tile so that for every integer i with $1 \leq i \leq b$ tile s_i dominates q tiles if $x_i \leq r$ and tile s_i dominates $q-1$ tiles if $x_i > r$. This describes the tiling P.

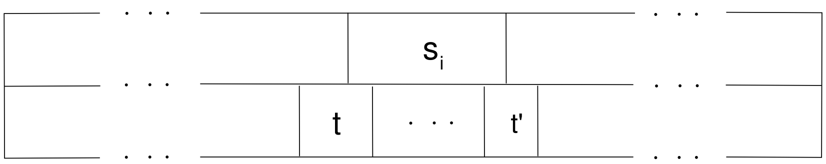
To see that P is a prime tiling with weight ratio $\frac{a}{b}$, it suffices to prove that the weight of every upper tile of P is a and the weight of every lower tile in P is b. For this it suffices to construct a family F of fat flows for P such that i) every flow in F consists of one upper tile and one lower tile, ii) every lower tile is contained in b of the flows in F and iii) every upper tile is contained in a of the flows in F. To this end define integers $y_1, \ldots, y_b$ with $0 \leq y_i < b$ such that the following holds.

5.3.1. $x_i + y_i \equiv a \mod b$ for $i = 1, \ldots, b$.

Since $x_{i+1} \equiv x_i + a \mod b$ and $x_{i+1} + y_{i+1} \equiv a \mod b$, $x_i + y_{i+1} \equiv 0 \mod b$ for $i = 1, \ldots, b-1$, which gives the following.

5.3.2. $x_i + y_{i+1} = b$ for $i = 1, \ldots, b-1$.

The family F of fat flows will be constructed in this paragraph. Suppose given an upper tile s_i. If $i > 1$, then let t be the leftmost lower tile which meets s_i. If $i < b$, then let t' be the rightmost lower tile which meets s_i.

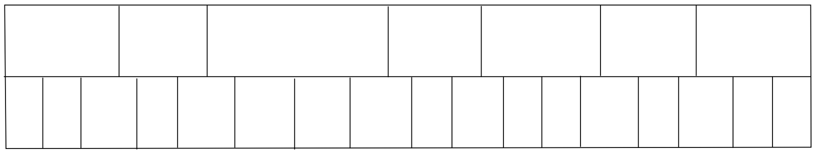


If $i > 1$, then let F include y_i paths consisting of s_i and t. If $i < b$, then let F include x_i paths consisting of s_i and t'. The family F will also include bq or $b(q-1)$ paths consisting of s_i and a tile below s_i depending upon whether $x_i \leq r$ or $x_i > r$.

It will now be verified that F satisfies the necessary conditions. It is clear that condition i) is satisfied. Condition ii) follows easily from line 5.3.2. To verify condition iii), let s_i be an upper tile. If $x_i \leq r$, then line 5.3.1 implies that $x_i + y_i = r$. Hence in this case $a = bq + x_i + y_i$, and so s_i is contained in a flows in F. If $x_i > r$, then 5.3.1 implies that $x_i + y_i = b + r$. Hence in this case $a = b(q-1) + x_i + y_i$, and again s_i is contained in a flows of F.

This completes the construction of a prime tiling with weight ratio $\frac{a}{b}$ if $\frac{a}{b} \geq 1$. The construction for rational numbers less than 1 is analogous. (Alternatively, the horizontal reflection of a prime tiling with weight ratio $\frac{a}{b}$ is a prime tiling with weight ratio $\frac{b}{a}$. See below for reflections.)

Example 5.3.3. A prime tiling with weight ratio $\frac{17}{7}$ can be constructed as follows. First note that $17 = 7 \cdot 2 + 3$. Next compute the least nonnegative residues of multiples of 3 modulo 7: $3, 6, 2, 5, 1, 4, 0$. Because $6, 5, 4 > 3$, the second, fourth and sixth upper tiles dominate one tile. The other upper tiles dominate two tiles. Thus the following is a prime tiling with weight ratio $\frac{17}{7}$.



There is a **vertical reflection** (that is, reflection in a vertical line) on the set of equivalence classes of 2-layer valence 3 tilings of quadrilaterals which is defined as follows. Given an equivalence class of 2-layer valence 3 tilings of quadrilaterals, it is possible to choose a representative T which is a tiling of a rectangle. The vertical reflection maps $[T]$ to the equivalence class represented by the tiling gotten by reflecting T through a vertical line parallel to the sides of T. A tiling T will be called **symmetric** if $[T]$ is fixed by vertical reflection. There is a **horizontal reflection** (that is, reflection in a horizontal line) on the set of equivalence classes of 2-layer valence 3 tilings of quadrilaterals which is analogous to vertical reflection.

Proposition 5.3.4. *Prime tilings are symmetric (under reflection in a vertical line).*

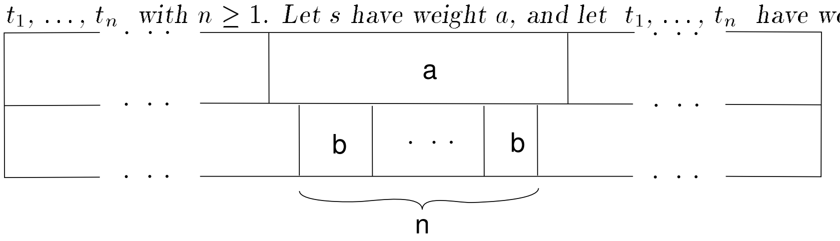
Proof. Suppose that P is a prime tiling with reduced weight ratio $\frac{a}{b}$. If Q is a representative of the image of $[P]$ under the vertical reflection, then Q has the property that each of its upper tiles has weight a and each of its lower tiles has weight b. Thus every prime factor of Q has weight ratio $\frac{a}{b}$. This easily proves Proposition 5.3.4.
♢ Proposition 5.3.4 ♢

5.4. Preparations for factorization algorithms.

Given a 2-layer valence 3 tiling T of a quadrilateral, one would like to be able to compute its optimal weight function. The unique prime factorization theorem provides a way to do this: computing the optimal weight function of a prime tiling merely involves counting tiles, and the optimal weight function of T is gotten by scaling the optimal weight functions of its prime factors so that their heights agree. Thus the problem of computing the optimal weight function of T reduces to computing the prime factorization of T. The results of this section will be used to factor 2-layer valence 3 tilings of quadrilaterals.

It is clear for each of the following results in this section that analogous results hold for representatives of the images of the equivalence class of the given tiling under the vertical and horizontal reflections. When the following results are cited, it will be left to the reader to apply reflections if necessary.

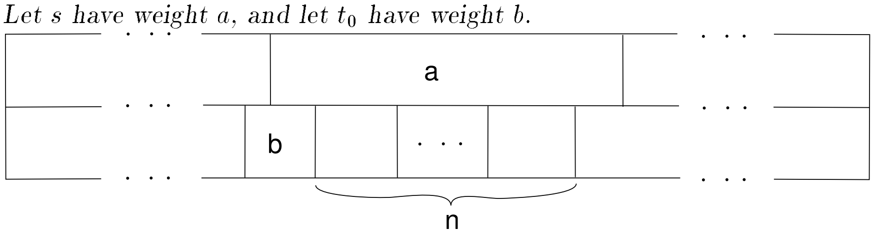
Lemma 5.4.1. *Suppose given an upper tile s in a 2-layer valence 3 tiling T of a quadrilateral which dominates n tiles $t_1, \ldots, t_n$ with $n \geq 1$. Let s have weight a, and let $t_1, \ldots, t_n$ have weight b.*



Then $n \leq \frac{a}{b} \leq m$, where m is the number of lower tiles meeting s. Since $m \leq n+2$, $n \leq \frac{a}{b} \leq n+2$. Moreover, $\frac{a}{b} = n$ if and only if s and $t_1, \ldots, t_n$ form a factor of T, and $\frac{a}{b} = n+2$ if and only if there is a lower tile t_0 immediately left of $t_1, \ldots, t_n$ and a lower tile t_{n+1} immediately right of $t_1, \ldots, t_n$ such that s and $t_0, \ldots, t_{n+1}$ form a factor of T.

Proof. Let w be the reduced integral optimal weight function of T. Let F be a fundamental family of parallel fat flows for T whose sum is w. It is clear that every flow in F which contains one of $t_1, \ldots, t_n$ also contains s, and so $nb \leq a$. Thus $n \leq \frac{a}{b}$. Moreover, equality holds if and only if the flows in F which contain one of $t_1, \ldots, t_n$ are exactly the flows which contain s. This occurs if and only if s and $t_1, \ldots, t_n$ form a factor of T. If some flow in F contains s but none of $t_1, \ldots, t_n$, then either t_0 or t_{n+1} exists and its weight is b. Now it is easy to see that $a \leq mb$. Thus $\frac{a}{b} \leq m$. The final assertion of the lemma is now easy to prove.
$\diamondsuit$ Lemma 5.4.1 $\diamondsuit$

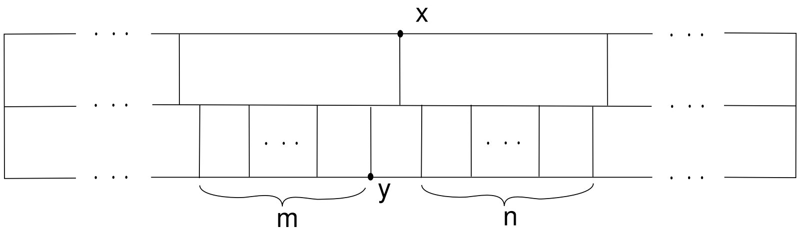
Lemma 5.4.2. *Suppose given an upper tile s in a 2-layer valence 3 tiling of a quadrilateral which dominates n tiles $t_1, \ldots, t_n$ with $n \geq 0$. Suppose that there exists a lower tile $t_0 \notin \{t_1, \ldots, t_n\}$ which meets the lower left vertex of s. Let s have weight a, and let t_0 have weight b.*



Then $\frac{a}{b} \leq m$, where m is the number of lower tiles meeting s. Since $m \leq n+2$, $\frac{a}{b} \leq n+2$. Moreover, $\frac{a}{b} = n+2$ if and only if there is a lower tile t_{n+1} immediately right of $t_1, \ldots, t_n$ such that s and $t_0, \ldots, t_{n+1}$ form a factor of T.

Proof. It is clear that if a minimal fat flow contains s, then the lower tile in it has weight at most b. As in the proof of Lemma 5.4.1, it easily follows that $a \leq mb$. Thus $\frac{a}{b} \leq m$. The rest of the lemma can be proved as in Lemma 5.4.1.
$\diamondsuit$ Lemma 5.4.2 $\diamondsuit$

Lemma 5.4.3. *Suppose given an upper tile in a 2-layer valence 3 tiling of a quadrilateral which dominates $m \geq 0$ tiles. Suppose that immediately to the right of this upper tile is another upper tile which dominates $n \geq m$ tiles. Define points x, y as in the diagram below.*



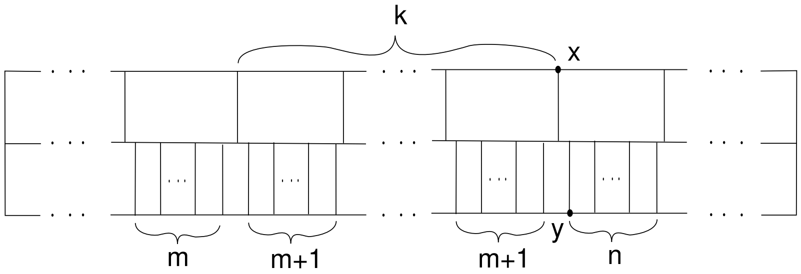
If there is a break from x to y, then the prime tilings on either side of this break are equivalent.

