On Spaces of Inscribed Triangles

Rich Schwartz

January 10, 2019

1 Introduction

A *Jordan loop* is the image of a circle under a continuous injective map into the plane. The purpose of this paper is to prove some results about triangles inscribed in Jordan Loops. Some results are known about this problem. For instance, the result due to M. Meyerson ([**M**],1980) says that all but at most 2 points of any Jordan loop are vertices of inscribed equilateral triangles. This result is sharp because two points of a suitable isosceles triangle are not vertices of inscribed equilateral triangles. Another result, due to M. Neilson ([**N**], 1992) says that, for any given shape λ, a dense set of points are vertices of inscribed triangles of shape λ. Here, a triangle of *shape* $\lambda \in \boldsymbol{C}$ (which we take to lie in the upper halfplane) is one that is similar to the triangle with vertices $0, 1, \lambda$.

One might be interested in questions about inscribed triangles because it relates to the famous Square Peg Conjecture of Toeplitz: Does every Jordan loop have 4 points which are vertices of a square? See [**Ma**] or [**P**] for historical surveys of this problem. One connection to triangles, for instance, is that every inscribed square gives rise to 4 inscribed right-angled isosceles triangles.

We say that a polygon P is *gracefully inscribed* in a Jordan loop J if every vertex of P lies on J and if the cyclic order on the vertices of P is the same whether it is computed with respect to the counter-clockwise cyclic order on P or the counter-clockwise cyclic order on J. For instance, if P and J are both convex, then P must be gracefully inscribed. In [**S**] we proved a result which is similar to Meyerson's Theorem but for rectangles and also a bit stronger.

Theorem 1.1 *For any Jordan loop J there is a connected subset β of gracefully inscribed rectangles such that all but at most 4 points of J are vertices of rectangles in β.*

The result is sharp, because a non-circular ellipse has 4 points which are not vertices of inscribed rectangles. One goal in this paper is to strengthen Meyerson's Theorem so that it looks more like Theorem 1.1.

Theorem 1.2 *For any Jordan loop J there is a connected subset β of gracefully inscribed equilateral triangles such that all but at most 2 vertices of J are vertices of equilateral triangles in β.*

The method to proof is similar to what we do in [**S**]. We approximate J by embedded polygons and take a limit of a structural result about the space of triangles inscribed in these polygonal approximations. The structural result works for any triangle shape, and we think that the techniques might eventually establish the same result as above for any shape λ, but we don't know how to do it yet.

We discuss what we can do along these lines. Let $G(J)$ denote the set of shape parameters λ for which the conclusion of theorem 1.2 is true. That is, there is a connected set β_λ of gracefully inscribed triangles of shape λ such that all but at most 2 points of J are vertices of triangles belonging to β_λ. We will describe our result in terms of angles rather than shape parameters. For $\theta \in (0, \pi/2)$ let $S(\theta)$ denote the set of shape parameters specifing triangles having an angle θ which is either the smallest angle or the largest angle (or both). Note that $S(\pi/3)$ just consists of the equilateral triangle shape. Every triangle shape belongs to at least one $S(\theta)$.

Theorem 1.3 *Let J be an arbitrary Jordan loop and let $\theta \in (0, \pi/2)$ be arbitrary. Then $G(J) \cap S(\theta) \neq \emptyset$.*

Theorem 1.3 contains Theorem 1.2 as a special case. Theorem 1.3 is, in turn, a special case of a stronger result, Theorem 2.12, a result whose statement we defer until §2.4.

Now we turn to our structural result. Let J be a polygon. We say that J is *angle-adapted* to λ if the minimum angle between any two edges of J exceeds the maximum angle of a triangle of shape λ. Let Ω_3 denote the set of ordered triples $(p_1, p_2, p_3) \in J^3$ consisting of distinct points. The space Ω_3 is a disjoint union of two open solid tori. Let $I(J, \lambda) \subset \Omega_3$ denote the set of triangles of shape λ that are inscribed in J. Assuming that $I(J,\lambda)$ is a 1-manifold, it makes sense to say whether or not a component of $I(J,\lambda)$ is essential – i.e. homologically nontrivial. Also, we say that a component is *graceful* if the corresponding triangles are gracefully embedded. The graceful components of $I(J,\lambda)$ lie in one component of Ω_3 and the ungraceful ones lie in the other.

Theorem 1.4 *Suppose that J is angle-adapted to λ and generically chosen. Then $I(J,\lambda)$ is a finite union of pairwise disjoint embedded polygons. Exactly one component of $I(J,\lambda)$ is essential. Moreover, the essential component is graceful.*

Theorem 1.4 is a quick corollary of a more precise result, Theorem 6.1, a result whose statement we defer until §6.1.

The proof of Theorem 1.4 (or equivalently of the sharper Theorem 6.1) is deceptively subtle. For this reason, it is worth pointing out why the obvious approach does not work. The obvious approach would be to start with some J_0 where the result is clear – e.g., a convex polygon – and then analyze the local changes to the space $I(J_t,\lambda)$, where J_t is a family of polygons that makes a generic interpolation between J_0 and some polygon J_1 of interest. A fairly easy argument shows that the overall homology class in $H_1(\Omega_3)$ does not change with t.

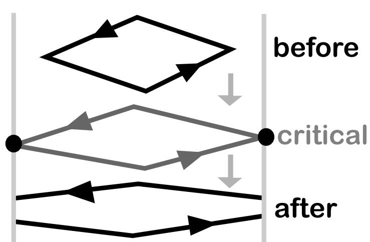

Figure 1.1: A locally allowed but globally bad topological transition.

The difficulty here is that *a priori* one inessential component could self-intersect at a critical point and turn into two oppositely oriented essential components. This would not change the picture homologically but it would change the number of essential components. Figure 1.1 shows what we mean. The picture takes place in a cylinder, with the vertical sides identified. (This cylinder is the projection of Ω_3 we get by taking the first two coordinates.)

The picture looks like an X at the critical point. Locally, this kind of topological transition is generic, but with just local information there does not seem to be a way to rule out the global change depicted.

We mention one last result. Say that a polygon J *supports an essential (un)graceful loop* if there is a polygonal loop in Ω_3 which represents a non-trivial class in $H_1(\Omega_3)$ and has the property that every associated inscribed triangle is (un) graceful. Unlike in the results above, the shape of the triangle is allowed to vary. We will prove the following result.

Theorem 1.5 *A polygon cannot support an essential ungraceful loop.*

Here is an outline of the rest of the paper. In §2 we give nice account of Meyerson's proof, partly for the sake of comparison and partly because one of his ideas, as explained in Lemma 2.2, inspired some of our constructions. Next, we deduce Theorems 1.2 and 1.3 from Theorem 1.4. (We separate out Theorem 1.2 for the sake of exposition.) The fact that there is just one essential component in Theorem 1.4 is crucial to the proofs of Theorems 1.2 and 1.3.

In §3 we prove Theorem 1.5. This argument is independent from the rest of the proofs in the paper.

In §4 we discuss what we mean by a *folding map*. This is an almost-everywhere non-singular piecewise linear map from the torus into the plane whose folding set – i.e. the edges over which the map reverses orientation – is a finite union of pairwise disjoint embedded polygonal loops. We bound on the number of pre-images of a Jordan loop such a map can have in terms of the number of essential components of its folding set.

In §5 we relate the problem of inscribing triangles to the idea of a folding map, then get a bound on the number of essential components of the associated folding set using a result about the fiber product.

In §6 we prove Theorem 6.1, the more precise version of Theorem 1.4. The results in §4-5 overcome the main difficulty in the proof, which is showing that $I(J,\lambda)$ has just one essential component. Theorem 1.5 takes care of the last statement of Theorem 6.1.

In §7 we take care of a technical detail left over from §2. This involves a brief study of how $I(J,\lambda)$ varies as λ varies.

I would like to thank Ramin Naimi, Igor Rivin, and Sergei Tabachnikov for discussions and questions about topics related to this paper.

2 Results for General Jordan Loops

2.1 Meyerson's Result

Here we give (essentially) Meyerson's proof that all but at most 2 points of a Jordan curve are vertices of an inscribed equilateral triangle. The proof we give departs somewhat from Meyerson's exact approach, but in spirit it is the same. We include this proof both because it is beautiful and because we will use the idea behind Lemma 2.2 later on.

Lemma 2.1 *Suppose that $p_0 \in J$ is some point, and there exist two other points p_1', p_2' in the region bounded by J such that $p_0p_1'p_2'$ is an equilateral triangle. Then J has an inscribed equilateral triangle with vertex p_0.*

Proof: Let B and U denote the bounded and unbounded components of $\boldsymbol{R}^2 - J$. Let ρ be the 60 degree rotation about p_0 such that $R(p_1') = p_2'$. Consider extending the rays p_0p_1' and p_0p_2' outward until they first hit J at points p_1, p_2. Without loss of generality p_0p_1 is not longer than p_0p_2. Hence $R(p_1) \in J \cup B$. Let q_1 be a point of J maximally far from p_0. We have $R(q_1) \in J \cup U$. So, by continuity there is some $r_1 \in J$ such that $R(r_1) \in J$. Our equilateral triangle has vertices $p_0, r_1, R(r_1)$. ♠

A *triod* is the image $T = h(Y)$ where Y is the letter Y and h is a homeomorphism into $\boldsymbol{R}^2$. Call the triod *good* if there is an equilateral triangle inscribed in the triod having one end of the triod as vertex, and otherwise *bad*. We call such triangles *end-inscribed triangles*. The key observation is that any 3 vertices of J are the endpoints of a triod that stays entirely in the region bounded by J. Just take one for the round disk and map it over.

Suppose for the moment that all triods are good. Choose any $a, b, c \in J$ and take a triod staying entirely inside J and having a, b, c as endpoints. Since this triod is good, there is an equilateral triangle inscribed in it having one of a, b, c as vertex, say a. But then the previous lemma applies to this triangle and shows that J has an inscribed equilateral triangle with a as vertex. So, to prove Meyerson's Theorem we just have to show that all triods are good.

We will prove that all triods are good in three steps: polygonal triods, end-straight triods, general triods. A triod is *polygonal* if it is the union of finitely many line segments.

Lemma 2.2 *A polygonal triod is good.*

Proof: Assume not, for the sake of contradiction. Let A denote the union of the first two legs of T. Let a be the endpoint of T not in A. For any $x \in T$ let A_x denote the result of rotating A by 60 degrees clockwise about x. When $x \in T - \partial A$, the we have $\partial A \cap A_x = \emptyset$ and $A \cap \partial A_x = \emptyset$. Otherwise we'd get the desired triangle. This means that the mod 2 intersection number I_x between A and A_x is well-defined and constant for all $x \in T - \partial A$.

Let b be an endpoint of A. The two arcs A and A_b meet only at b and make a 60 degree angle. So, by compactness, A_x and A cross exactly once, at x, for x sufficiently close to b. Hence $I_x = 1$ for all $x \in T - \partial A$. In particular, $I_a = 1$. But then we have an inscribed equilateral triangle with vertex a. ♠

A triod is *end straight* the triad is polygonal sufficiently near the ends.

Lemma 2.3 *An end straight triod is good.*

Proof: Let T be end-straight. We can approximate T by a sequence $\{T_n\}$ of polygonal triods having the same final segments. By the previous lemma, T_n has an end-inscribed equilateral triangle Δ_n. Not all points of Δ_n can be on the same final segment of T_n. Note also that $T_n \to T$ and T is embedded. Combining these two observations, we see that there is a uniform positive lower bound to the size of Δ_n. Hence we can take a Hausdorff limit and find the end-inscribed equilateral triangle on T. (In the next section we define Hausdorff limits.) ♠

Lemma 2.4 *An arbitrary triod is good.*

Proof: Now let T be an arbitrary triod, with ends a, b, c. For any large integer n, move out along the triple point of T until you reach the first point that is exactly $1/n$ from a. Call this point a'. Likewise define b', c'. Let T_n be the triod obtained by adding the segments aa', bb', cc' and erasing the arcs of T which join a to a', etc. If n is large enough, all points of $T' - (aa')$ are further than $1/n$ from a. Etc.

By construction T_n is end-straight. Let Δ_n be an end-inscribed triangle on T_n. Note that Δ_n cannot have a as a vertex, and another vertex on aa'. So, either Δ_n is inscribed in T, and we're done, or else (after relabeling) Δ_n has a as a vertex and one point in bb'. Letting $n \to \infty$ and taking a Hausdorff limit, we get an inscribed equilateral triangle with both a and b as vertices. ♠

2.2 Hausdorff Convergence

In this section we deduce Theorem 1.2 from Theorem 1.4. We begin with a discussion about Hausdorff convergence.

Let X be some metric space and let $\mathcal{X}$ denote the set of compact subsets of X. The *Hausdorff distance* between two elements A, B of $\mathcal{X}$ is the infimal $\epsilon > 0$ such that each point of A is within ϵ of B and *vice versa*. This notion of distance makes $\mathcal{X}$ into a compact metric space. Whenever we speak about convergence of compact subsets of another space, we always refer to convergence with respect to the Hausdorff metric.

Lemma 2.5 *Suppose that a sequence $\{A_n\}$ of compact connected arcs in X converges to a set A of X. Then A is connected.*

Proof: Suppose A is disconnected. There are disjoint open sets $U_1, U_2 \subset X$ whose union contains A such that each U_j intersects A nontrivially. Since A is compact, there is some positive $\epsilon > 0$ such that all points of $A \cap U_j$ are at least ϵ from $X - U_j$. In particular, each point of A is at least ϵ from each points in $X - U_1 - U_2$. For all sufficiently large n, the arc A_n intersects both U_1 and U_2 and therefore contains a point $x_n \in X - U_1 - U_2$. But then x_n is at least ϵ from A. Xhis contradicts the fact that $A_n \to A$ in the Hausdorff metric. ♠

2.3 Equilateral Triangles Revisited

Now let J be a Jordan loop. We think of J as the image $\alpha(\boldsymbol{R}/\boldsymbol{Z})$ where $\alpha : \boldsymbol{R}/\boldsymbol{Z} \to \boldsymbol{C}$ is continuous and injective.

Lemma 2.6 *The Jordan loop J can be approximated by a generic sequence of polygons that are angle-adapted to shape λ. More precisely, there is a sequence $\{\alpha_n\}$ of maps from $\boldsymbol{R}/\boldsymbol{Z}$ into $\boldsymbol{C}$ which converges uniformly to α such that the image of α_n satisfies the hypotheses of Theorem 1.4 with respect to λ.*

Proof: It is well known that any Jordan loop can be approximated by a sequence of polygons in the parametrized sense. See [**T**]. To arrange the needed angle condition, we can cut off the corners of the polygons in our sequence, repeatedly, and increasingly near the vertices, until all the interior angles are large enough. At the same time we modify the maps to reflect this change. We then make small generic perturbations. ♠

Let $J_n = \alpha_n(\boldsymbol{R}/\boldsymbol{Z})$. We take

$$\lambda = \frac{1 + i\sqrt{3}}{2}. \tag{1}$$

This is the equilateral triangle shape. Let $I_n = I(J_n, \lambda)$ denote the space of equilateral triangles inscribed in J_n. Let β_n be the unique essential component of I_n. Let ϵ_n denote the side length of the smallest triangle associated to β_n. If the sequence $\{\epsilon_n\}$ is uniformly bounded away from 0, then we can take a limit and find an equilateral triangle inscribed at each point of J whose side length is uniformly large.

Consider the case when $\epsilon_n \to 0$. Given a labeled equilateral triangle Δ inscribed in J_n, let Δ_{12} denote the side of Δ which connects vertex Δ_1 to vertex Δ_2. One of the two arcs of J_n contains the vertex Δ_3. Let $\mu(\Delta)$ denote the measure of this arc, according to the parametrization α_n. What we mean is that there is some interval $I \subset \boldsymbol{R}/\boldsymbol{Z}$ such that our arc is $\alpha_n(I)$, and $\mu(\Delta)$ is the length of I.

Lemma 2.7 *The range of μ on β_n converges to $(0,1)$ as $n \to \infty$.*

Proof: Here is the crucial part of the proof. The act of cyclically permuting the labels of a triangle acts on the space I_n. In particular, this action maps essential components to essential components. Since there is exactly one essential component, we see that β_n is invariant under cyclic relabeling.

The rest of the proof is continuity. Let τ_n be the smallest equilateral triangle associated to β_n. Let t_n be the 3-element subset of $\boldsymbol{R}/\boldsymbol{Z}$ such that $\alpha_n(t_n) = \tau_n$. Given that the side length of τ_n tends to 0 and that the parametrizations converge, the diameter of the smallest interval of $\boldsymbol{R}/\boldsymbol{Z}$ containing t_n tends to 0. But then $\mu(\tau_n)$ is either close to 0 or close to 1. In the former case, $\mu(\tau_n')$ is close to 1 for a suitable relabeling τ_n' of τ_n. In the latter case, $\mu(\tau_n')$ is close to 0 for a suitable cyclic relabelling τ_n' of τ_n. In either case, we find two triangles τ_n and τ_n' such that μ is close to 0 on one of them and close to 1 on the other. Since $\mu(\beta_n)$ is connected, this set achieves all values between these two extremes. ♠

Let $I_n = [1/n, 1 - 1/n]$. We re-index so that $I_n \subset \mu(\beta_n)$.

Lemma 2.8 *There are closed $\beta_n(3) \subset ... \subset \beta_n(n)$ such that μ maps $\beta_n(k)$ to I_k for $k = 3, ..., n$.*

Proof: We start with a point $\tau \in \beta_n$ such that $\mu(\tau) = 1/2$ and then we let $\beta_n(k)$ be the smallest arc of β_n containing τ such that μ maps $\beta_n(k)$ to I_k. These arcs automatically have the desired containment properties. ♠

Lemma 2.9 *For fixed k there is some uniform $\epsilon_k > 0$ such that each point of $\beta_n(k)$ corresponds to an equilateral triangle of diameter greater than ϵ_k.*

Proof: Let $\{\tau_n\}$ be a supposed equence of counter-examples. Let $\{t_n\}$ be the corresponding 3-element set of $\boldsymbol{R}/\boldsymbol{Z}$. The distance between the first 2 points of t_n tends to 0, forcing the side length of τ_n to 0. But then one of the sides of τ_n must subtend nearly the whole of J_n. This forces $\mu(\tau_n)$ either to 0 or to 1. ♠

Now we are ready to take a limit.

Lemma 2.10 *$I(J, \lambda)$ contains a connected set β such that $\mu(\beta) = (0, 1)$.*

Proof: Passing to a subsequence we can assume that, for each k, the sequence $\{\beta_n(k)\}$ is a Cauchy sequence in the Hausdorff metric. Let $\beta(k)$ denote the limit. The set $\beta(k)$ is connected, by Lemma 2.5. Moreover, each point of $\beta(k)$ corresponds to an equilateral triangle inscribed in J. Finally, $\mu(\beta(k)) = I_k$ by continuity. We have $\beta(3) \subset \beta(4) \subset \beta(5)$.... The nested union of connected sets is connected, so the union $\beta = \bigcup \beta(k)$ has the desired properties. ♠

Suppose that there are three points of J which are not vertices of equilateral triangles associated to β. These three points divide J into three intervals. Since β is connected, the jth vertex of any triangle associated to β lies in the same interval, independent of the triangle. If each interval contains a vertex, we get a positive lower bound to the value of μ on β. If some interval is empty, we get a finite upper bound. Either way, we have a contradiction. Note that we are taking limits of gracefully inscribed triangles, so the triangles associated to β are graceful.

2.4 The General Case

Now we prove Theorem 1.3. We first discuss the additional technical detail that we need for the proof. Let J be a polygon that is angle-adapted to two shape parameters λ_0 and λ_1. Let $\overline{\lambda_0\lambda_1}$ be the line segment connecting λ_0 to λ_1. Define

$$I(J, \lambda_1, \lambda_2) = \bigcup_{\lambda \in \overline{\lambda_0\lambda_1}} I(J, \lambda). \qquad (2)$$

Lemma 2.11 *$I(J, \lambda_0, \lambda_1)$ has a path connected subset B that contains the essential component of $I(J, \lambda_j)$ for $j = 0, 1$.*

Let $\boldsymbol{H}$ denote the upper half plane. Given a Jordan loop J, we let $G(J) \subset \boldsymbol{H}$ be as in Theorem 1.3, the set of parameters λ such that Theorem 1.2 holds true with respect to λ. The operation of cyclically relabeling changes the parameter λ to the parameter $\rho(\lambda) = (\lambda - 1)/\lambda$. The map ρ has order 3 and λ_0 is its unique fixed point in $\boldsymbol{H}$. The set $G(J)$ is automatically invariant under ρ. In this section we prove the following result. the following result.