Proof. Suppose that there is a break from x to y. Let the reduced weight ratio of the prime tiling left, respectively right, of this break be $\frac{a}{b}$, respectively $\frac{c}{d}$. Then Lemmas 5.4.1 and 5.4.2 applied to these prime tilings show that

$$\frac{a}{b} \leq m+1 \leq n+1 \leq \frac{c}{d}.$$

Since Proposition 5.1.2 states that $\frac{a}{b} \geq \frac{c}{d}$, it follows that the two prime tilings have equal weight ratios, and so the two prime tilings are equivalent.
♢ Lemma 5.4.3 ♢

Proposition 5.4.4. *Suppose given an upper tile in a 2-layer valence 3 tiling of a quadrilateral which dominates $m \geq 0$ tiles. Suppose that immediately to the right of this upper tile are $k \geq 0$ upper tiles each of which dominates $m+1$ tiles. Suppose that immediately to the right of these upper tiles is an upper tile which dominates $n \geq m+2$ tiles. Define points x, y as in the diagram below.*



Then the joint from x to y is a break.

Proof. Let the reduced weight ratio of the prime tiling P containing the leftmost of the $k+2$ upper tiles under consideration be $\frac{a}{b}$. Similarly, let the reduced weight ratio of the prime tiling Q containing the rightmost of the $k+2$ upper tiles under consideration be $\frac{c}{d}$. According to Lemmas 5.4.1 and 5.4.2

$$\frac{a}{b} \leq m+2 \leq n \leq \frac{c}{d}.$$

If $\frac{c}{d} = n$, then Lemma 5.4.1 shows that there is a break joining x and y as desired. Hence it may be assumed that $\frac{c}{d} > n$. Thus $P \neq Q$, and so there is a break between P and Q.

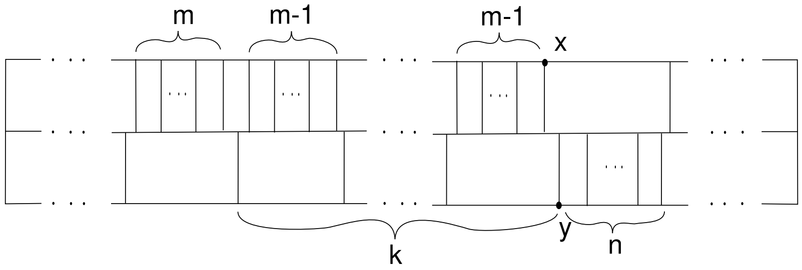
Suppose that the joint from x to y is not a break.

In this paragraph it will be shown by contradiction that the break at the left side of Q is not a right break. Suppose that it is a right break. Then Lemma 5.4.1 shows that $\frac{c}{d} \leq m+2 \leq n$. This contradicts the assumption that $\frac{c}{d} > n$.

Thus the break at the left side of Q is a left break. Now Lemma 5.4.3 shows that Q is equivalent to the prime tiling to its left. This is easily seen to be impossible using Proposition 5.3.4, which states that Q is symmetric.
♢ Proposition 5.4.4 ♢

The next proposition can be viewed as an extension of Proposition 5.4.4. It might be said that Proposition 5.4.5 extends Proposition 5.4.4 to the case in which m is negative and n is positive.

Proposition 5.4.5. *Suppose given a lower tile in a 2-layer valence 3 tiling of a quadrilateral which dominates m tiles with $m \geq 1$. Suppose that immediately to the right of this lower tile are $k \geq 0$ lower tiles each of which dominates $m-1$ tiles. Suppose that immediately to the right of these upper tiles is an upper tile which dominates n tiles with $n \geq 1$. Define points x, y as in the diagram below.*



Then the joint from x to y is a break.

Proof. The proof of Proposition 5.4.4 easily extends to prove Proposition 5.4.5.
♢ Proposition 5.4.5 ♢

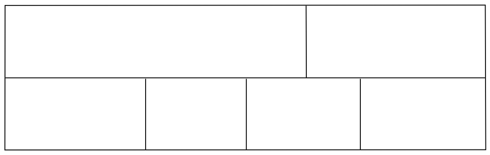
5.5. The left factor algorithm.

An algorithm, called the **left factor algorithm**, for computing prime factorizations of 2-layer valence 3 tilings of quadrilaterals will be described in this section. The left factor algorithm consists of three phases. The first phase is described immediately below.

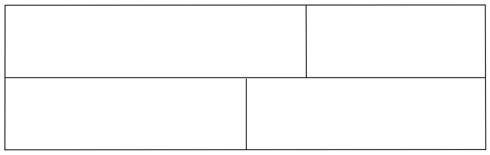
Let T be a 2-layer valence 3 tiling of a quadrilateral. Using Propositions 5.4.4 and 5.4.5, the first phase of the left factor algorithm expresses T as a product of factors which have a very special form as follows. Suppose that T contains two tiles s, respectively t, which dominate m, respectively n, tiles. Suppose that $n \geq m+2$ if s, t are in the same layer of T, and suppose that $m > 0$, $n > 0$ if s, t are in different layers of T. Further suppose that s, t are closest such tiles. The following two sentences are easy to see under these assumptions. If s, t are in the same layer and there are tiles between s and t in this layer, then $n = m + 2$ and every tile between s and t in this layer dominates $m+1$ tiles. If s, t are in different layers and there are tiles "between" s and t, then $m = n = 1$ and every tile "between" s and t dominates no tile. Thus under these assumptions either Proposition 5.4.4 or Proposition 5.4.5 provides an easily computable factorization of T. The first phase of the left factor algorithm consists of repeatedly applying Propositions 5.4.4 and 5.4.5 in this manner to obtain a factorization of T for which every factor S has the following properties: i) either no lower tile of S dominates a tile or no upper tile of S dominates a tile; ii) if no lower tile of S dominates a tile, then there exists an integer k such that every upper tile dominates either k or $k+1$ tiles; iii) if no upper tile of S dominates a tile, then there exists an integer k such that every lower tile dominates either k or $k+1$ tiles. Thus (using the horizontal reflection) the first phase of the left factor algorithm easily reduces computing the prime factorization of T to computing prime factorizations of tilings S such that no lower tile of S dominates a tile and there exists an integer k for which every upper tile of S dominates either k or $k+1$ tiles.

A function will be defined in this paragraph which will be used by the second phase of the left factor algorithm to simplify these latter tilings S. Let $\mathcal{T}$ denote the set of equivalence classes of 2-layer valence 3 tilings of quadrilaterals. Let $\mathcal{T}^+$ denote the set of equivalence classes $[T]$ in $\mathcal{T}$ such that every upper tile in T dominates at least one tile but T is not a prime tiling with weight ratio 1, that is, T does not consist of one upper tile and one lower tile. Define the **deletion map** $\delta : \mathcal{T}^+ \to \mathcal{T}$ so that for every equivalence class $[T]$ in $\mathcal{T}^+$ the equivalence class $\delta([T])$ is represented by the tiling gotten from T by deleting one tile below every upper tile.

Example 5.5.1. The equivalence class of the tiling



is in $\mathcal{T}^+$, and its image under the deletion map δ is represented by the following.



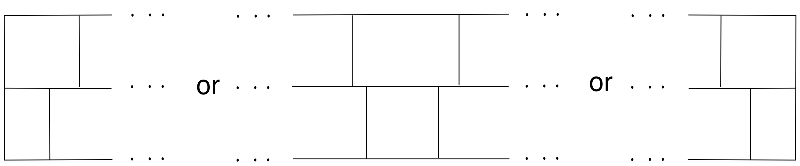
The deletion map is introduced because of the following proposition.

Proposition 5.5.2. *If P is a prime tiling with weight ratio $\frac{a}{b}$ such that $[P]$ lies in $\mathcal{T}^+$, then $\delta([P])$ is the equivalence class of prime tilings with weight ratio $\frac{a}{b}-1$. Furthermore, given an equivalence class $[T]$ in $\mathcal{T}^+$ such that T is the product of prime tilings $P_1, \ldots, P_n$ in order from left to right, then $[P_1], \ldots, [P_n]$ lie in $\mathcal{T}^+$ and $\delta([T])$ is the product of $\delta([P_1]), \ldots, \delta([P_n])$ in order from left to right.*

Proof. Let P be a prime tiling with reduced weight ratio $\frac{a}{b}$ such that $[P]$ lies in $\mathcal{T}^+$. It is clear that $a \geq b$. Moreover, $a > b$ because the equivalence class of prime tilings with weight ratio 1 is not in $\mathcal{T}^+$. Let $a = bq+r$, where q, r are integers with $0 \leq r < b$. Suppose that $q > 1$. Then since $a - b = b(q-1) + r$ with $q - 1 \geq 1$, the construction of prime tilings in Section 5.3 shows that $\delta([P])$ is represented by a prime tiling with weight ratio $\frac{a-b}{b} = \frac{a}{b} - 1$. Now suppose that $q = 1$. In this case because every upper tile of P dominates at least one tile, the construction of prime tilings shows that every upper tile of P dominates exactly one tile and that $r = b - 1$. It easily follows that if $q = 1$, then $\delta([P])$ is represented by a prime tiling with weight ratio $\frac{b-1}{b} = \frac{r}{b} = \frac{b+r}{b} - 1 = \frac{a}{b} - 1$. This proves the first assertion of Proposition 5.5.2.

In this paragraph it will be proved that if P is a prime factor of a tiling T such that $[T]$ lies in $\mathcal{T}^+$, then $[P]$ is also in $\mathcal{T}^+$. It is easy to see that every upper tile in P dominates at least one tile. Thus what must be proved is that the weight ratio of P is not 1. Suppose that the weight ratio of P is 1. Then because every upper tile in T dominates at least one tile, P occurs in T in one of the following three forms.

5.5.3.



a) b) c)

If P occurs in T as in a) of Figure 5.5.3, then Lemma 5.4.3 shows that the prime tiling Q to the right of P is equivalent to P. But Q does not occur in T in one of the three forms given in Figure 5.5.3. Thus P does not occur in T as in a) of Figure 5.5.3. Similar arguments conclude the proof that the weight ratio of P is not 1, and so $[P]$ lies in $\mathcal{T}^+$.

It is now easy to complete the proof of Proposition 5.5.2 by using Proposition 5.1.2 and the fact that δ reduces weight ratios of equivalence classes of prime tilings by 1.
$\diamondsuit$ Proposition 5.5.2 $\diamondsuit$

The description of the left factor algorithm now turns to the second phase. Let T be a 2-layer valence 3 tiling of a quadrilateral for which there exists a positive integer k such that every upper tile of T dominates either k or $k+1$ tiles. Assuming that T is not the prime tiling with weight ratio 1, Proposition 5.5.2 shows that computing the prime factorization of T is equivalent to computing the prime factorization of a tiling S which represents $\delta([T])$. Note that Example 5.5.1 shows for $k = 1$ that a lower tile of S might dominate an upper tile. If this occurs, then apply the first phase of the left factor algorithm to S. The second phase of the left factor algorithm replaces T by a tiling which represents the image of $[T]$ under the largest possible iterate of δ and applies the first phase of the algorithm to this tiling.

Thus the factorization of a 2-layer valence 3 tiling T of a quadrilateral easily reduces to the case in which no lower tile of T dominates a tile and every upper tile of T dominates at most one tile. It is interesting that this seems to be the most difficult case. The next lemma gives information about the location of breaks for such tilings T.

Lemma 5.5.4. *Let T be a 2-layer valence 3 tiling of a quadrilateral such that no lower tile of T dominates a tile and every upper tile of T dominates at most one tile. Let b be a left break of T, and let s, respectively t, be the upper tile just left, respectively right, of b. Then s dominates one tile and t dominates no tile. An analogous result holds for right breaks.*

Proof. The proof will proceed by induction on the number k of prime factors of T. If $k = 1$, then the lemma is vacuously true, so suppose that $k > 1$ and that the lemma is true for smaller values of k.

Suppose that s dominates m tiles and t dominates n tiles with $m, n \in \{0, 1\}$. The proof of the induction step will proceed by contradiction. Suppose that $m \leq n$. Let P be the prime tiling immediately left of b, and let Q be the prime tiling immediately right of B. Lemma 5.4.3 shows that $[P] = [Q]$. Let P, Q have weight ratio $\frac{a}{b}$. Lemmas 5.4.1 and 5.4.2 show that

$$\frac{a}{b} \leq m + 1 \leq n + 1 \leq \frac{a}{b}.$$

Thus $\frac{a}{b} = m + 1 = n + 1$. This implies that P consists of s, the m tiles which s dominates and one more lower tile which meets the lower left vertex of P. Similarly, Q consists of t, the n tiles which t dominates and one more lower tile which meets s and t.

Now suppose that $m = n = 0$. Then the right side of Q is not the right side of T, for otherwise the lower tile of Q dominates the upper tile of Q, contrary to assumption. Thus the right side of Q is a left break of the tiling S which consists of the tiles between b and the right side of T. The induction hypothesis applies to S and shows that Q dominates one tile. This contradiction proves the induction step in this case.

It remains to consider the case in which $m = n = 1$. This case can be proved just as the case in which $m = n = 0$ by working with the factor of T left of b.
$\diamondsuit$ Lemma 5.5.4 $\diamondsuit$

Now let T be a 2-layer valence 3 tiling of a quadrilateral such that no lower tile of T dominates a tile and every upper tile of T dominates at most one tile. The third phase of the left factor algorithm, which computes the prime factorization of T will now be given. The left factor algorithm is named for this third and most difficult phase. Call a joint of T a **potential break** if it is not ruled out by Lemma 5.5.4. If T has no potential breaks, then T is prime. Otherwise, let $b_1, \ldots, b_n$ be the potential breaks of T.

Now the left factor algorithm searches for the leftmost prime factor of T as follows. Let T_1 be the tiling which consists of the tiles of T between the left side of T and b_1. The next step of the algorithm is to determine whether or not T_1 is prime. An easy, although possibly inconclusive, test to apply is to test T_1 for symmetry. If T_1 is not symmetric, then T_1 is not prime by Proposition 5.3.4. If T_1 is symmetric, then suppose that T_1 has a lower tiles and b upper tiles. If T_1 is prime, then its weight ratio is $\frac{a}{b}$. It is only slightly more difficult than testing for symmetry to check whether or not T_1 is a prime tiling with weight ratio $\frac{a}{b}$, using the construction of prime tilings given in Section 5.3. If T_1 is not prime, then redefine T_1 to be the tiling which consists of the tiles between the left side of T and b_2 (the right side of T if $n = 1$). Continue in this manner until a prime tiling T_1 is found. This tiling T_1 is a potential leftmost prime factor of T.

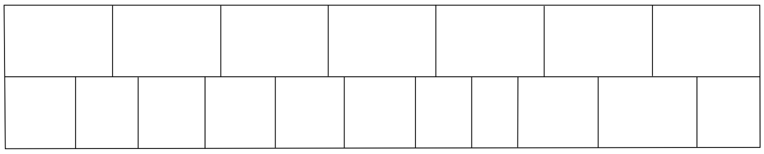
If $T_1 = T$, then there is nothing more to do. If $T_1 \neq T$, then find as above a potential leftmost prime factor T_2 of the tiling S which consists of the tiles of T between the right side of T_1 and the right side of T.

Use Proposition 5.1.2 to test whether or not the join of T_1 and T_2 is a product. If the join of T_1 and T_2 is not a product, then continue the search for a suitable T_2, that is, a prime tiling T_2 such that the join of

T_1 and T_2 is a product and T_2 consists of the tiles between the left side of S and a potential break of S or the right side of S. If no suitable T_2 is found, then T_1 is not the leftmost prime factor of T, so continue the search for a leftmost prime factor of T.

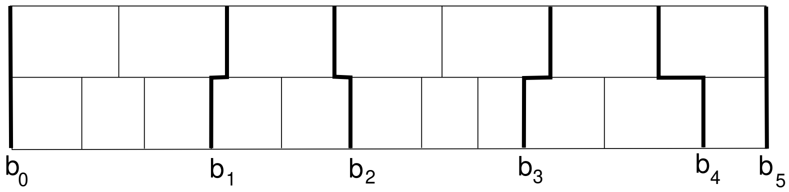
Suppose that a suitable tiling T_2 is found. If T is the product of T_1 and T_2, then there is nothing more to do. Otherwise, search for a suitable T_3. It is clear that this process continues until the prime factorization of T is found. This completes the description of the left factor algorithm.

Example 5.5.5. Following is an application of the third phase of the left factor algorithm. Let T be the tiling shown below.



No lower tile of T dominates a tile and every upper tile of T dominates at most one tile. Thus the third phase of the left factor algorithm applies to T.

The potential breaks of T are labeled b_1, b_2, b_3, b_4 below, and the sides of T are labeled b_0, b_5.

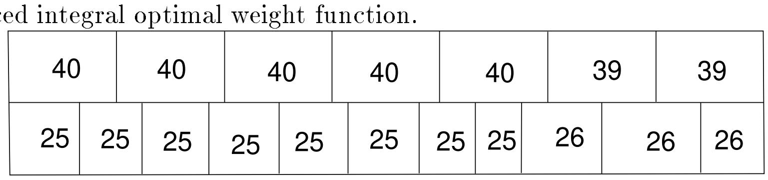


For integers i, j with $0 \leq i < j \leq 5$, let T_{ij} be the tiling which consists of the tiles in T between b_i and b_j. The table below describes how the left factor algorithm computes the prime factorization of T. Every box is either empty or it contains some T_{ij}. If T_{ij} is not symmetric, then "NS" appears. If T_{ij} is prime, then "P" appears with the weight ratio of T_{ij}. If T_{ij} is prime, but the join of the previous potential prime factor of T and T_{ij} is not a product, then "NP" appears. If T_{ij} is prime and the join of the previous potential prime factor of T and T_{ij} is a product, then move right in the table. Otherwise, move down and left.

$T_{0,1}$ P,3/2	$T_{1,2}$ P,2,NP	
	$T_{1,3}$ P,5/3,NP	
	$T_{1,4}$ P,7/4,NP	
	$T_{1,5}$ NS	
$T_{0,2}$ P,5/3	$T_{2,3}$ P,3/2,NP	
	$T_{2,4}$ P,5/3	$T_{4,5}$ P,1,NP
	$T_{2,5}$ NS	
$T_{0,3}$ P,8/5	$T_{3,4}$ P,2,NP	
	$T_{3,5}$ P,3/2	

Thus T is the product of $T_{0,3}$ and $T_{3,5}$, prime tilings with weight ratios $\frac{8}{5}$ and $\frac{3}{2}$, respectively.

Here is the reduced integral optimal weight function.

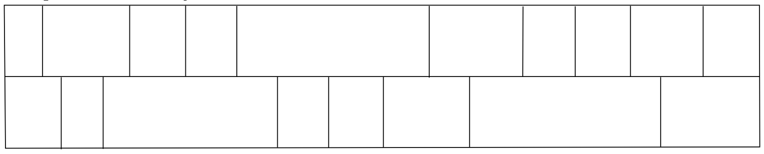


5.6. The joint ratio algorithm.

In this section another algorithm, called the joint ratio algorithm, for computing prime factorizations of 2-layer valence 3 tilings of quadrilaterals will be described. Joint ratios are defined next.

Suppose given a 2-layer valence 3 tiling T of a quadrilateral. Define the **right joint ratio** for a given right joint of T to be the number of lower tiles in T left of the joint divided by the number of upper tiles in T left of the joint. Define a right joint ratio for the right side of T as well; this is the number of lower tiles in T divided by the number of upper tiles in T. Define **left joint ratios** in the same way using left joints instead of right joints. The left joint ratio for the right side of T equals the right joint ratio for the right side of T.

Example 5.6.1. The right and left joint ratios for



are

$$\frac{1}{1} \quad \frac{3}{4} \quad \frac{6}{5} \quad \frac{7}{8} \quad \frac{8}{10}$$

and

$$\frac{2}{2} \quad \frac{5}{5} \quad \frac{6}{6} \quad \frac{7}{9} \quad \frac{8}{10}.$$

The next theorem is the basis for the joint ratio algorithm.

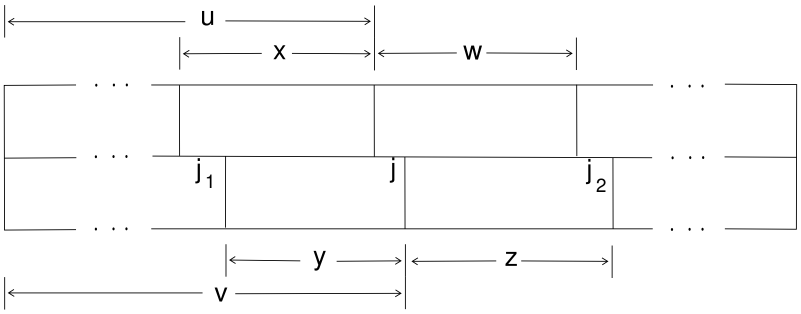
Theorem 5.6.2. *Suppose given a 2-layer valence 3 tiling T of a quadrilateral.*

1) The minimal right joint ratios of T occur at breaks or the right side of T.

2) The maximal left joint ratios of T occur at breaks or the right side of T.

3) If the joint ratio of the right side of T is both the unique minimal right joint ratio and the unique maximal left joint ratio of T, then T is a prime tiling.

Proof. The proof will begin by establishing some notation. Suppose that T has a right joint j. Let j_1 be a right joint of T left of j. If there is no right joint of T left of j, then let j_1 be the left side of T. Let j_2 be a right joint of T right of j. If there is no right joint of T right of j, then let j_2 be the right side of T. Let u, respectively v, be the number of upper, respectively lower, tiles left of j. Let x, respectively y, be the number of upper, respectively lower, tiles between j and j_1. Let w, respectively z, be the number of upper, respectively lower, tiles between j and j_2.



Thus the right joint ratios for j, j_1, j_2 are $\frac{v}{u}$, $\frac{v-y}{u-x}$, $\frac{v+z}{u+w}$, except that j_1 has no right joint ratio if it is the left side of T. This leads to line 5.6.3, which is easily verified.

5.6.3.

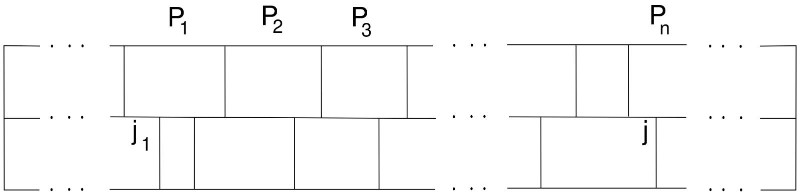
$$\frac{v}{u} \le \frac{v-y}{u-x} \iff uv - vx \le uv - uy \iff vx \ge uy \iff \frac{v}{u} \ge \frac{y}{x}$$

$$\frac{v}{u} \le \frac{v+z}{u+w} \iff uv + vw \le uv + uz \iff vw \le uz \iff \frac{v}{u} \le \frac{z}{w}$$

Now consider statement 1) of Theorem 5.6.2. Suppose that the right joint ratio of the above joint j is minimal and that j is not a break. This will lead to a contradiction. Choose j_1 and j_2 to be right breaks or sides of T so that there are no right breaks between j_1 and j_2. Because the right joint ratio for j is less than or equal to the right joint ratios for j_1 (if it exists) and j_2, $\frac{v}{u} \leq \frac{v-y}{u-x}$ (if meaningful) and $\frac{v}{u} \leq \frac{v+z}{u+w}$. From line 5.6.3 it follows that

5.6.4.
$$\frac{v}{u} \geq \frac{y}{x} \text{ and } \frac{v}{u} \leq \frac{z}{w}.$$

The ratio $\frac{y}{x}$ will be investigated in this paragraph. Let $P_1, \ldots, P_n$ be the consecutive prime factors of T for which j_1 is the left side of P_1 and j is contained in P_n.