Theorem 2.12 *Let J be an arbitrary Jordan loop and let $\lambda \in \boldsymbol{H}$ be arbitrary shape parameter. Any continuous path in $\boldsymbol{H}$, from λ to $\rho(\lambda)$, intersects $G(J)$.*

Proof of Theorem 1.3: By symmetry, Theorem 2.12 is also true with ρ^2 in place of ρ. Let λ be the parameter specifying an isosceles triangle with angles $\theta, \theta, \pi - 2\theta$. We can connect λ to one of $\rho(\lambda)$ or $\rho^2(\lambda)$ by a continuous path such that every point on the path lies in $S(\theta)$. ♠

Now we prove Theorem 2.12. We choose a sequence $\{J_n\}$ of polygons approximation the Jordan loop J. By compactness we can arrange that these polygons are angle-adapted to every parameter we consider.

Given a triangle τ inscribed in J_n, we let $\mu_k(n, \tau)$ denote the parametric measure of the arc of J_n subtended by the kth side of τ. Let

$$\mu_k(n, \lambda) = \sup_{\tau \in \beta_n} \mu_k(n, \tau). \qquad (3)$$

Here β_n is the essential component of $I(J_n, \lambda)$. Let S_k denote the set of parameters $\psi \in \boldsymbol{H}$ such that $\sup_n \mu_k(\psi, n) = 1$.

Lemma 2.13 *$G(J)$ contains all $\lambda \in \boldsymbol{H} - (S_1 \cup S_2 \cup S_3)$.*

Proof: For such parameters, there is a uniform lower bound to the diameter of a triangle of β_n. Here β_n is the essential component of $I(J_n, \lambda)$. The same argument as in the proof of Theorem 1.2 finishes the proof. ♠

We call a pair (λ_0, λ_1) *potent* if there are inequal indices $j \neq k$ such that $\lambda_0 \in S_j$ and $\lambda_1 \in S_k$.

Lemma 2.14 *If (λ_0, λ_1) is potent then any continuous path connecting λ_0 to λ_1 intersects $G(J)$.*

Proof: Call the path Γ as above. By the previous result, it suffices to consider the case when every point of Γ lies in $S_1 \cup S_2 \cup S_3$. If λ_2 lies midway between λ_0 and λ_1 in terms of the parametrization, then one of the pairs (λ_0, λ_2) or (λ_1, λ_2) is potent. Iterating this fact, we find a limit point $\gamma \in \Gamma$ and two infinite sequences of parameters $\{\lambda_{1,m}\}$ and $\{\lambda_{2,m}\}$ converging to γ such that $(\lambda_{1,m}, \lambda_{2,m})$ is potent for all m. Passing to a subsequence, and relableling, we can assume that $\lambda_{1,m} \in S_1$ for all m and $\lambda_{2,m} \in S_2$ for all m. We pass to a further subsequence so that

$$\mu_j(n, \lambda_{j,n}) > 1 - 1/n. \qquad (4)$$

Setting $\mu = \mu_1$, we now see that the restriction of μ to the path connected set

$$B_n = I(J_n, \lambda_{1,n}, \lambda_{2,n})$$

contains the interval $[1/n, 1 - 1/n]$. But then we can choose a path $\beta_n \subset B_n$ such that the restriction of μ to β_n has the same range. We have now set up the same conditions that Lemma 2.7 establishes in the equilateral case, except that the shapes of the triangles in β_n vary between the two nearby parameters $\lambda_{1,n}$ and $\lambda_{2,n}$. This change makes no difference to our limiting arguments, because these parameters converge to γ. The rest of the proof is as in the equilateral case. Hence $\gamma \in G(J)$. ♠

Now we return to the exact hypotheses of Theorem 2.12. If we happen to have $\lambda \in \boldsymbol{H} - S_1 - S_2 - S_3$, then $\lambda \in G(J)$ and Lemma 2.13 finishes the proof. Otherwise $\lambda \in S_k$ for some k. But then $\rho(\lambda) \in S_{k+1}$, with indices taken mod 3. But then $(\lambda, \rho(\lambda))$ is potent, and Lemma 2.14 finishes the proof.

3 Graceful Essential Loops

3.1 Main Result

In this chapter we prove Theorem 1.5. The proof comes down to the following two lemmas.

Lemma 3.1 (Existence) *Given a polygon J, there exists a polygon J' arbitrarily close to J in the Hausdorff metric which supports an essential graceful loop.*

Lemma 3.2 (Exclusion) *No polygon can support both an essential graceful loop and an essential ungraceful loop.*

Assume these results momentarily. If some polygon J supports an essential ungraceful loop, then so does every sufficiently nearby polygon, including one like J' from Lemma 3.1. But then J' supports both an essential ungraceful loop and an essential graceful loop. This contradicts Lemma 3.2. This completes the proof of Theorem 1.5.

3.2 Existence

In this section we prove Lemma 3.1. By adding extra vertices and an extra very short side if necessary, and rotating, we can find a polygon J' as close as we like to J with the following properties: J' has an edge e_1' in the x-axis and $J' - e_1'$ lies above the x-axis. Moreover, the two edges e_0' and e_2' adjacent to e_1' are shorter than e_1'. Label so that e_0', e_1', e_2' go left-to-right.

Let Ce_j' be the center of e_j'. Start with points a, b, c respectively at Ce_0', Ce_1', Ce_2', we do the following 5 steps, moving one vertex at a time monotonically counterclockwise around J'.

1. Move a to $e_0' \cap e_1'$. (Note: $\overline{ab}$ ends up in the x-axis.)
2. Move c all the way around J' to Ce_0'. (Note: c stays above the x-axis.)
3. Move b to $e_1' \cap e_2'$. (Note: $\overline{ab}$ stays in the x-axis.)
4. Move a to Ce_1'. (Note: $\overline{ab}$ stays in the x-axis.)
5. Move b to Ce_2'. (Note: c, a, b stay ordered left-to-right.)

This gives a continuous graceful family of triangles which does the permutation $a, b, c \to c, a, b$. Repeat this 3 times to get the desired loop.

3.3 Exclusion

In this section we prove Lemma 3.2.

Lemma 3.3 *Let Z be a closed cylinder and let $B \subset Z$ be a compact subset with finitely many connected components. If no connected component of B separates the two boundary components of Z, then neither does B.*

Proof: We can assume that $Z = (\boldsymbol{R}/Z) \times [0,1]$, the standard flat cylinder. Let $B = B_1 \cup ... \cup B_n$ be the decomposition into connected components. By compactness, we can slightly enlarge B without changing the separation status of any of the sets. So, without loss of generality, we can assume that each B_i is a finite union of squares. So, ∂B is a finite disjoint union of polygons, and ∂B_i is a finite disjoint union of polygons, and $\partial B = \bigcup \partial B_i$. Since B_i does not separate, all components of ∂B_i are trivial in $H_1(Z)$. But then all components of ∂B are trivial in $H_1(Z)$. But then B does not separate the boundary components of Z. ♠

Let Ω_3^+ denote the subset of Ω_3 consisting of triples (p_1, p_2, p_3) which appear in the counterclockwise order along J. Let Ω_3^- be the other component. All the essential graceful loops lie in Ω_3^+. Any essential ungraceful loop in Ω_3^- determines an *opposite loop* in Ω_3^+, which we get by mapping the point (p_1, p_2, p_3) to (p_3, p_2, p_1). Assuming we have both an essential graceful loop γ_1 and an essential ungraceful loop, γ_2', we let γ_2 be the opposite loop in Ω_3^+. Thus, both γ_1 and γ_2 are essential loops in Ω_3^+. For each of exposition, we will assume that our loops γ_1 and γ_2 both represent 1 in $H_1(\Omega_3^+)$. The general case is essentially the same, except that we base the construction on suitable multiples of the loops.

We say that an *extreme point* of J is one which cannot be the middle point of a triple $p_1, p_2, p_3 \in J$ of collinear points. Any vertex of the convex hull of J is an extreme point, and hence J has at least 3 such. Let v_1, v_2, v_3 be 3 extreme points. Let $\phi_j : \Omega_3^+ \to J$ be the map with action $\phi_j(p_1, p_2, p_3) = p_j$. We define

$$\Delta_j = \phi_j^{-1}(v_j). \tag{5}$$

Here Δ_j is a polygonal disk which intersects every essential loop in Ω_3^+.

Lemma 3.4 *There exists a cylinder $Z \subset \Omega_3^+$ such that $\partial Z = \gamma_1 \cup \gamma_2$ and for each $j = 1, 2, 3$ the set $Z \cap \Delta_j$ contains an arc δ_j connecting γ_1 and γ_2.*

Proof: The projection $\pi : \Omega_3^+ \to \Omega_2$ maps these loops to embedded loops in the annulus, and from this we conclude that our colored loops represent the element ± 1 in $H_1(\Omega_3^+) = \boldsymbol{Z}$. In particular, each γ_i intersects each Δ_j. This means that we can connect γ_1 to γ_2 by a polygonal arc $\delta_j \subset \Delta_j$.

Since $\Delta_i \cap \Delta_j$ is a polygonal arc for $i \neq j$, we can choose δ_1 so that it only intersects $\Delta_2 \cup \Delta_3$ finitely many times. Then we can choose δ_2 to that it avoids δ_1 and intersects Δ_3 in finitely many points. Finally we can choose δ_3 so that it avoids both δ_1 and δ_2. In short, we can choose these arcs to be pairwise disjoint. The union $\gamma_1 \cup \gamma_2 \cup \delta_1 \cup \delta_2 \cup \delta_3$ looks like a kind of cyclical ladder: It is also the union of 3 null-homologous polygonal loops, every two of which meet along an arc. We fill in these polygonal loops to disks. This gives us our cylinder Z. ♠

Each point of Ω_3^+ corresponds to a labeled triangle inscribed in J. Let B correspond to those triangles in which the points are collinear. The triangles corresponding to γ_1 are all counter-clockwise ordered, and the triangles corresponding to γ_2 are all clockwise ordered. Hence B separates γ_1 from γ_2.