Let the reduced weight ratios of $P_1, \ldots, P_n$ be $\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}$. Because the breaks between the P_i's are left breaks, Proposition 5.1.2 shows that
$$\frac{a_1}{b_1} \geq \cdots \geq \frac{a_n}{b_n}.$$
For $i = 1, \ldots, n-1$ let $x_i = b_i$ and $y_i = a_i$. Choose x_n and y_n so that $\sum_{i=1}^n x_i = x$ and $\sum_{i=1}^n y_i = y$. Then $\frac{y_i}{x_i} = \frac{a_i}{b_i} \geq \frac{a_n}{b_n}$, and so

5.6.5.
$$y_i b_n \geq x_i a_n \quad \text{for} \quad i = 1, \ldots, n-1.$$
Furthermore, because P_n is a prime tiling containing the right joint j having x_n tiles with weights a_n in its upper layer left of j and y_n tiles with weights b_n in its lower layer left of j,

5.6.6.
$$y_n b_n > x_n a_n.$$
Combining lines 5.6.5 and 5.6.6 gives $y b_n > x a_n$, and so

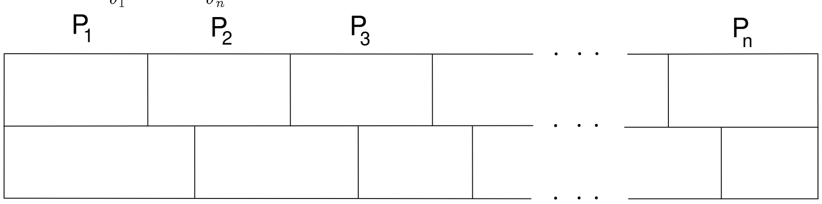
5.6.7.
$$\frac{y}{x} > \frac{a_n}{b_n}.$$

Combining lines 5.6.4 and 5.6.7 gives $\frac{v}{u} > \frac{a_n}{b_n}$. An argument just like that of the previous paragraph gives $\frac{z}{w} < \frac{a_n}{b_n}$, and so $\frac{v}{u} < \frac{a_n}{b_n}$. This contradiction concludes the proof of 1).

Statement 2) can be proved just as statement 1) was proved, or it can be reduced to statement 1) by means of the horizontal reflection.

Now consider statement 3). Suppose that T is not prime. It must be shown that the joint ratio of the right side of T is not both the unique minimal right joint ratio and the unique maximal left joint ratio of T.

First suppose that all of the breaks of T are right breaks. Let $P_1, \ldots, P_n$ be the prime factors of T with reduced weight ratios $\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}$.



Arguing as in the paragraph following line 5.6.4, it follows that

$$\frac{a_1}{b_1} \leq \cdots \leq \frac{a_n}{b_n},$$

hence $b_i a_1 \leq a_i b_1$ for $i = 1, \ldots, n$, hence $xa_1 \leq yb_1$, where $x = \sum_{i=1}^{n} b_i$ and $y = \sum_{i=1}^{n} a_i$, and so $\frac{a_1}{b_1} \leq \frac{y}{x}$. This shows that the right joint ratio $\frac{y}{x}$ of the right side of T is at least as large as the right joint ratio $\frac{a_1}{b_1}$ of the break between P_1 and P_2, as desired.

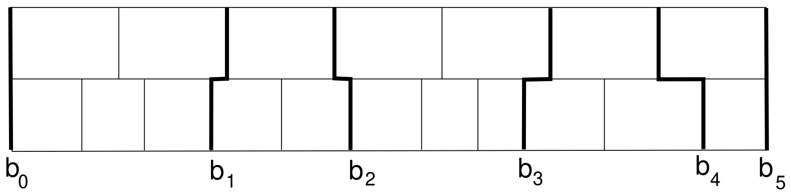
The argument above proves 3) if all of the breaks of T are right breaks. It is now easy to see that 3) is also true if all of the breaks of T are left breaks.

Finally suppose that T has both right breaks and left breaks. By applying the horizontal reflection if necessary, it may be assumed that the leftmost break of T is a right break. The argument of the next-to-last paragraph shows that the right joint ratio of this leftmost break of T is less than or equal to the left joint ratio of the leftmost left break of T. Since the right joint ratios are minimized and the left joint ratios are maximized, if the minimal value equals the maximal value, then these values equal the joint ratios of the leftmost right and left breaks of T. This finishes the proof of 3).
◇ Theorem 5.6.2 ◇

Theorem 5.6.2 provides a **joint ratio algorithm** for computing factorizations of 2-layer valence 3 tilings of quadrilaterals as follows. Compute all right joint ratios of T. All minimal values of these joint ratios occur at breaks or the right side of T. Compute all left joint ratios of T. All maximal values of these joint ratios occur at breaks or the right side of T. If no breaks of T arise from this minimization and maximization, then T is prime. Otherwise, perform analogous minimizations and maximizations with the factors of T determined by all of the breaks found above. This process eventually ends with the prime factorization of T. This completes the description of the joint ratio algorithm.

In practice it is easier to apply the first two phases of the left factor algorithm than the joint ratio algorithm to a general 2-layer valence 3 tiling T of a quadrilateral. In other words, it is best to reserve the joint ratio algorithm for tilings of the sort handled by the third phase of the left factor algorithm. In this case it is only necessary to compute joint ratios of the potential breaks given by Lemma 5.5.4.

Example 5.6.8. Consider the tiling T of Example 5.5.5, given below.



The right joint ratios for b_2, b_4, b_5 are $\frac{5}{3}$, $\frac{10}{6}$, $\frac{11}{7}$. The minimum is $\frac{11}{7}$, which yields no breaks. The left joint ratios for b_1, b_3, b_5 are $\frac{3}{2}$, $\frac{8}{5}$, $\frac{11}{7}$. The maximum is $\frac{8}{5}$. Thus b_3 is a break.

Now consider $T_{0,3}$, using the notation of Example 5.5.5. The right joint ratios for b_2, b_3 are $\frac{5}{3}$, $\frac{8}{5}$. The minimum is $\frac{8}{5}$. The left joint ratios for b_1, b_3 are $\frac{3}{2}$, $\frac{8}{5}$. The maximum is $\frac{8}{5}$. Thus $T_{0,3}$ is prime with weight ratio $\frac{8}{5}$.

Finally, consider $T_{3,5}$. The right joint ratios for b_4, b_5 are $\frac{2}{1}$, $\frac{3}{2}$. The minimum is $\frac{3}{2}$. The only left joint ratio is $\frac{3}{2}$. Thus $T_{3,5}$ is prime with weight ratio $\frac{3}{2}$.

The weight ratio algorithm shows that T is the product of $T_{0,3}$ and $T_{3,5}$.

6. Approximating combinatorial moduli.

How does one approximate the combinatorial moduli of a shingled ring? In Section 7 we shall explain a connection between combinatorial moduli and the recognition of discrete groups of constant negative curvature in dimension 3. This treatment requires that a whole sequence of combinatorial moduli be approximated.

Recall that if S is a finite collection of sets that covers a ring or quadrilateral R in the plane or 2-sphere, then we have combinatorial moduli $M_{\sup}(R,S)$ and $m_{\inf}(R,S)$. Our aim is to approximate these two numbers.

The conformal or quasiconformal compatibility of an entire sequence of covers is measured by the notion of *conformality*. A sequence $S(1)$, $S(2)$, ... of locally finite covers of the 2-sphere S^2 or the plane E^2 with mesh approaching 0 is said to be **conformal** if there is a positive number K having the following properties:

Axiom (1). *Approximate moduli are almost well-defined.* That is, if R is a ring in S^2 (E^2), then there is a positive integer I and a positive number m such that, for all $i \geq I$, the combinatorially defined moduli

$$m_{\inf}(R, S(i)) \quad \text{and} \quad M_{\sup}(R, S(i))$$

both lie in the interval $[m, K \cdot m]$. The number m is called an **approximate modulus** for R.

Axiom (2). *Points of S^2 (E^2) are encircled by small rings of large approximate modulus.* That is, given $p \in S^2$ (E^2), a neighborhood N of p in S^2, and a positive number M, there is a ring R encircling p in N such that any approximate modulus m for R is at least M.

6.1. Skinny approximation.

In this section we obtain estimates on moduli from the existence of skinny paths with special properties.

Given a tiling of a quadrilateral or ring, the finite Riemann mapping theorem shows that there exists a grid consisting of flows and cuts such that the paths in one direction are fat, the paths in the other direction are skinny and the modulus is the number of cuts divided by the number of flows. Theorem 6.1.1 states that if we allow the paths in both directions to be skinny, then the number of cuts divided by the number of flows lies between our two moduli.

Theorem 6.1.1. *Let T be a tiling of a quadrilateral or ring X. Let $(f_1, \ldots, f_C)$ be a family of parallel skinny flows of T, and let $(c_1, \ldots, c_H)$ be a family of parallel skinny cuts of T for some positive integers C and H. Let $(\alpha_1, \ldots, \alpha_C)$ be a family of underlying piecewise smooth arcs for $(f_1, \ldots, f_C)$, and let $(\beta_1, \ldots, \beta_H)$ be a family of underlying piecewise smooth arcs or simple closed curves for $(c_1, \ldots, c_H)$. Suppose that every α_i meets every β_j exactly once, and given a tile t in T, the number of α_i's meeting the interior of t equals the number of β_j's meeting the interior of t and every such α_i meets every such β_j in the interior of t. Then*

$$m_f \leq \frac{H}{C} \leq M_f.$$

Proof. Define a weight function w on T as follows. If t is a tile in T, then $w(t)$ is the number of α_i's which meet the interior of t.

We first estimate A_w as follows.

$$A_w = \langle w, w\rangle = \langle \sum_{i=1}^{C} f_i, \sum_{j=1}^{H} c_j\rangle = \sum_{i=1}^{C}\sum_{j=1}^{H}\langle f_i, c_j\rangle \leq \sum_{i=1}^{C}\sum_{j=1}^{H} 1 = CH.$$

Next let p be a minimal fat cut for w. Then

$$C_{w,f} = \langle p, w\rangle = \langle p, \sum_{i=1}^{C} f_i\rangle = \sum_{i=1}^{C}\langle p, f_i\rangle \geq \sum_{i=1}^{C} 1 = C.$$

Thus

$$\frac{H}{C} = \frac{CH}{C^2} \geq \frac{A_w}{C_{w,f}^2} \geq m_f,$$

which gives the first inequality of Theorem 6.1.1.

Next let p be a minimal fat flow for w. Then

$$H_{w,f} = \langle p, w\rangle = \langle p, \sum_{j=1}^{H} c_j\rangle = \sum_{j=1}^{H}\langle p, c_j\rangle \geq \sum_{j=1}^{H} 1 = H.$$

Thus

$$\frac{H}{C} = \frac{H^2}{CH} \leq \frac{H^2_{w,f}}{A_w} \leq M_f,$$

which gives the second inequality of Theorem 6.1.1.

This proves Theorem 6.1.1.
$\diamondsuit$ Theorem 6.1.1 $\diamondsuit$

It will be convenient in investigating Axiom (2) of Cannon [8] to allow degenerate rings in which one of the boundary components is a point. That is, a **degenerate ring** is a closed topological disk X together with a distinguished point x_0 in its interior. Suppose S is a shingling of a degenerate ring X with distinguished point x_0. A **fat cut** is a fat path which has an underlying topological path separating the boundary of X and the point x_0. A **skinny flow** is a skinny path whose shingles can be ordered such that one of its extreme shingles contains x_0 and the other extreme shingle meets the boundary of X. With these definitions, one defines the moduli m_f and M_s as for a ring.

Proposition 6.1.2. *Let X be a quadrilateral or ring or degenerate ring, and let T be a tiling of X. Let $w_1, \dots, w_k$ be weight functions on T with disjoint supports for some positive integer k. Suppose that for each $i \in \{1, \dots, k\}$, $C_{w_i,f} \geq 1$. Let $A = \sum_{i=1}^{k} A_{w_i}$. Then*

$$m_f \leq \frac{A}{k^2}.$$

Proof. Let $w = \sum_{i=1}^{k} w_i$. Then $A_w = \sum_{i=1}^{k} A_{w_i} = A$. Let p be a minimal fat cut for w. Then $C_{w,f} = \langle p, w\rangle = \langle p, \sum_{i=1}^{k} w_i\rangle \geq k$. Thus $m_f \leq \frac{A_w}{C^2_{w,f}} \leq \frac{A}{k^2}$.
$\diamondsuit$ Proposition 6.1.2 $\diamondsuit$

Proposition 6.1.3. *Let X be a quadrilateral or ring or degenerate ring, and let T be a tiling of X. Let $f_1, \dots, f_k$ be disjoint skinny flows for T for some positive integer k. Suppose that for each $i \in \{1, \dots, k\}$, f_i contains at most N tiles. Then*

$$m_f \leq \frac{N}{k}.$$

Proof. For each $i \in \{1, \dots, k\}$, let w_i be the weight function defined by $w_i(t) = 1$ if $t \in f_i$ and $w_i(t) = 0$ otherwise. The proof follows immediately from Proposition 6.1.2.
$\diamondsuit$ Proposition 6.1.3 $\diamondsuit$

6.2. The averaging trick. We prove the results of this section using the averaging trick shown to us by Mladen Bestvina which occurs in the proof of Theorem 7.1 in Cannon [8]. In this section we let X denote a quadrilateral or ring.

Let T be a tiling of X. Line 2.4.5.2 shows that $M_s = m_f \leq m_s = M_f$. We wish to know to what extent equality can fail in the inequality. We view T as giving a cellular decomposition of X, and so we can speak of valences of vertices of T. We will see that a bound on the valences of the vertices of T gives a bound on the quotient M_f/m_f.

First consider the simplest case in which every vertex of T has valence 3, except that in case of a quadrilateral the four corner vertices have valence 2. It is easy to see in this case that every skinny path is in fact a fat path (It is here that we use the fact that the four corner vertices have valence 2.) essentially because if two tiles have a vertex in common, then they have an edge in common. It easily follows that optimal weight functions for fat flows are optimal weight functions for skinny flows. This proves the following.

6.2.1. If T is a tiling of a quadrilateral or ring such that every vertex of T has valence 3, except that in case of a quadrilateral the four corner vertices have valence 2, then $M_s = m_f = m_s = M_f$.

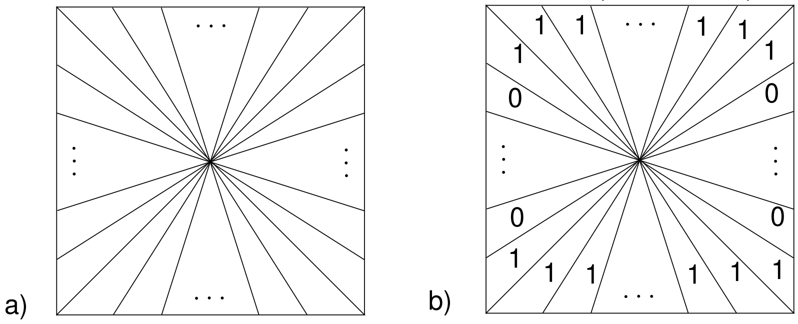


Figure 6.2.3.

Example 6.2.2. This example shows that in general the inequality in line 2.4.5.2 is strict. Let T be the tiling of a square given in a) of Figure 6.2.3. This tiling is invariant under a 90° rotation of the square. Let the valence of the central vertex be $v \geq 8$.

First consider skinny flows for T. It is easy to see that the weight function given in b) of Figure 6.2.3 lies in the cone spanned by its minimal skinny flows. Thus this weight function w is an optimal weight function for skinny flows by line 2.3.6. Hence

$$M_s = \frac{H_{w,s}^2}{A_w} = \frac{4}{\frac{v}{2}+4} = \frac{8}{v+8}.$$

We can compute M_f in the same way, but we proceed as follows. Since T is invariant under a 90° rotation, the problem of optimizing skinny cuts for T is equivalent to the problem of optimizing skinny flows for T. It easily follows that $m_s = M_s^{-1}$. Hence line 2.4.5.2 shows that

$$M_s = m_f = \frac{8}{v+8} < \frac{v+8}{8} = m_s = M_f.$$

Thus $\frac{M_f}{m_f} = \frac{(v+8)^2}{64}$.
◊ Example 6.2.2 ◊

Although Example 6.2.2 shows that in general the inequality in line 2.4.5.2 is strict, the following result shows that the quotient M_f/m_f in Example 6.2.2 is almost the largest possible.

Theorem 6.2.4. *Let T be a tiling of a quadrilateral or ring X. Let v be a positive integer such that the valence of every vertex in T is at most v. Then*

$$\frac{M_f}{m_f} \leq 4v^2.$$

Proof. Let $(p_1, \ldots, p_k)$ be a fundamental family of parallel skinny flows for T, which exists by line 2.4.4.9. Let $(\alpha_1, \ldots, \alpha_k)$ be a family of parallel underlying arcs for $(p_1, \ldots, p_k)$. For every integer i with $1 \leq i \leq k$ let $\varphi(\alpha_i)$ be the set of all tiles in T which meet α_i, so that $\varphi(\alpha_i)$ is a fat flow for $i = 1, \ldots, k$.

Now let w be an optimal weight function for fat flows of T. Let $L_w(p)$ denote the length of a path p relative to w. Then since $H_{w,f} \leq L_w(\varphi(\alpha_i))$ for $i = 1, \ldots, k$,

$$H_{w,f} \leq \frac{1}{k}\sum_{i=1}^{k} L_w(\varphi(\alpha_i)) = \frac{1}{k}\sum_{i=1}^{k}\sum_{t\in\varphi(\alpha_i)} w(t) = \frac{1}{k}\sum_{t\in T} w'(t)w(t),$$

where w' is the weight function on T for which $w'(t)$ is the number of arcs α_i which meet the tile t. This gives the following, where the second inequality comes from the Cauchy-Schwarz inequality.

6.2.5.

$$M_f = \frac{H_{w,f}^2}{A_w} \leq \frac{1}{k^2 A_w}\left(\sum_{t\in T} w'(t)w(t)\right)^2 \leq \frac{1}{k^2 A_w} A_{w'}A_w = \frac{A_{w'}}{k^2}$$

Our next goal is to estimate the weight function w'. Let $w'' = \sum_{i=1}^{k} p_i$, so that w'' is an optimal weight function for skinny flows of T. For every vertex u in T and tile s in T define $w_u(s)$ as follows. First, $w_u(s) = 0$ if $u \notin s$. If $u \in s$, then $w_u(s)$ is the number of arcs $\alpha_1, \ldots, \alpha_k$ which contain u and meet the interior of s.

Now let t be a tile in T. Since the arcs $\alpha_1, \ldots, \alpha_k$ are parallel, it is not difficult to see the following. There exist vertices l_t and r_t in t such that if i is an integer for which α_i meets t, then either α_i contains l_t or α_i contains r_t or α_i meets the interior of t. See Figure 6.2.6.

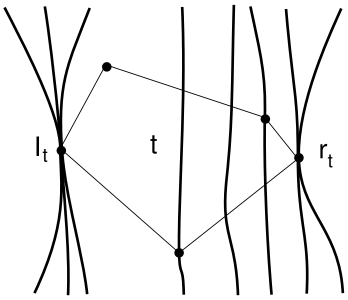


Figure 6.2.6. The tile t, the vertices l_t, r_t and the arcs which meet t.

Define $\psi(t)$ to be the set of tiles in $T \setminus \{t\}$ which contain l_t or r_t. Then,

$$w'(t) \leq w''(t) + \sum_{s\in\psi(t)}\sum_{u\in\{l_t,r_t\}} w_u(s).$$

We now use this inequality to estimate $A_{w'}$. Since l_t and r_t have valence at most v, it is easy to see that the number of nonzero terms in the right side of the last inequality of the form $w''(t)$ or $w_u(s)$ is at most $2v$. Thus the Cauchy-Schwarz inequality shows that

$$w'(t)^2 \leq 2v\left(w''(t)^2 + \sum_{s\in\psi(t)}\sum_{u\in\{l_t,r_t\}} w_u(s)^2\right).$$

Hence

$$\begin{aligned} A_{w'} &\leq 2v\left(A_{w''} + \sum_{t\in T}\sum_{s\in\psi(t)}\sum_{u\in\{l_t,r_t\}} w_u(s)^2\right) \\ &\leq 2v\left(A_{w''} + \sum_{s\in T}\sum_{u\in s}\sum_{t\in T\setminus\{s\},u\in t} w_u(s)^2\right) \\ &\leq 2v\left(A_{w''} + (v-1)\sum_{s\in T}\sum_{u\in s} w_u(s)^2\right). \end{aligned}$$

Let $s \in T$, and let u be a vertex in s with $w_u(s) \neq 0$. Then one of the arcs α_i meets the interior of s and meets the boundary of s in two points, u and some other point u'. We say that the point in $\{u, u'\}$ which lies closer to the bottom of T along α_i lies in the bottom of s. We say that the point in $\{u, u'\}$ which lies closer to the top of T along α_i lies in the top of s. It is clear that the sum of the terms $w_u(s)$ either over all vertices u in the bottom of s or over all vertices u in the top of s is at most $w''(s)$. Hence the sum of the terms $w_u(s)^2$ either over all vertices u in the bottom of s or over all vertices u in the top of s is at most $w''(s)^2$. Thus

$$A_{w'} \leq 2v \left(A_{w''} + (v-1) \sum_{s \in T} 2w''(s)^2 \right) \leq 2v \left(A_{w''} + 2(v-1) A_{w''} \right) \leq 4v^2 A_{w''}.$$

Combining this estimate for $A_{w'}$ with line 6.2.5 gives that $M_f \leq 4v^2 \frac{A_{w''}}{k^2}$. Line 2.3.4 shows that $A_{w''} = kH_{w'',s}$, and so $\frac{A_{w''}}{k^2} = \frac{H^2_{w'',s}}{A_{w''}} = M_s$. Hence $M_f \leq 4v^2 M_s = 4v^2 m_f$, the equality coming from line 2.4.3.6. This proves Theorem 6.2.4.
♢ Theorem 6.2.4 ♢

We next prove a theorem which provides a comparison between the moduli of two different shinglings of a quadrilateral or ring.