Lemma 3.5 *A connected component β of $B \cap Z$ separates γ_1 from γ_2.*

Proof: There is a uniform positive lower bound to the distance $\|p_i - p_j\|$ when $(p_1, p_2, p_3) \in Z$ and $i \neq j$. From this we conclude that $B \cap Z$ is compact. The set B itself is a finite union of algebraic varieties and so $B \cap Z$ has only finitely many connected components. Since B separates γ_1 from γ_2, the intersection $B \cap Z$ must separate γ_1 from γ_2 in Z. By Lemma 3.3, some component β of B also separates γ_1 from γ_2 in Z. ♠

Since β is connected, there is some index $j \in \{1, 2, 3\}$ such that the restriction of ϕ_j to β always picks out the point of the triple that is between the other two – i.e., the middle point. Since $Z \cap \Delta_j$ contains an arc connecting γ_1 to γ_2, the set β must intersect Δ_j. But then $v_j \in \phi_j(\beta)$. We have shown that v_j is the middle point of a triple of collinear points in J. This contradicts the fact that v_j is an extreme point. This completes the proof of Lemma 3.2.

4 Folding Maps

4.1 Main Result

Let $\boldsymbol{T}$ be the square torus. Suppose that $\Delta : \boldsymbol{T} \to \boldsymbol{C}$ is a piecewise linear map that is almost-everywhere non-singular. So, $\boldsymbol{T}$ has a triangulation and the restriction of Δ to the interior of each triangle is an affine isomorphism onto its image. We define the folding set B to be the set of edges e in the triangulation of $\boldsymbol{T}$ such that Δ is orientation reversing on one side of e and orientation preserving on the other. Informally, the map Δ locally folds the domain over the folding set. We call Δ a *folding map* if B is a finite union of pairwise disjoint embedded polygonal loops.

Let Δ be a folding map. We say that a polygon $K \subset \boldsymbol{C}$ is *adapted* to Δ there is a primitive homology class $\Theta \in H_1(\boldsymbol{T})$ such that the following is true:

- $A = \Delta^{-1}(K)$ is a finite union of pairwise disjoint embedded polygonal loops, all transverse to B.
- When suitably oriented, each essential component of A represents the class Θ in $H_1(\boldsymbol{T})$. We call this the *positive orientation*.
- The restriction of Δ to each essential component of A, given the positive orientation, is a degree 1 map onto K.
- When suitably oriented, each essential component of B represent the class Θ in $H_1(\boldsymbol{T})$.

The purpose of this chapter is to prove the following result.

Lemma 4.1 (Component) *The number of essential A-components is at most the number of essential B-components.*

4.2 Proof in a Special Case

Here we prove the Component Lemma under the assumption that the essential components of A are disjoint from B.

We will assume that the conclusion of the Component Lemma is false and derive a contradiction. The essential components of A divide $\boldsymbol{T}$ into a number of annuli, one of which is shown at left in Figure 4.1. The B curves are drawn in black and the A curves are drawn in various shades of grey. Since the annuli boundaries are disjoint from B, each B component must be contained in some annulus. Since there are more annuli than B components, one of the annuli does not contain an essential B component. The annulus X shown in Figure 4.1 has this feature. We have colored $\boldsymbol{T}$ white or grey according as the map Δ is orientation preserving or reversing.

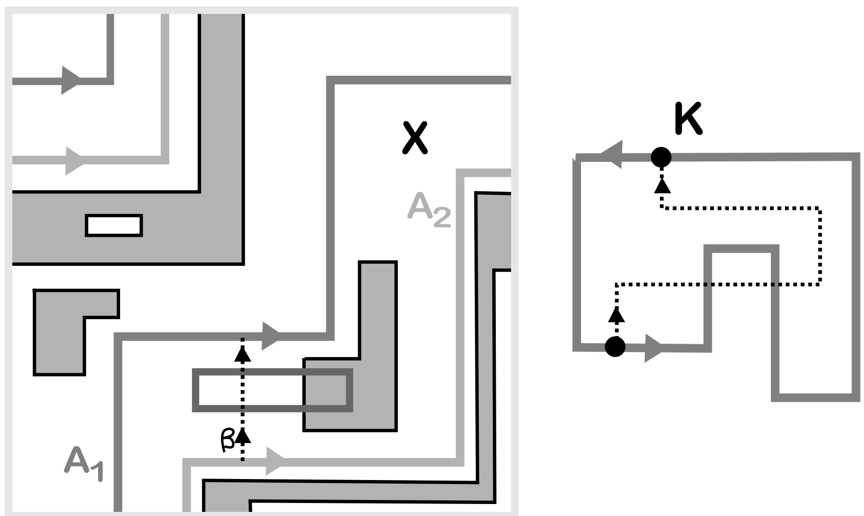

Figure 4.1: One of the annuli.

Given that X contains no essential component of B, one can connect the two boundary components of X by a polygonal path β that avoids B set and therefore remains entirely in a monochrome (white or grey) region. We have indicated β with a dotted path that remains in the white region.

Let $\gamma = \Delta(\beta)$. The curve γ starts out on K, moves away, crosses K an even number of times, then returns to K. These crossings come from the places where β crosses inessential components of $\Delta^{-1}(K)$. The right side of Figure 4.1 shows the situation. If β stays in the white (orientation preserving) region, then $\Delta(\beta)$ must start by moving into the bounded region of $\boldsymbol{C} - K$, as shown in Figure 4.2. In other words, we can say that $\Delta(\beta)$ starts out by moving *inside*. Counting the crossings, we see that γ must reach the second endpoint on K from the inside. That is, all points of γ sufficiently near the second endpoint must also be in the bounded component of $\boldsymbol{C} - K$. But this is a contradiction: The map Δ would have to be orientation reversing at the second endpoint of β. A similar argument with *outside* replacing *inside*.

4.3 Proof in the General Case

Now we consider the general case. The proof goes by induction on the number of intersections between the essential components of A and the components of B. The special case above is the base case for the induction. For the inductive step, the idea is to find an "innermost intersection" and modify Δ near the relevant segment of A so as to eliminate 2 intersection points and retain the number of essential A and B components and the general nature of the map.

Let a be an essential component of a. The restriction $\Delta : a \to K$ is a degree 1 map, but it need not be monotone. The map is monotone on the intervals bounded by the finitely many points of $a \cap B$. There are an even number $a_1, ..., a_n$ of such intervals. The restriction $\Delta|a_i$ is monotone, and winds $4w_i$ units around K, which we normalize below to have length 4. The number w_i is positive iff $\Delta(a_i)$ winds counter-clockwise around K. The only constraint we have is that $w_1 + ... + w_n = 1$ and the signs alternate.

To make our construction cleaner, we pre-compose Δ with a piecewise linear homeomorphism so that the essential components of A are geodesics, and we post-compose by a piecewise linear homeomorphism so that K is the unit square centered at the origin.

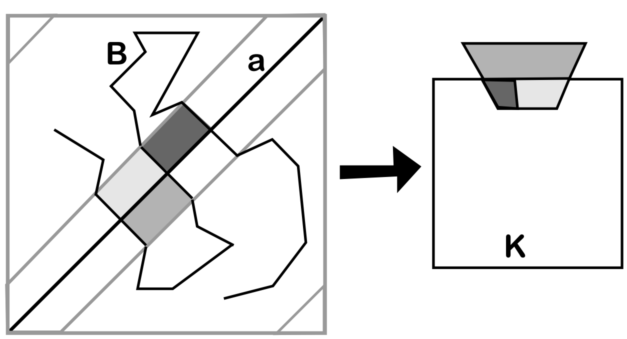

Figure 4.2: The restriction of Δ to the thin strip σ.

The transversality condition guarantees that we make small modifications to Δ near a so that it has a product structure in the sense that

$$(\Delta \circ t_\epsilon)|_a = (d_{1+\epsilon} \circ \Delta)|_a. \tag{6}$$

Here t_ϵ is translation by ϵ normal to a and $d_{1+\epsilon}$ is dilation by $1+\epsilon$ about the origin. The cleanest way to do this is to slice $\boldsymbol{T}$ open along a and graft in the desired product map. Figure 4.2 indicates the action of Δ on a few of the pieces in a small geodesic strip σ with centerline a. (On the right, the dark piece is folded over the light one.)

We now explain a modification of Δ where we change Δ inside σ. Let σ_+ and σ_- be the two components of $\sigma - a$. Let $\Delta_\pm$ be the restriction of Δ to $\partial\sigma_\pm$. We first define a new piecewise linear and degree 1 map $f : a \to K$. We then take a new map Δ_f in $\sigma_\pm$ so that it maps $t_\epsilon(a)$ to $d_{1+\epsilon}(K)$ and interpolates between f and $\Delta_\pm$. That is, we use σ to implement homotopies between f and $\Delta_\pm$. Finally, we set $\Delta_f = \Delta$ outside σ. By construction, Δ_f is a fold map and $A - A_f = \Delta_f^{-1}(K)$. The folding set B_f agrees with B outside σ but inside σ the two sets B and B_f might be wildly different.

Our first trick is to choose f so that it has the same fold points as Δ_a. All that we change are the values of the numbers $w_1, ..., w_n$. We can then find piecewise linear homotopies which result in $B_f = B$. We just change the amounts that the various arcs wind around in a piecewise linear way. Indeed, we can choose f to have an arbitrary sequence $w_1', ..., w_n'$ such that $\text{sign}(w_i) = \text{sign}(w_i')$ and $w_1' + ... + w_n' = 1$.

Our second trick requires some preparation. Since all the essential loops are in the same homology class, we can find a digon D whose interior is disjoint from $A \cup B$ and whose boundary has one arc in an essential component a of A and the other edge in B, as shown in Figure 4.3.

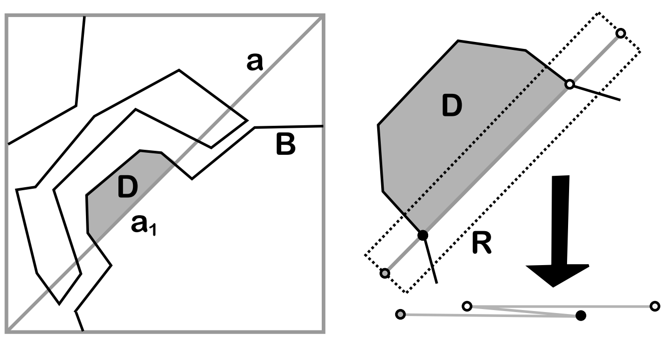

Figure 4.3: An innermost digon and the region of modification.

Let a_1 be the interval of a which is the boundary component of D. We apply our first trick so that $|w_1| < 1/2$ and $|w_1| < |w_i|/4$ for all $i \neq 1$. (We then modify to regain the product structure near a.) Once we do this, we can find a rectangular neighborhood R of a_1, as shown on the right side of Figure 4.3, so that the $\Delta(a_1) \subset \Delta(R \cap a)$, and $R \cap B$ is a union of two segments which cut cross R. The rightmost part of Figure 4.3 indicates the action of Δ on $a \cap R$.

If we take R very thin, then R lies in the kind of strip σ discussed above, with the top and bottom boundarties of R respectively lying in the top and bottom boundaries of σ. We now do the first trick again. This time we make the map $f : a \to K$ agree with $\Delta|_a$ outside R and be monotone in $a \cap R$. The conditions on the winding numbers makes this possible. The point is that the restriction of Δ to $a \cap R$ starts out moving a long way in one direction, doubles back a little bit when it hits the left endpoint of a_1, then doubles back again and moves a long way in the initial direction once it hits the right endpoint of a_1.

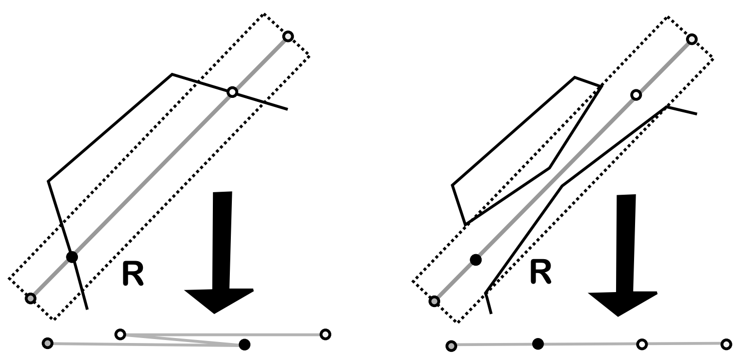

Figure 4.4: Changing the fold lines.

We construct Δ_f using homotopies which are the identity outside R. These homotopies just undo the little fold in the most straightforward way. Figure 4.4 indicates the modification. The bottom part of Figure 4.4 indicates the action of the two maps Δ and f on $a \cap R$. As indicated in Figure 4.4, this operation creates one more inessential B-component but does not change the number of essential B-components. Moreover, it decreases the number of intersection points by 2. This gives a new counter-example which, by induction, cannot exist. This contradiction establishes the general case of the result.

5 Folding Maps and Polygons

5.1 Main Result

We call a polygon *obtuse* if it is angle-adapted to i. This is to say that every angle between consecutive edges is obtuse. Suppose, for $j = 1, 2$, that

$$\alpha_j : \boldsymbol{R}/\boldsymbol{Z} \to J_j \subset \boldsymbol{C} \tag{7}$$

is a piecewise linear parametrization of an embedded obtuse polygonal loop J_j.

Recall that $\boldsymbol{T} = (\boldsymbol{R}/\boldsymbol{Z})^2$. We introduce the *difference map* $\Delta : \boldsymbol{T} \to \boldsymbol{C}$. The map is

$$\Delta(s_1, s_2) = \alpha_1(s_1) - \alpha_2(s_2). \tag{8}$$

Lemma 5.1 (Folding Lemma) *Suppose that both J_1 and J_2 are obtuse and no side of J_1 is parallel to a side of J_2. Then Δ is a folding map and the fold set B has precisely 2 essential components. When suitable oriented, these essential components each represent $(1,1)$ in $\boldsymbol{H}_1(\boldsymbol{T})$.*

Our proof of the Folding Lemma relies on a certain result about fiber products.

We call the map $f : \boldsymbol{R}/\boldsymbol{Z} \to \boldsymbol{R}/\boldsymbol{Z}$ a *nice map* if f has degree 1, is piecewise linear, and is not constant on any interval. We call the points where f locally reverses direction the *fold points*. We call $t = f(s)$ a *fold value* if s is a fold point for f. Given two nice maps $f_1, f_2 : \boldsymbol{R}/\boldsymbol{Z} \to \boldsymbol{R}/\boldsymbol{Z}$ we can form the *fiber product*

$$H(f_1, f_2) = \{(s_1, s_2) \in \boldsymbol{T} |\ f_1(s_1) = f_2(s_2)\}. \tag{9}$$

We call two nice maps f_1 and f_2 *unrelated* if they have no common fold values. For completeness, we include in this chapter a proof of the following (possibly well-known) result.

Lemma 5.2 (Fiber Product) *Suppose that f_1 and f_2 are unrelated nice maps. Then $H(f_2, f_2)$ is a polygonal 1-manifold which has exactly one connected component that is homologically nontrivial in $\boldsymbol{T}$. When suitably oriented, the one nontrivial component represents $(1,1)$ in homology $H_1(\boldsymbol{T})$.*

5.2 Proof of the Folding Lemma

We keep the notation from above.

Lemma 5.3 *B is a finite union of pairwise disjoint embedded polygonal loops.*

Proof: If $(s_1, s_2) \in B$, then at least one of the two points $\alpha_1(s_1)$ or $\alpha_2(s_1)$ is a vertex. This follows from the no-parallel-sides assumption on J_1 and J_2. Call $(s_1, s_2) \in B$ *special* if both $\alpha_1(s_1)$ and $\beta_2(s_1)$ are vertices. Otherwise call $(s_1, s_2) \in B$ *ordinary*.

Suppose that (s_1, s_2) is ordinary, and $\alpha_1(s_1)$ is a vertex of J_1. There is a unique edge e of J_2 such that $\alpha_2(s_2) \in e$. But then the whole vertical segment $s \times \alpha_2^{-1}(e)$ lies in B. Likewise if $\alpha_1(s_1)$ is not a vertex of J_1 and $\alpha_2(s_2)$ is a vertex, then B contains a horizontal segment. This shows that B is a finite union of closed horizontal and vertical line segments together with perhaps a finite union of isolated points. The potential isolated points correspond to the special points.

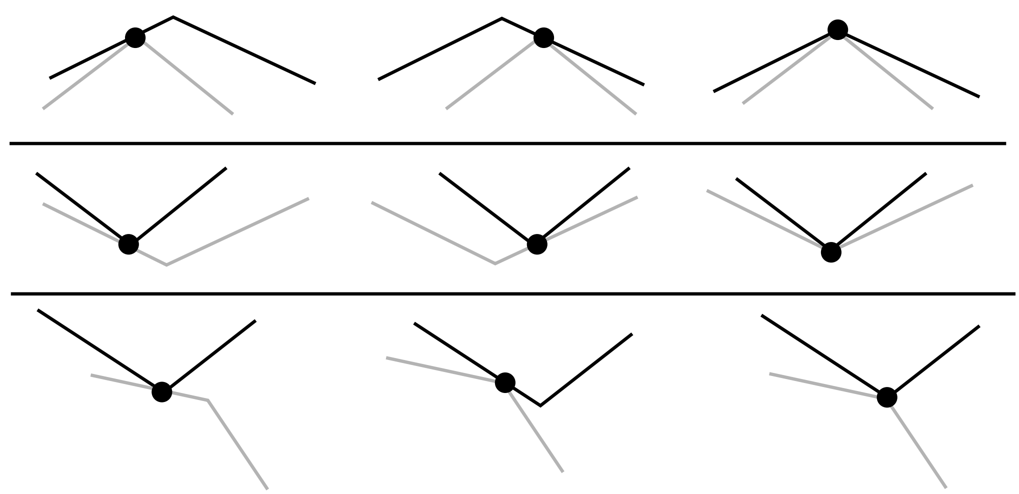

Figure 5.1: Variations near the special points.

Suppose that (s_1, s_2) is a special point. We define the *relevant angle* at J_1 to be the angle at the vertex $\alpha_1(s_1)$ on the side that locally contains $J_2 + \Delta(s_1, s_2)$. Likewise we define the relevant angle of J_2. If the relevant angle of J_1 exceeds π, then in a neighborhood of (s_1, s_2), the set B is a union of 2 horizontal segments. If the relevant angle of J_2 exceeds π, then in a neighborhood of (s_1, s_2), the set B is a union of 2 vertical segments. If the relevant angles are less less than π then the obtuse angle condition guarantees that in a neighborhood of (s_1, s_2) the set B is a union of one horizontal edge and one vertical edge. Figure 5.1 shows the 3 cases and indicates the variation around (s_1, s_2).

This analysis shows that B has no isolated points, and then every special point is contained in exactly 2 closed segments of B. Hence B is a finite union of pairwise disjoint embedded polygonal loops. ♠

Lemma 5.4 *B has exactly 2 components which are nontrivial in $\boldsymbol{H}_1(\boldsymbol{T})$. When they are suitably oriented, these components each represent $(1,1)$ in $\boldsymbol{H}_1(\boldsymbol{T})$.*

Proof: Looking at the orientations, at each point $(s_1, s_2) \in B$, we can ask whether the two curves are moving in the same direction or in the opposite direction at the relevant intersection point. We write $B = B_+ \cup B_-$ according to the two possibilities. This is a partition of the components of B into two types. We will show that B_+ has exactly one component that is nontrivial in $H_1(\boldsymbol{T})$, and this component represents $(1,1)$ when when suitably oriented. The argument for B_- is very similar.

Morally, the proof works like this: We compose α_j with the unit tangent map to get a nice degree 1 map f_j from $\boldsymbol{R}/\boldsymbol{Z}$ to the unit circle, which we identify with $\boldsymbol{R}/\boldsymbol{Z}$. We then recognize B_+ as the fiber product of f_1, f_2 and we finish the proof by quoting the Fiber Product Lemma. Since we tangent map is not quite well defined for polygons, we have to use an approximation argument.