Theorem 6.2.7. *Let X be a quadrilateral or ring. Let S and T be two shinglings of X such that every shingle in S and T is the closure of its interior. Suppose that there exists a positive integer K such that for every shingle s in S the number of shingles in T which meet the interior of s is at most K. Likewise suppose that there exists a positive integer L such that for every shingle t in T the number of shingles in S which meet the interior of t is at most L. We augment our usual notation for moduli with parentheses and an S or T to indicate the shingling being used, so for example $M_f(S)$ is the fat flow modulus of X relative to S. Then*

$$\frac{1}{KL} M_s(T) \leq M_s(S) \leq M_f(S) \leq KLM_f(T).$$

Proof. We first prove that $M_s(T) \leq KLM_s(S)$. To begin, define a function φ from the set of flows of S to the set of flows of T so that if p is a flow of S, then $\varphi(p)$ is the set of all shingles in T which meet the interior of some shingle in p. To see that $\varphi(p)$ is a flow, first note that given a shingle s in S, the shingles in T which meet the interior of s cover s. Hence if p is a flow of S, then the shingles in $\varphi(p)$ cover the union of the shingles in p. It easily follows that if p is a flow of S, then $\varphi(p)$ is a flow of T.

Now let w be an optimal weight function for skinny flows of T. Let $L_w(p)$ denote the length of a path p of T relative to w. Let $p_1, \ldots, p_k$ be a fundamental family of skinny flows of S. Since $\varphi(p_i)$ is a flow of T, line 2.4.1.1 shows that $\varphi(p_i)$ contains a skinny flow of T, and so $H_{w,s} \leq L_w(\varphi(p_i))$ for $i = 1, \ldots, k$. Hence

$$H_{w,s} \leq \frac{1}{k} \sum_{i=1}^{k} L_w(\varphi(p_i)) = \frac{1}{k} \sum_{i=1}^{k} \sum_{t \in \varphi(p_i)} w(t) = \frac{1}{k} \sum_{t \in T} w'(t) w(t),$$

where w' is the weight function on T for which $w'(t)$ is the number of flows $\varphi(p_i)$ containing the shingle t. This gives the following, where the second inequality comes from the Cauchy-Schwarz inequality.

6.2.8.

$$M_s(T) = \frac{H^2_{w,s}}{A_w} \leq \frac{1}{k^2 A_w} \left(\sum_{t \in T} w'(t) w(t) \right)^2 \leq \frac{1}{k^2 A_w} A_{w'} A_w = \frac{A_{w'}}{k^2}$$

Our next goal is to estimate $A_{w'}$. Let $w'' = \sum_{i=1}^{k} p_i$, so that w'' is an optimal weight function for skinny flows of S. Given a shingle s in S, let $\tau(s)$ be the set of shingles in T which meet the interior of s.

Likewise, given a shingle t in T, let $\sigma(t)$ be the set of shingles in S which meet the interior of t. Then $w'(t) \leq \sum_{s\in\sigma(t)} w''(s)$. Since $|\sigma(t)| \leq L$, the Cauchy-Schwarz inequality gives that $w'(t)^2 \leq L \sum_{s\in\sigma(t)} w''(s)^2$. Hence

$$\begin{aligned} A_{w'} = \sum_{t\in T} w'(t)^2 &\leq L \sum_{t\in T} \sum_{s\in\sigma(t)} w''(s)^2 \\ &= L \sum_{s\in S} \sum_{t\in\tau(s)} w''(s)^2 \leq KL \sum_{s\in S} w''(s)^2 = KLA_{w''}. \end{aligned}$$

This and line 6.2.8 give that $M_s(T) \leq KL\frac{A_{w''}}{k^2}$. Line 2.3.4 shows that $A_{w''} = kH_{w'',s}$, and so $\frac{A_{w''}}{k^2} = \frac{H^2_{w'',s}}{A_{w''}} = M_s(S)$. Thus $M_s(T) \leq KLM_s(S)$.

The inequality $M_f(S) \leq KLM_f(T)$ can be proved in essentially the same way. Here we take $(p_1, \ldots, p_k)$ to be a fundamental family of fat flows of T. Let $(\alpha_1, \ldots, \alpha_k)$ be a family of underlying arcs for $(p_1, \ldots, p_k)$. Define $\varphi(\alpha_i)$ to be the set of shingles in S which meet α_i, so that $\varphi(\alpha_i)$ is a fat flow of S for $i = 1, \ldots, k$. It is easy to see that the above argument shows in this case that $M_f(S) \leq KLM_f(T)$.

Finally, the inequality $M_s(S) \leq M_f(S)$ comes from line 2.4.5.1.

This proves Theorem 6.2.7.
$\diamondsuit$ Theorem 6.2.7 $\diamondsuit$

6.3. Examples.

6.3.1. Barycentric subdivision.

Let T_0 be any triangulation of the Euclidean plane, viewed as a tiling. Recursively define a sequence $\{T_i\}$ by defining T_i to be the second barycentric subdivision $B^2(T_{i-1})$ of T_{i-1} for $i > 0$.

Theorem 6.3.1.1. *The sequence T_0, T_1, $\ldots$ is not conformal. In fact, it fails to satisfy either of Axioms (1) and (2).*

Before proving Theorem 6.3.1.1, we make some preliminary estimates.

Let x_0 be a vertex of T_0, let R_0 be the closed star of x_0, let e be an edge of T_0 and let R_1 be the closed star of e. See Figure 6.3.1.2. Let n be the number of tiles in R_0. We consider R_0 as a degenerate ring with distinguished point x_0. We estimate the moduli of R_0 and R_1 with respect to the sequence $\{T_i\}$.

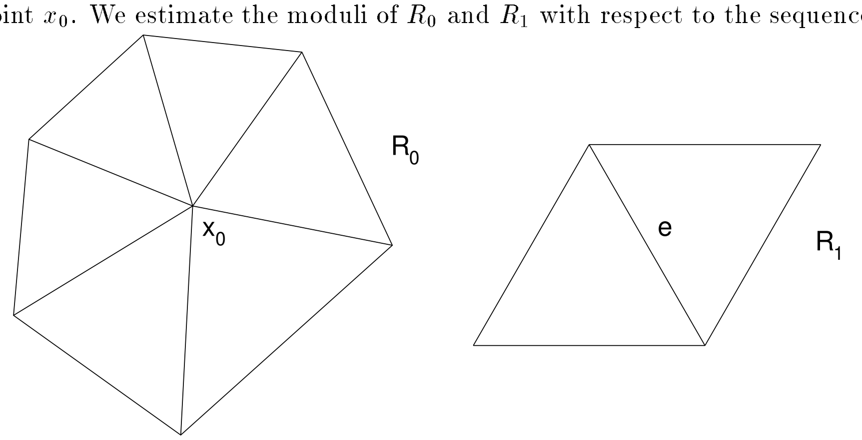


Figure 6.3.1.2.

First use constant weights $w_i \equiv 1$ on the tilings $T_i|R_1$ of R_1 to estimate the suprema $M_f(R_1, T_i)$ and $M_s(R_1, T_i)$ from below and the infima $m_f(R_1, T_i)$ and $m_s(R_1, T_i)$ from above. It is straightforward to show that $\lim_{i\to\infty} M_f(R_1, T_i) = \lim_{i\to\infty} M_s(R_1, T_i) = 0$ and $\lim_{i\to\infty} m_f(R_1, T_i) = \lim_{i\to\infty} m_s(R_1, T_i) = \infty$.

We leave the proofs to the reader: the point is that areas grow by powers of 36 while lengths, both fat and skinny, grow by approximately powers of 4. Hence one learns absolutely nothing asymptotically from constant weights.

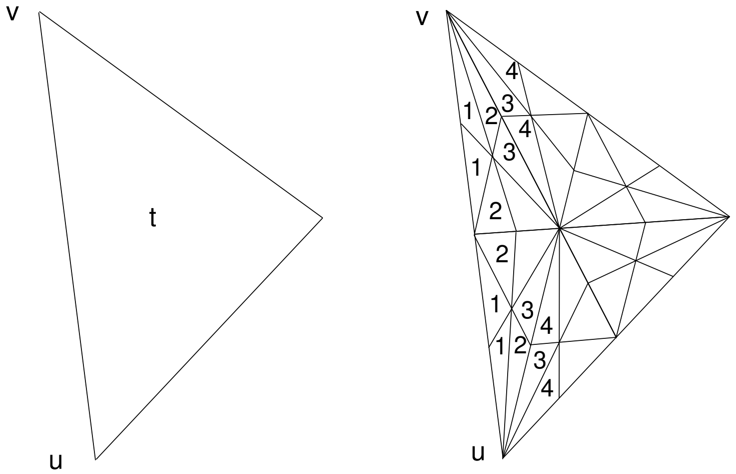


Figure 6.3.1.3.

We next estimate moduli by using skinny paths. The key observation is the following. Let t be a tile of T_i for some i, and let u, v be distinct vertices of t. Then there are 4 disjoint skinny paths $\alpha_1, \dots, \alpha_4$ in $B^2(t)$, labeled 1, 2, 3 and 4 in Figure 6.3.1.3, such that for $j \in \{1, \dots, 4\}$ α_j has 4 tiles and they can be ordered such that one of the extreme tiles contains u and the other extreme tile contains v.

6.3.1.4. For every nonnegative integer i, $m_f(R_0, T_i) \le \frac{1}{n}$.

Proof. There are n disjoint skinny flows of the tiling $T_0|R_0$ of the degenerate ring R_0, each consisting of one tile. It follows inductively from the above observation that if $i \ge 0$, there are $n4^i$ disjoint skinny flows $\alpha_1, \dots, \alpha_{n4^i}$ in the tiling $T_i|R_0$ of R_0, and each α_j, $j \in \{1, \dots, n4^i\}$, contains 4^i tiles. It follows from Proposition 6.1.3 that if $i \ge 0$, $m_f(R_0, T_i) \le \frac{4^i}{n4^i} = \frac{1}{n}$.
◊ 6.3.1.4 ◊

6.3.1.5.

$$\lim_{i \to \infty} m_f(R_1, T_i) = 0.$$

Proof. First consider the tiling $T_1|R_1$ of the quadrilateral R_1. See Figure 6.3.1.6. For each $j \in \{1, \dots, 8\}$, let $w_{1,j}$ be the weight function defined on $T_1|R_1$ by $w_{1,j}(t) = 0$ if t is not labeled j, $w_{1,j}(t) = 1/2$ if t is colored grey and is labeled j, and $w_{1,j}(t) = 1$ if t is colored white and is labeled j. Then $w_{1,1}, \dots, w_{1,8}$ have disjoint supports, and for $j \in \{1, \dots, 8\}$ $C_{w_{1,j},f} = 1$. Furthermore, $A_{w_{1,j}} = 4$ if $j \in \{1, \dots, 6\}$ and $A_{w_{1,7}} = A_{w_{1,8}} = 3$. By Proposition 6.1.2, $m_f(R_1, T_1) \le \frac{6\cdot 4 + 2\cdot 3}{8^2} = \frac{15}{32}$. Note that for each $j \in \{1, 3, 5, 7\}$, the tiles in the support of $w_{1,j} + w_{i,j+1}$ can be grouped in blocks of 4 tiles of equal weights as in Figure 6.3.1.7.

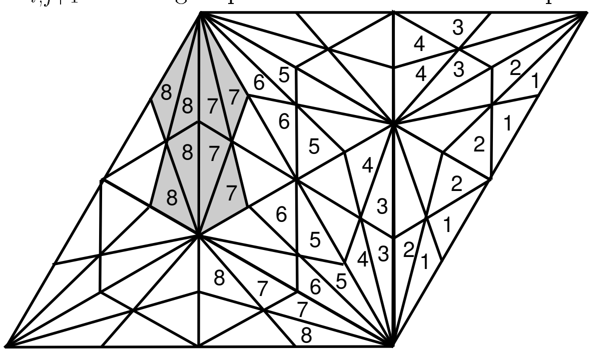


Figure 6.3.1.6.