For each positive integer n, we approximate the map α_j by a piecewise circular map $\alpha_{j,n}$ which stays within $1/n$ of α_j in the sup-norm. Even though we may not be able to arrange that $\alpha_{j,n}$ is everwhere differentiable, we can arrange that the image $J_{j,n}$ is continuously differentiable. In other words, the tangent map is not necessarily continuous, but the unit tangent map is continuous and indeed piecewise linear. To make the approximation, we replace the edges of J_j by circular arcs of very small curvature and the vertices of J_j by circular arcs of very large curvature, and we make sure that the tangent vectors match at the endpoints. We then take a map which has constant speed on each circular arc and we adjust the speeds to get the close sup-norm approximation.

The unit tangent map U is well defined for $J_{j,n}$. We let $f_{j,n} = U \circ \alpha_{j,n}$. For any $s \in \boldsymbol{R}/\boldsymbol{Z}$, the point $\alpha_{j,n}(s)$ converges to $\alpha_j(s)$ as $n \to \infty$. If $\alpha_j(s)$ lies in the interior of an edge e of J_j then $f_{j,n}(s)$ converges to the counter-clockwise unit vector parallel to e. If $\alpha_j(s)$ is a vertex v of J_j then $f_{j,n}(s)$ converges on a subsequence to a unit vector tangent to a line through v which does not locally separate J_j.

From this description, the fiber product $B_{+,n} = H(f_{1,n}, f_{2,n})$ converges to the set B_+ in the Hausdorff metric as $n \to \infty$. The convergence is stronger than that: If $\{p_n\}$ and $\{q_n\}$ are two sequences of points in $B_{+,n}$ that converge to the same point of B_+, then some arc in $B_{+,n}$ of length ℓ_n connects p_n to q_n, and $\ell_n \to 0$. Call this the *short arc property*. The short arc property is local and the proof only involves inspecting what happens for pairs of consecutive edges of J_1 and J_2.

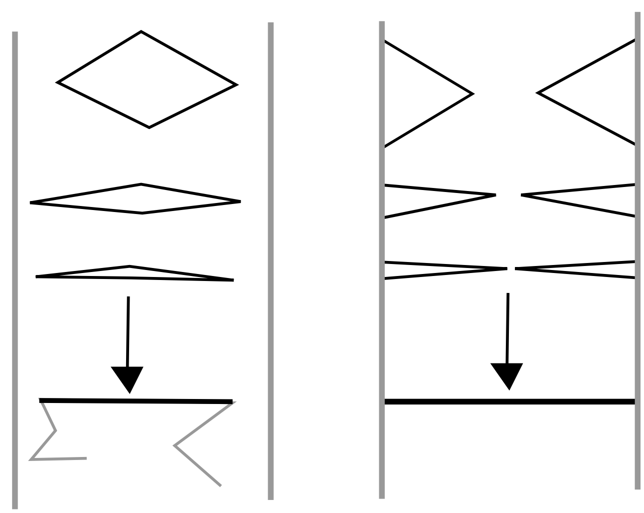

Figure 5.2: Some bad Hausdorff convergence

The short arc property rules out the kinds of Hausdorff convergence depicted in Figure 5.2. The pictures are meant to take place inside a cylinder, so the vertical sides in each half of the picture are identified. On the left, a sequence of inessential loops converges to a proper arc of some loop. On the right, a sequence of inessential loops converges to an essential loop. This kind of convergence involves a kind of "folding over" which brings together points that are not joined by short arcs.

By the Fiber Product Lemma, $B_{+,n}$ has one essential component β_n. The sequence $\{\beta_n\}$ converges on a subsequence to a subset of some component

β of B_+. The component β must be essential, because otherwise for large n the essential loop β_n would fail to intersect some representative of (say) $(1,-1)$ in $H_1(\boldsymbol{T})$ that avoids β. Since β_n represents $(1,1)$ in $H_1(\boldsymbol{T})$ for all n, the component β, when suitably oriented, must represent $(1,1)$ as well.

Let β' be any other component of B_+. Since there is a uniform upper bound on the number of components of $B_{+,n}$ (in terms of the number of sides of the polygons) and since all of β' is contained in the Hausdorff limit of $B_{+,n}$, there is some sequence $\{\beta'_n\}$ of components converges to an uncountable closed subset β'' of β'. By the Fiber Product Lemma, β'_n is inessential because $\beta'_n \neq \beta_n$. Either $\beta'' = \beta'$ or β'' is a proper arc. In the latter case, we have the first kind of convergence depicted in Figure 5.2. This is impossible, so $\beta'' = \beta'$. If β' is essential, we have the second kind of convergence depicted in Figure 5.3. This is impossible, so β' is inessential. ♠

5.3 Proof of the Fiber Product Lemma

We will break the proof into 4 steps. Let $H = H(f_1, f_2)$ be the fiber product of f_1 and f_2.

Lemma 5.5 *H is a polygonal 1-manifold.*

Proof: Given two partitions $\{I_i\}$ and $\{J_j\}$ of $\boldsymbol{R}/\boldsymbol{Z}$ into intervals, we can take the product and get a partition of $\boldsymbol{T}$ into rectangles $\{R_{ij}\}$ with $R_{ij} = I_i \times J_j$. Since f_1 and f_2 are unrelated, we can choose these partitions so that the restriction of each function to each interval is linear and injective, and no vertex of an R_{ij} belongs to H. The locations of the fold points force us to choose certain breaks in the partitions, but otherwise we choose the breaks generically.

Let $H_{ij} = H \cap R_{ij}$. By construction H_{ij} is either the emptyset or a line segment which connects the interior point of some edge of R_{ij} to the interior point of some other edge of R_{ij}. Consider the picture around an endpoint p of H_{ij}. Let R' be the rectangle adjacent to R_{ij} across the edge containing p. Since $H \cap R'$ is not the emptyset, $H \cap R'$ has the structure just mentioned. In particular, H_{ij} meets a unique line segment of H at p. This shows that H is a polygonal 1-manifold. ♠

Lemma 5.6 *H has an orientation with the following properties:*

- *Whenever the generic vertical geodesic $x = x_0$ intersects H at a point (x_0, y), the relevant segment points to the right if and only if $f_2'(y) > 0$.*
- *Whenever the generic horizontal geodesic $y = y_0$ intersects H at a point (x, y_0), the relevant segment points to the top if and only if $f_1'(x) > 0$.*

Proof: Here is the construction. If $f_1(I_j)$ and $f_2(J_j)$ are not disjoint, then they overlap in one of 4 possible ways. At the same time, there are 4 possible orientations for these segments, depending on the signs of the derivatives f_1' and f_2'. All in all, there are 16 different possibilities. For each of these possibilities, we choose an orientation for the corresponding segment of H, according to the scheme shown in Figure 5.3.

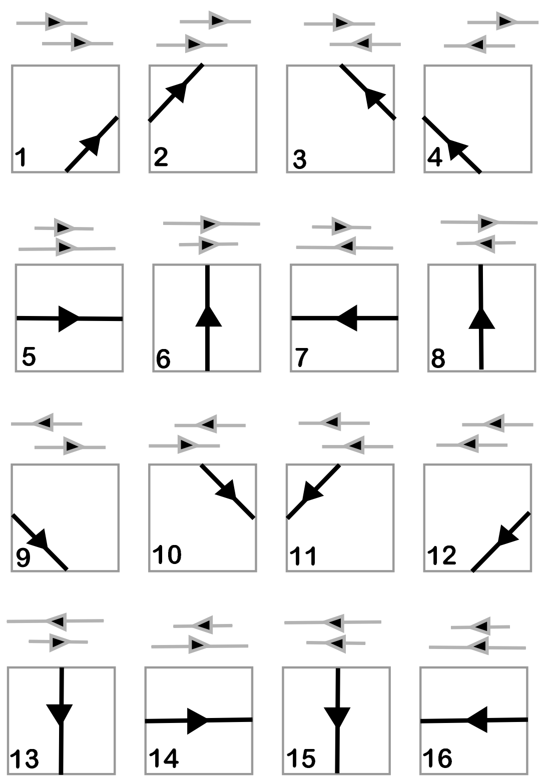

Figure 5.3: The orientation on the fiber product

A case-by-case check shows that this scheme defines a consistent orientation on H. Figure 5.4 shows how Cases 1,2 fit together and how Cases 1,3 fit together.

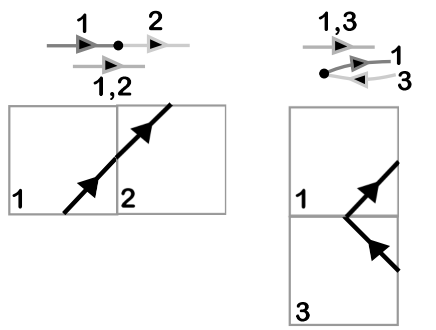

Figure 5.4: Adjacent pairs of segments and their tiles.

Now we check how the geodesic $x = x_0$. intersects one of our tiles. In all cases, the arrow in Figure 5.3 points to the right if and only if the lower of the two segments (corresponding to $f_2(J_j)$ points to the right. Similarly, we check how the geodesic $y = y_0$ intersects our tiles. In all cases, the arrow in Figure 5.3 points up if and only if the upper of the two segments (corresponding to $f_1(I_i)$) points to the right. ♠

From now on we equip H with the orientation given above, and we call it the *natural orientation*. Since H is oriented, it makes sense to ask which homology class H represents in $H_1(\boldsymbol{T})$.

Lemma 5.7 *H represents the element $(1, 1)$ in $H_1(\boldsymbol{T})$.*

Proof: If suffices to show that the geodesics $x = x_0$ amd $y = y_0$ each intersect H once, counting the orientations. Consider $x = x_0$. Each intersection point with this geodesic corresponds to a parameter value y where $f_2(y) = f(x_0)$. The orientation points to the right if and only if $f_2'(y) > 0$. But the number of times $f_2'(y) > 0$ is one more than the number of times $f_2'(y) < 0$ because f_2 has degree 1. In other words, $f_2(\boldsymbol{R}/\boldsymbol{Z})$ crosses a point righwards one more time than it crosses leftwards. This proves our claim for the geodesic $x = x_0$. A similar argument works for the geodesic $y = y_0$. ♠

Now we know that H represents $(1, 1)$ in $H_1(\boldsymbol{T})$. Two distinct and nontrivial homology classes in $\boldsymbol{T}$ intersect unless they represent the same homology classes or their sum is 0 in homology. Since H is an embedded 1-manifold, all the homologically nontrivial components of H represent either $(1, 1)$ or $(-1, -1)$. Moreover, the number of $(1, 1)$ representatives is one more than the number of $(-1, -1)$ representatives. The last step finishes the proof.

Lemma 5.8 *An arbitrary non-trivial component h of H represents $(1,1)$ in homology.*

We can find a piecewise linear map $a : \boldsymbol{R}/\boldsymbol{Z} \to \boldsymbol{T}$ such that $a = (a_1, a_2)$ parametrizes h, and each a_j is a degree 1 map. Define $b = f_j \circ a_1$. This map is independent of j and has degree 1.

The parametrization a gives h a second orientation of h which we call the *forced orientation*. The component h represents the element $(1,1)$ in $H_1(\boldsymbol{T})$ with respect to the forced orientation. So, to finish the proof, we need to show that the forced orientation and the natural orientation coincide.

Given $t \in \boldsymbol{R}/\boldsymbol{Z}$ we can compare the signs of $f_2'(a_2(t))$ and $a_1'(t)$. The former quantity determines the direction that h points across the vertical line through $a(t)$. The latter quantity determines the direction that h points across the vertical line through $a(t)$. The two orientations agree iff the two quantities have the same sign. By the Chain Rule, $f_2'(a_2(t))$ is positive if and only if $a_2'(t)$ and $b'(t)$ have the same sign. Therefore the two orientations agree if there is any point t such that

$$a_1'(t)a_2'(t)b'(t) > 0 \tag{10}$$

Note that $a_j(s) = a_j(t)$ implies that $b(s) = b(t)$. This is because $b = f_j \circ a_j$.

We will suppose that Equation 10 fails for all t and we will derive a contradiction. We can find lifts $A_1, A_2, B : \boldsymbol{R} \to \boldsymbol{R}$ of a_1, a_2, b respectively. Each function F is such that $F(x+1) = F(x) + 1$. The lifted functions also satisfy the same property as above: If $A_j(s) = A_j(t)$ then $B(s) = B(t)$. Moreover $A_j' = a_j'$ and $B' = b'$. So, $A_1'(t)A_2'(t)B'(t) < 0$ whenever all these derivatives are defined. In particular, these derivatives cannot all be positive.

Say that a point $t \in \boldsymbol{R}$ is a *peak* if the function $B(t) - t$ has a global maximum at t. A peak exists because the function $B(t) - t$ is periodic. Let t_0 be a peak. By construction, $B(s) < B(t_0)$ for all $s < t_0$. For $\epsilon > 0$ sufficiently small, we have $B'(t_0 - \epsilon) \geq 1 > 0$. We pick ϵ so small that no derivative changes sign on $[t_0 - \epsilon, t_0]$. Since not all derivatives are positive, have $A_j'(t_0 - \epsilon) < 0$ for some j. By the Fundamental Theorem of Calculus, $A_j(t_0 - \epsilon) > A_j(t_0)$. Since $A_j(t_0 - 1) < A_j(t_0)$ there is some $s \in (t_0 - 1, t_0)$ such that $A_j(s) = A_j(t_0)$. But then $B(s) = B(t_0)$. This is a contradiction. ♠

6 Spaces of Inscribed Triangles

6.1 The Main Result

Say that a λ-*triangle* is a triangle having shape λ. This means that it is similar (in an orientation preserving way) to the triangle with vertices $0, 1, \lambda$.

A λ-triangle defines 3 similarities: S_k rotates side $k+1$ to side $k-1$, with indices taken mod 3. Let $L = (L_1, L_2, L_3)$ be the triple of lines. We say that L is *adapted* to λ if there is at least one index k such that $S_k(L_{k+1})$ is not parallel to L_{k-1}. This definition only depends on λ and not on the representative triangle. Within the 6-dimensional space of choices for L, the set of L that are not adapted to λ has codimension 2.

Now let J be a polygon, with edges $E_1, ..., E_n$ and lines $L_1, ..., L_n$ extending these edges. We say that a triple (i, j, k) of indices is *bad* if $E_1 \cap E_2 \cap E_3$ is nonempty. Otherwise we call the triple *good*. We say that J is *adapted* to λ if

1. J is angle-adapted to triangles of shape λ. (See Theorem 1.4.)
2. For each good triple (i, j, k), the triple (L_i, L_j, L_k) is adapted to λ.
3. At most one vertex of any J-inscribed λ-triangle is a vertex of J.

Recall that $I(J, \lambda) \subset \Omega_3$ is the space of triangles inscribed in J.

Theorem 6.1 *Suppose that J is adapted to λ. Then $I(J, \lambda)$ is a finite union of pairwise disjoint embedded polygons. Exactly one component of $I(J, \lambda)$ is essential. Moreover, the essential component is graceful.*

6.2 Manifold Structure

In this section, we explain why the local conditions above guarantee that $I(J, \lambda)$ is a finite union of pairwise disjoint polygons. Before we start, we mention that the angle condition is necessary. If J has some angles that are smaller than an angle of a λ-triangle, then $I(J, \lambda)$ contains points corresponding to arbitrarily small triangles. In this case the space isn't even compact.

Let $E = (E_1, E_2, E_3)$ be a triple of segments and let $L = (L_1, L_2, L_3)$ be the triple of lines extending them. We say that a λ-triangle is *inscribed* in E

(respectively L) if the kth vertex of the triangle lies in L_k (respectively E_k) for $k = 1, 2, 3$. We allow the degenerate possibility that the the triangle is just 3 identical points. Let $I(L, \lambda)$ denote the space of triangles of shape λ inscribed in L. Likewise define $I(L, E)$. We say that a *brick* is the product of 3 segments.

Lemma 6.2 *Suppose that L is adapted to λ. Then the set $I(L, \lambda)$ is a straight line. Hence $I(E, \lambda)$ is the intersection of a straight line and the brick $E_1 \times E_2 \times E_3$.*

Proof: After re-labeling, suppose that $S_1(L_2)$ and L_3 are not parallel. here S_1 is the first similarity associated to some triangle of shape λ. We pick p_1 and adjust our choice of triangle of shape λ so that $S_1(p_1) = p_1$. Then there are unique points $p_2 \in L_2$ and $p_2 \in L_3$ such that p_1, p_2, p_3 is inscribed in L, namely $p_3 = S_1(L_2) \cap L_3$ and $p_2 = S_1^{-1}(p_3)$. The point p_2, p_3 vary linearly with p_1. This constructs a line of inscribed triangles, and every other inscribed triangle is among the ones we constructed. ♠

Lemma 6.3 *If J is adapted to λ then $I(J, \lambda)$ is a finite union of pairwise disjoint embedded polygons in $\boldsymbol{C}^3$.*

Proof: The product $J \times J \times J \subset \boldsymbol{C}^3$ has a decomposition into bricks. Each brick is the triple product of edges of J. We call a brick in this decomposition *bad* if it corresponds to a bad triple of edges, and otherwise *good*. There are at most 2 edges involved in a bad triple, and in this case the edges are consecutive. The angle condition on J guarantees that $I(J, \lambda)$ is disjoint from all the bad bricks.

Let B be a good brick that intersects $I(J, \lambda)$ and let $b = B \cap I(J, \lambda)$. By Lemma 6.2, the set b is a line segment, the intersection of a straight line with B. Since J has no singularly inscribed triangles of shape λ, the endpoints of b must lie in the interiors of the facets of B. (A *facet* of a brick is one of the 6 two-dimensional faces.)

Let b_0 be one of the endpoints of b. There is a unique second brick B' such that $b_0 \in B'$. The brick B' is also good because it intersects $I(J, \lambda)$. But then $b' = B' \cap I(J, \lambda)$ is also a line segment that has b_0 as an endpoint. This analysis shows that $I(J, \lambda)$ is an embedded graph with straight line edges and all vertex degrees equal to 2. ♠

6.3 The Parity Argument

We continue with the assumption that J is adapted to λ. We know from Lemma 6.3 that $I(J,\lambda)$ is a finite union of polygons. We just have to establish the other two statements. We start by showing that $I(J,\lambda)$ has an odd number of essential components.

Lemma 6.4 *$I(J,\lambda)$ has an odd number of essential components.*

Proof: Let $\Omega_3^{\pm}$ be the two components of Ω_3. Let Ω_2 be the set of ordered and unequal pairs $(p_1,p_2) \in J \times J$. Let

$$\pi : \Omega_3 \to \Omega_2$$

be the map which forgets the third point. The map π induces an isomorphism from $H_1(\Omega_3^{\pm}) \to H_1(\Omega_2)$ and moreover π is injective on $I = I(J,\lambda)$. Here The reason here is that the first two points of a triangle of shape λ determine the whole triangle.

We compactify Ω_2 by adding two boundary components. The points near one component have the form (p_1,p_2) where p_2 immediately follows p_1 in the cyclic order of J, and the points near the other component have the reverse property. Let $\overline{\Omega}_2$ denote this compactification. Note that an embedded polygon in Ω_2 is homologically nontrivial if and only if it separates the two boundary components of $\overline{\Omega}_2$.

We partition Ω_2 into two sets as follows: Given $(p_1,p_2) \in \Omega_2$ we choose $p_3 \in \boldsymbol{C}$ so that (p_1,p_2,p_3) is a triangle in C_λ. We color (p_1,p_2) red (respectively blue) if p_3 lies in the unbounded (respectively bounded) component of $\boldsymbol{C} - J$. With this scheme, Ω_2 is partitioned into the red points, the blue points, and the points of $\pi(I)$.

Given that J is adapted to λ, all points sufficiently near one boundary component of $\overline{\Omega}_2$ have one color and all points sufficiently near the other boundary component have the other color. A generic polygonal path connecting one component to the other must therefore intersect $\pi(I)$. But such a path intersects each inessential component an even number of times and each essential component an odd number of times. Hence there are an odd number of essential components. ♠

6.4 One Essential Component

Let κ be the number of essential components in $I(J, \lambda)$. Here we prove that $\kappa = 1$.