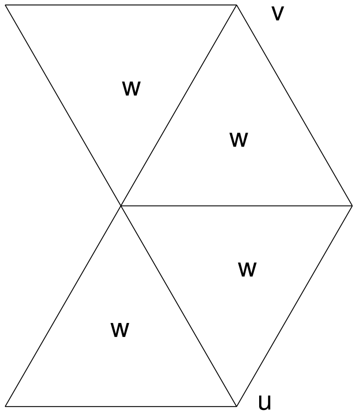


Figure 6.3.1.7.

We show inductively that if $i \in \mathbf{N}$, there are 2^{2i+1} disjointly supported weight functions $w_{i,1}, \ldots, w_{i,2^{2i+1}}$ on $T_i|R_1$ such that $\sum_{j=1}^{2^{2i+1}} A_{w_{i,j}} = 2(15)^i$ and for $j \in \{1, \ldots, 2^{2i+1}\}$, $C_{w_{i,j},f} = 1$, the tiles in the support of $w_{i,j} + w_{i,j+1}$ for j odd can be grouped in fundamental blocks of 4 tiles of equal weights as in Figure 6.3.1.7, and if two fundamental blocks for $w_{i,j} + w_{i,j+1}$ have nonempty intersection the intersection contains one of the vertices labeled as in Figure 6.3.1.7.

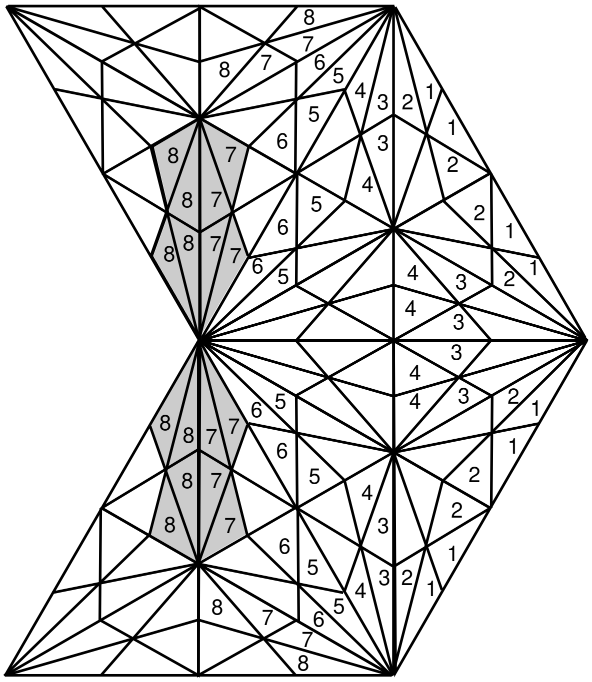


Figure 6.3.1.8.

We have already shown this for $i = 1$, so suppose that the above is true for some positive integer i. Let j be an integer such that $2j+1 \in \{1, \ldots, 2^{2i+1}\}$. Define weight functions $w_{i+1,8j+k}$, $k \in \{1, \ldots, 8\}$, as follows. Let F be a fundamental block in the support of $w_{i,2j+1} + w_{i,2j+2}$ with weight w as shown in Figure 6.3.1.7. Let t be a tile in $B^2(F)$. First suppose that F contains two tiles in the support of $w_{i,2j+1}$ and two tiles in the support of $w_{i,2j+2}$. See Figure 6.3.1.8. In this case $w_{i+1,8j+k}(t) = 0$ if t is not labeled k, $w_{i+1,8j+k}(t) = w/2$ if t is colored grey and is labeled k, and $w_{i+1,8j+k}(t) = w$ if t is colored white and is labeled k. Now suppose that all four tiles in F are in the support of $w_{i,2j+1}$. In this case $w_{i+1,8j+k}(t) = 0$ if $k \in \{5, 6, 7, 8\}$, and if $k \in \{1, 2, 3, 4\}$, then $w_{i+1,8j+k}(t) = 0$ if t is not labeled k or $k+4$, $w_{i+1,8j+k}(t) = w/2$ if t is colored grey and is labeled $k+4$, and $w_{i+1,8j+k}(t) = w$ if t is colored white and is labeled k or $k+4$. Finally suppose that all four tiles in F are in the support of $w_{i,2j+2}$. In this case $w_{i+1,8j+k}(t) = 0$ if $k \in \{1, 2, 3, 4\}$, and if $k \in \{5, 6, 7, 8\}$, then $w_{i+1,8j+k}(t) = 0$ if t is not labeled k or $k-4$, $w_{i+1,8j+k}(t) = w/2$ if t is colored grey and is labeled k, and $w_{i+1,8j+k}(t) = w$ if t is colored white and is labeled k or $k-4$. Then $C_{w_{i+1,j},f} = 1$ for every j, and the weights in the support of $w_{i+1,j} + w_{i+1,j+1}$ for j odd can be grouped in fundamental blocks of equal weight as in Figure 6.3.1.7. The sum of the $w_{i,j}$-areas of the fundamental block F is $4w^2$, and the sum of the $w_{i+1,j}$-areas of F is $14 \cdot 4 \cdot w^2 + 4 \cdot 4 \cdot \frac{w^2}{4} = 60w^2$. Hence $\sum_{j=1}^{2^{2i+3}} A_{w_{i+1,j}} = 2(15)^{i+1}$. This completes the proof of the induction step.

By Proposition 6.1.2, if $i \in \mathbf{N}$ then $m_f(R_1, T_i) \leq \frac{2(15)^i}{(2^{2i+1})^2} = \frac{1}{2}\left(\frac{15}{16}\right)^i$. Hence $\lim_{i\to\infty} m_f(R_1, T_i) = 0$.
$\diamondsuit$ 6.3.1.5 $\diamondsuit$

6.3.1.9.

$$\lim_{i\to\infty} m_f(R_0, T_i) = 0.$$

Proof. The proof is very similar to the proof of line 6.3.1.5. The details are left to the reader.
$\diamondsuit$ 6.3.1.9 $\diamondsuit$

Proof of Theorem 6.3.1.1. If $\{T_i\}$ satisfies the first axiom of conformality, then there is a positive real number K such that if R is a ring, then there is a positive real number m such that for i sufficiently large, $m_f(R, T_i)$ and $M_f(R, T_i)$ both lie in the interval $[m, Km]$. Suppose R is any ring in R_0 that is a union of tiles in one of the tilings in $\{T_i\}$, and that separates x_0 and the boundary of R_0. The proofs of lines 6.3.1.5 and 6.3.1.9 were based on constructing disjointly supported weight functions on $T_i|R_0$ and then using Proposition 6.1.2. Since the hypotheses of Proposition 6.1.2 are also satisfied by the restrictions of these weight functions to $T_i|R$, the same argument shows that $\lim_{i\to\infty} m_f(R, T_i) = 0$ and hence it is impossible for $m_f(R, T_i)$ to lie in a K-interval $[m, Km]$ for i sufficiently large. Hence $\{T_i\}$ does not satisfy the first axiom of conformality.

The sequence $\{T_i\}$ fails the first axiom in a stronger fashion. Let $R = R_0 \setminus \mathrm{star}(x_0, B^2(R_0))$. Then $\lim_{i\to\infty} m_f(R, T_i) = 0$ as shown above. There are four parallel skinny cuts in $T_1|R$ constructing by using the labeling of Figure 6.3.1.3 in each of the n tiles of $T_0|R_0$. An adaptation of the arguments used in proving lines 6.3.1.5 and 6.3.1.9 shows that $\lim_{i\to\infty} M_f(R, T_i) = \infty$, and hence for any real number K, $m_f(R, T_i)$ and $M_f(R, T_i)$ do not lie in a K-interval if i is sufficiently large.

If $\{T_i\}$ satisfies the second axiom of conformality, then given a positive real number M, there is a ring R in R_0 separating x_0 and the boundary of R_0 with $m_f(R, T_i) \geq M$ for i sufficiently large. Let j be an integer large enough so that $R \subset R' = R_0 \setminus \mathrm{star}(x_0, B^{2j}(R_0))$. It is not difficult to see that $m_f(R, T_i) \leq m_f(R', T_i)$ for every i. Thus it is impossible that $m_f(R, T_i) \geq M$ for i sufficiently large because we have shown that $\lim_{i\to\infty} m_f(R', T_i) = 0$.
$\diamondsuit$ Theorem 6.3.1.1 $\diamondsuit$

6.3.2. Square tilings.

Let T_0 denote the tiling of the plane by squares of unit edge with vertices on the integer lattice. Let T_{i+1} denote the tiling formed by subdividing each square of T_i into four squares of equal size.

Theorem 6.3.2.1. *The sequence T_0, T_1, $\ldots$ of square tilings of the plane is conformal.*

Proof. We outline two proofs, one geometric and the other combinatorial.

The geometric proof simply quotes a theorem of Cannon [8]. The square tilings of the plane are obviously "almost round" in the sense of Section 7 of [8]; hence the sequence of square tilings gives combinatorial moduli comparable to the standard analytic moduli by Theorem 7.1 of [8]. Hence the sequence is conformal as claimed.

The combinatorial proof will not be complete since we prove Axiom 1 only for special rings. However, the proof we give illustrates important techniques and can be extended by fussy effort to a complete proof. The proof of Axiom 1 will be modeled on the proof of Theorem 7.3 of [8]. The proof of Axiom 2 will use the bounded valence theorem, Theorem 6.2.4.

Axiom 1 for special rings. Let R be a ring that is a union of tiles of T_0, no one of which intersects both ends of R. Let $M_i = M_{\mathrm{sup}}(R, T_i)$ and $m_i = m_{\mathrm{inf}}(R, T_i)$. We claim that there is a constant $K > 0$, independent of i and of R, such that

6.3.2.2.
$$\frac{1}{K} m_0 \leq m_i \leq M_i \leq K M_0.$$

The bounded valence theorem, Theorem 6.2.4, and line 2.4.5.1 imply the existence of a constant $L > 0$, independent of i and R, such that

6.3.2.3.
$$m_i \leq M_i \leq L m_i.$$

Hence m_i and M_i lie in the interval $[m, K^2 L \cdot m]$, where $m = (1/K)m_0$, i is arbitrary and K and L are independent of R, so that Axiom 1 of conformality is satisfied for these special rings R.

In order to establish the claim, we take weight functions σ and τ realizing M_0 as $H^2(R, \sigma)/A(R, \sigma)$ and m_0 as $A(R, \tau)/C^2(R, \tau)$. We define weight functions σ_i and τ_i on $T_i|R$ as follows. If t' is a tile of $T_i|R$, define

$$\sigma_i(t') = \max\{\sigma(t) \mid t \in T_0|R \text{ and } t' \subset \mathrm{Nbd}(t, 1/2)\},$$

where $\mathrm{Nbd}(X, \epsilon)$ denotes the closed Euclidean ϵ-neighborhood of X in the plane, with the sup norm. Define τ_i by a similar equation which uses τ instead of σ.

An obvious calculation shows that $A(R, \rho_i) \leq 2^{2i+2} A(R, \rho)$ for $\rho = \sigma$ and for $\rho = \tau$. The real point of the choice of the weight functions σ_i and τ_i shows up in the following length estimate. Let α be a path joining the ends of R and let β be a path separating the ends of R, with $L_{\sigma_i}(\alpha) = H(R, \sigma_i)$ and $L_{\tau_i}(\beta) = C(R, \tau_i)$. If α intersects a tile $t \in T_0|R$, then because t does not intersect both ends of R, α crosses the annulus between t and $E^2 \setminus \mathrm{Nbd}(t, 1/2)$. The intersection with this annulus contributes at least $2^{i-1} \cdot \sigma(t)$ to the σ_i-length of α. Every tile weight in this contribution might be counted with respect to as many as three tiles t. We conclude that

$$H(R, \sigma_i) = L_{\sigma_i}(\alpha) \geq \frac{1}{3} \cdot 2^{i-1} \cdot L_\sigma(\alpha) \geq \frac{1}{3} \cdot 2^{i-1} \cdot H(R, \sigma).$$

Similarly,

$$C(R, \tau_i) = L_{\tau_i}(\beta) \geq \frac{1}{3} \cdot 2^{i-1} \cdot L_\tau(\beta) \geq \frac{1}{3} \cdot 2^{i-1} \cdot C(R, \tau).$$

We conclude that

$$M_i \geq \frac{H^2(R, \sigma_i)}{A(R, \sigma_i)} \geq \frac{(1/9)2^{2i-2} H^2(R, \sigma)}{2^{2i+2} A(R, \sigma)} = \frac{1}{144} M_0.$$

Similarly

$$m_i \leq \frac{A(R, \tau_i)}{C^2(R, \tau_i)} \leq \frac{2^{2i+2} A(R, \tau)}{(1/9)2^{2i-2} C^2(R, \tau)} = 144 m_0.$$

We use 6.3.2.3 to deduce that

$$m_0 \leq M_0 \leq 144 M_i \leq 144 L m_i$$

and

$$M_0 \geq m_0 \geq \frac{1}{144} m_i \geq \frac{1}{144L} M_i.$$

These latter two inequalities yield 6.3.2.2 provided that $K \geq 144L$.

Axiom 2. Pick a tiling T_i. Let R be a ring formed by taking a square with n^2 tiles from T_i and deleting a concentric square with $(m-1)^2$ tiles, $m - 1 < n$. We estimate the modulus of R as follows. Divide the tiles of $T_i|R$ into layers, one tile thick, circling R. Successive layers, from inside out, will have

$$4m, \ 4m + 8, \ 4m + 16, \ \ldots, \ 4n - 4$$

tiles. Define a weight function τ on T_i so that the tiles t in the layer having $4k$ tiles each have the weight $\tau(t) = 1/4k$. Then

$$Lm_{\inf}(R, T_i) \geq M_{\sup}(R, T_i) \geq \frac{H^2(R, \tau)}{A(R, \tau)}.$$

$$H(R, \tau) = \frac{1}{4m} + \frac{1}{4m+8} + \cdots + \frac{1}{4n-4} = A(R, \tau).$$

We deduce that

$$\begin{aligned}\frac{H^2(R,\tau)}{A(R,\tau)} &= \frac{1}{4m} + \frac{1}{4m+8} + \cdots + \frac{1}{4n-4} \\ &\geq \int_m^n \frac{dx}{8x} \\ &= \frac{1}{8}\ln(n/m),\end{aligned}$$

so that $m_{\inf}(R, T_i)$ is large for n/m large.

We leave as an exercise the proof, which we do not need, that

$$m_{\inf}(R, T_i) = A(R, \tau)/C^2(R, \tau).$$

7. Variable negative curvature versus constant curvature groups.

Thurston's geometrization conjecture [21] would require that a 3-manifold admitting a Riemannian metric of variable negative curvature also admit a metric of constant negative curvature. The finite Riemann mapping theorem of Section 3 and the combinatorial Riemann mapping theorem of Cannon [8] have potential application to the finding of a constant curvature metric. This expository section explains the connection.

We assume that G is a discrete group that is **negatively curved** (in the large). That is, G is **word hyperbolic** in the sense of Gromov (see [15], [7], [14], [11], [2], [3]). In order to apply the mapping theorems, we need to assume that G has as its space $S_\infty(G)$ at infinity the 2-sphere S^2 (see, for example, [15] or [20]), and we need to extract from G finite coverings of this 2-sphere. Both the space at infinity and the requisite finite coverings are defined in terms of a finitely generated Cayley group graph Γ for G (see [5], or [6], or any of a number of other places for a description of Cayley graphs). Points of $S_\infty(G)$ are equivalence classes $[R]$ of geodesic rays $R : [0, \infty) \to \Gamma$, rays R and S being equivalent if $d(R(t), S(t))$ is bounded for $t \in [0, \infty)$. Both the topology at infinity and the requisite coverings of the space at infinity are defined in terms of what we call **combinatorial half spaces** and **combinatorial disks at infinity** (see [7] for motivations and discussion). Fix a base vertex O of Γ. Let $R : [0, \infty) \to \Gamma$ be a geodesic ray such that $R(0) = O$. Let n be a positive integer. The **half space** $H(R, n)$ is defined by the equation

$$H(R, n) = \{\ x \in \Gamma | d(x, R([n, \infty))) \leq d(x, R([0, n)))\},$$

and the **combinatorial disk** $D(R, n)$ is defined by the equation

$$D(R, n) = \{\ [R'] \mid R'(0) = O \text{ and } \lim_{t\to\infty} d(R'(t), \Gamma \setminus H(R, n)) = \infty\}.$$

For n a positive integer, we define $O(n)$ to be the collection of all combinatorial disks of the form $D(R, n)$ where R is a geodesic ray with initial point O.

Eric Swenson [20] has proved a number of important facts about these sets. A few of them are summarized in the following theorem.

Theorem 7.1. *[Swenson, 20]. Let G be a negatively curved group, Γ a finitely-generated Cayley graph for G, and O a base vertex for Γ.*

(1) The disks $D(R, n)$ based at O form a basis for a compact metric topology on $S_\infty(G)$, and this topology is equivalent with Gromov's topology.

(2) For each positive integer n, the set $O(n)$ of disks $D(R,n)$ based at O forms a finite open cover of $S_\infty(G)$; and, as n varies, these coverings $O(n)$ have uniformly bounded degree. Furthermore, the mesh of these coverings goes to 0 as $n \to \infty$. Consequently, $\dim(S_\infty(G))$ is finite.

(3) The coverings $O(n+1)$ can be derived from the coverings $O(n)$ by a linearly recursive rule which can be interpreted as subdivision.

Conjecture. If $S_\infty(G)$ is a k-sphere for some k, then each open disk $D(R,n)$ is contractible, or at least has trivial homology.

We say that a discrete group G has **constant negative curvature** (in the large), or that G is **hyperbolic**, if G acts cocompactly, properly discontinuously, and isometrically on some hyperbolic space H^n. Cannon and Cooper [9] have characterized groups of constant negative curvature in terms of the geometry of a finitely-generated Cayley group graph Γ:

Theorem 7.2. *[9]. A discrete group G has constant negative curvature (in the large) in dimension 3 if and only if a finitely-generated Cayley group graph Γ for G is quasi-isometric with H^3.*

Cannon and Cooper are under the impression that their proof works also in higher dimensions.

Cannon and Swenson have characterized 3-dimensional discrete groups of constant negative curvature in terms of the sequence $O(1)$, $O(2)$, ... of covers defined above.

Theorem 7.3. *[10]. A discrete group G acts cocompactly, properly discontinuously, and isometrically on H^3 if and only if the following conditions are satisfied:*

(1) The group G is negatively curved (in the large) and has as its space $S_\infty(G)$ at infinity the 2-sphere.

(2) The sequence $O(1)$, $O(2)$, ... of covers defined above is conformal in the sense of Section 6.

The proof involves a good deal of the geometry of negatively-curved spaces from Swenson's thesis [20], the combinatorial Riemann mapping theorem [Cannon, 8], and the Sullivan-Tukia result [Sullivan, 19 (see, in particular, p. 468)] on uniformly quasiconformal group actions on S^2. The Cannon-Swenson characterization, Theorem 7.3, reduces the 3-dimensional problem broached by Theorem 7.2 to a (difficult) 2-dimensional problem, namely the approximation of combinatorial moduli. Much remains to be done.

References

[1] Ahlfors, L. V., *Conformal Invariants: Topics in Geometric Function Theory*, McGraw-Hill, New York, 1973.

[2] Alonso, J. M., Brady, T., Cooper, D., Ferlini, V., Lustig, M., Mihalik, M., Shapiro, M., Short, H., *Notes on word hyperbolic groups*, Group Theory from a Geometrical Viewpoint: 21 March - 6 April 1990, ICTP, Trieste, Italy (E. Ghys, A. Haefliger and A. Verjovsky, eds.), World Scientific, Singapore, 1991, pp. 3-63.

[3] Bowditch, B. H., *Notes on Gromov's hyperbolicity criterion for path-metric spaces*, Group Theory from a Geometrical Viewpoint: 21 March - 6 April 1990, ICTP, Trieste, Italy (E. Ghys, A. Haefliger and A. Verjovsky, eds.), World Scientific, Singapore, 1991, pp. 64-167.

[4] Brooks, R. L., Smith, C. A. B., Stone, A. H., Tutte, W. T., *The dissection of squares into squares*, Duke Math. J. **7** (1940), 312–340.

[5] Cannon, J. W., *The combinatorial structure of cocompact discrete hyperbolic groups*, Geometriae Dedicata **16** (1984), 123–148.

[6] Cannon, J. W., *Almost convex groups*, Geometriae Dedicata **22** (1987), 197–210.

[7] Cannon, J. W., *The theory of negatively curved spaces and groups*, Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces (T. Bedford, M. Keane and C. Series, eds.), Oxford University Press, Oxford, 1991, pp. 315–369.

[8] Cannon, J. W., *The combinatorial Riemann mapping theorem*, submitted to *Acta Mathematica*.

[9] Cannon J. W., Cooper, D., *A characterization of compact hyperbolic and finite-volume groups in dimension three*, to appear in *Trans. Amer. Math. Soc.*.

[10] Cannon, J. W., Swenson, E. L., *Recognizing constant curvature discrete groups*, in preparation.

[11] Coornaert, M., Delzant, T., Papadopoulos, A., *Géométrie et théorie des groupes*, Springer Lecture Notes Series 1441, Springer-Verlag, New York Berlin Heidelberg, 1990.

[12] Dudeney, H. E., *The Canterbury Puzzles*, (This is a reprint of the 1907 edition.), Dover, New York, 1958.

[13] Gardner, M., *The Second Scientific American Book of Mathematical Puzzles and Diversions*, University of Chicago Press, Chicago, 1987.

[14] E. Ghys and P. de la Harpe (eds.), *Sur les Groupes Hyperboliques d'aprés Mikhael Gromov*, Progress in Mathematics. Vol 83, Birkhaüser, Zurich, 1990.

[15] Gromov, M., *Hyperbolic groups*, Essays in Group Theory (S. M. Gersten, ed.), Springer-Verlag, New York Berlin Heidelberg, 1987, pp. 75–264.

[16] Lehto, O., Virtanen, K. I., *Quasiconformal Mappings in the Plane*, Springer-Verlag, New York, 1973.

[17] Poincaré, H., *The Value of Science*, Dover, New York, 1958.

[18] Schramm, O., *Square tilings with prescribed combinatorics*, preprint.

[19] Sullivan, D., *On the ergodic theory at infinity of an arbitrary discrete group of hyperbolic motions*, Riemann Surfaces and Related Topics: Proceedings of the 1978 Stony Brook Conference (I. Kra and B. Maskit, eds.), Princeton University Press, Princeton, 1981, pp. 465–496.

[20] Swenson, E. L., *Negatively curved groups and related topics*, Dissertation, BYU (1993).

[21] Thurston, W. P., *Three dimensional manifolds, Kleinian groups and hyperbolic geometry*, Bull. Amer. Math. Soc. **6, # 3** (1982), 357–381.

Department of Mathematics, BYU, Provo, UT 84602, U.S.A.

Department of Mathematics, VPI& SU, Blacksburg, VA 24061-0123, U.S.A.

Department of Mathematics, EMU, Ypsilanti, MI 48197, U.S.A.