Lemma 6.5 *There exist polygons J_1, J_2, K, all similar to J, such that the difference map $\Delta : J_1 \times J_2 \to \boldsymbol{C}$ is a folding map adapted to K. Moreover, $\Delta^{-1}(K)$ has $\kappa + 1$ essential components, and the restriction of Δ to each essential component has degree 1 with respect to its positive orientation and the counter-clockwise orientation of K.*

Proof: Define

$$J_1 = J, \qquad J_2 = \frac{\lambda}{\lambda - 1} J, \qquad K = \frac{1}{1 - \lambda} J. \tag{11}$$

Define

$$T_p(z) = \lambda(z - p) + p. \tag{12}$$

A triple of points (p_1, p_2, p_3) has shape λ if and only if the points are unequal and $T_{p_1}(p_2) = p_3$. Consider now a pair $(p_1, q_2) \in J_1 \times J_2$. We let

$$p_2 = \frac{\lambda - 1}{\lambda} q_2 \in J_1. \tag{13}$$

We compute

$$T_{p_1}(p_2) = (1 - \lambda)(p_1 - q_2). \tag{14}$$

Therefore, p_1, p_2, and $p_3 = T_{p_1}(p_2)$ are the vertices of a triangle in $I(J, \lambda)$ if and only if

$$q_2 \neq \frac{\lambda}{\lambda - 1} p_1, \qquad p_1 - q_2 \in K. \tag{15}$$

Assuming that we have picked parametrizations α_1 and α_2 of J_1 and J_2 respectively, we let s_1, s_2 be such that $p_1 = \alpha_1(s_1)$ and $q_2 = \alpha_2(s_2)$. Equation 15 tells us that $\Delta(s_1, s_2) \in K$ if and only if one of two conditions holds:

1. $q_2 = \lambda/(\lambda - 1)p_1$, or
2. p_1 and p_2 are the first two vertices of a triangle in $I(J, \lambda)$.

All the pairs (s_1, s_2) corresponds to an essential component of $\Delta^{-1}(K)$. The remaining pairs correspond to the components of $I(J, \lambda)$. So, $\Delta^{-1}(K)$ has $\kappa + 1$ essential components.

The map $(s_1, s_2) \to (p_1, q_2) \to (p_1, p_2)$ gives an affine homeomorphism between all but one of the components of $\Delta^{-1}(K)$ and the components of I. The one remaining component is an embedded polygon disjoint from all these. This polygon represents $(1, 1)$ in $H_1(\boldsymbol{T})$; it is essentially the "main diagonal" when J_2 is identified with J_1. Therefore $\Delta^{-1}(K)$ is a finite union of pairwise disjoint embedded polygons. when positively oriented, each essential component represents $(1, 1)$ in $H_1(\boldsymbol{T})$.

As we trace out a $(1, 1)$ component in the positive orientation, the corresponding point p_1 winds once around J_1 counter-clockwise. But then so do p_2 and p_3. Since $p_3 = (1 - \lambda)(p_1 - p_2)$, we see that $p_1 - p_2$ winds counter-clockwise around K. Hence Δ has degree 1 with respect to the preferred orientations. ♠

First assume that J is obtuse. Then the Folding Lemma says that the folding set B for Δ is a disjoint union of embedded polygons, exactly two of which are essential. These essential polygons represent $(1, 1)$ in $H_1(\boldsymbol{T})$, just like the components of $A = \Delta^{-1}(K)$. The sides of B are horizontal and vertical. Taking J generic, we can arrange that the sides of A are never horizontal or vertical. If a vertex of A lies in B or a vertex of B lies in A, then some triangle of shape λ is inscribed in J in such a way that it uses two of the vertices of J. This does not happen. Hence A and B have transverse intersection. All in all, the map Δ and the polygon K satisfy the hypotheses of the Component Lemma. Since B has exactly 2 essential folding components, the Component Lemma tells us that $\kappa + 1 \leq 2$. Since κ is odd, we have $\kappa = 1$.

Now we deduce the general case from the restricted case just treated. If J is not obtuse, we can choose a sequence of obtuse and λ-adapted polygons $\{J_n^*\}$ which converge to J. We get J_n^* by cutting off the corners of P very near to the vertices. The space $I(J_n^*, \lambda)$ has exactly one essential component. The corresponding spaces $I(J_n^*, \lambda)$ converge to $I(J, \lambda)$ in the Hausdorff metric as $n \to \infty$. The convergence is stronger, in the sense that almost every point of $I(P, \lambda)$ has the short arcs property discussed in the proof of the Folding Lemma. This is enough to push through the same limiting argument as the one given there. This completes the proof in the unrestricted case.

7 Unfinished Business

7.1 Almost Adapted Polygons

In this chapter we prove Lemma 2.11.

Let λ be a shape parameter. For convenience of exposition, we assume that λ is not the equilateral parameter. For $m = 0, 1, 2, 3$ we say that a triangle τ is *m-inscribed* in J if m of its vertices are also vertices of J. Generically, all triangles of shape λ inscribed in J are either 0-inscribed or 1-inscribed. We say that J is *almost adapted* to λ if

1. J is angle-adapted to triangles of shape λ. (See Theorem 1.4.)

2. For each good triple (i, j, k), the triple (L_i, L_j, L_k) is adapted to λ.

3. J has one 2-inscribed triangle of shape λ and no 3-inscribed triangles of shape λ.

Remark: In the equilateral case, there would necessarily be three 2-inscribed triangles of shape λ, thanks to the relabeling operation. This is why we are ignoring this case.

Lemma 7.1 *Suppose that J is almost adapted to λ. Then $I(J, \lambda)$ is an embedded graph whose edges are straight line segments. With at most one exception, every vertex has degree 2. The one exceptional vertex, if it exists, has degree 0 or 4.*

Proof: The same analysis as in the proof of Lemma 6.3 shows that $I(J, \lambda)$ is a finite union of line segments having disjoint interiors, together with perhaps one isolated point, the one corresponding to the singularly inscribed triangle. Call this the *singular point*. The endpoints of the segments are the vertices. The nonsingular vertices lie in the interiors of brick facets and hence all have degree 2. We just have to analyze the singular point.

If the singular point is isolated it has degree 0. Otherwise it has degree between 1 and 4 because it can only be incident to segments in the incident bricks. Any graph has an even number of vertices of odd degree, and there are no other vertices of odd degree. Hence, the singular vertex must have even degree at most 4. ♠

7.2 Proof of Lemma 2.11

Suppose that $\{\lambda_t\}$ is a linear variation of the shape parameter, interpolating between λ_0 and λ_1. We take a generic polygon J which is angle-adapted to λ_t for all $t \in [0,1]$, and adapted to λ_t for all but finitely many values $0 < t_1, ..., t_k < 1$, and almost adapted at the special values. Let $I_t = I(J, \lambda_t)$.

Let $\Upsilon_0, ..., \Upsilon_k$ be components of the topological space

$$[0,1] - (t_1 \cup ... \cup t_k).$$

Here Υ_0 and Υ_k are half-open intervals and the other Υ_j are open intervals. Let

$$C_j = \bigcup_{t \in \Upsilon_j} \beta_t.$$

Here β_t is the essential component of I_t.

Inspecting the proof of Lemma 6.3 we see that β_t varies continuously for $t \in \Upsilon_j$. Hence C_j is a topological (and indeed piecewise algebraic) cylinder. These sets are path connected. In view of Lemma 7.1, here are the only possible changes to I_t as t passes through a special value.

- An inessential component is born or dies.
- Two inessential components merge into one, or the reverse.
- The essential component merges with an inessential component to form the new essential component, or the reverse.

Theorem 6.1 prevents an inessential component from splitting into two essential components.

From this analysis of the topological transitions, we see that $\overline{C}_j \cap \overline{C}_{j+1}$ is either a polygonal loop in I_{t_j} or two polygonal loops of I_{t_j} joined together at the exceptional vertex. Hence

$$B = \bigcup_{j=0}^{k} \overline{C}_k$$

is a connected subset of $I(J, \lambda_0, \lambda_1)$. By construction $\beta_0, \beta_1 \subset B$. This completes the proof.

7.3 Some Extra Discussion

In the argument above, we varied λ but not J. Were we to vary J as well as λ, we would see the same kinds of topological transitions. These transitions would take place as we moved around in the space $S(N, \lambda)$ of N-gons that are angle adapted to λ. However, I do not know a proof that $S(N, \lambda)$ is connected. Without connectivity, the space is not so useful. Here we describe a more flexible space.

Say that J is *weakly angle-adapted* to λ if the minimum *interior* angle of J exceeds the maximum interior angle of a triangle of shape λ. Let $S_+(N, \lambda)$ denote the set of N-gons that are weakly angle-adapted to the parameter λ. If $J \in S_+(N, \lambda)$, then the very small triangles inscribed in J are all ungracefully inscribed. For this reason, Theorem 1.4 holds for generic $J \in S_+(N, \alpha)$ provided we replace the space $I(J, \lambda)$ with the space $I_+(J, \lambda)$ of gracefully inscribed triangles. The following lemma is the foundation for a good deformation theory.

Lemma 7.2 *$S_+(N, \lambda)$ is path connected.*

Proof: The subset of convex N-gons in $S_+(N, \lambda)$ is clearly connected. If J is non-convex, let ΨJ denote the subset of the convex hull of J lying in the unbounded component of $\boldsymbol{C} - J$. We can triangulate ΨJ using just vertices of J and consider a leaf of the dual tree. This gives us vertices v, v_2, v_3 of J such that the triangle $\Delta(v, v_2, v_2)$ has its interior disjoint from the interior of J. We then make the deformation shown in Figure 7.1, to a polygon that has v_2 in the middle of the edge $\overline{vv_3}$.

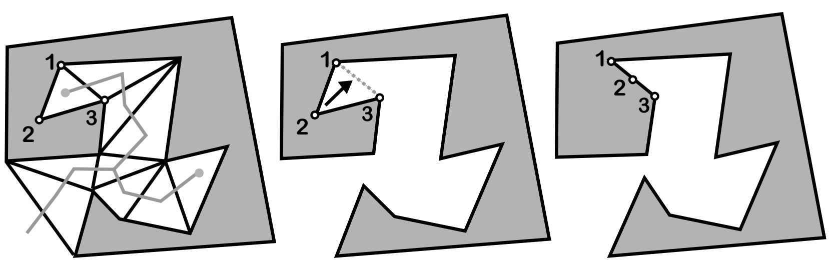

Figure 7.1: Simplifying the polygon

This deformation only decreases the interior angle at v_2, but it decreases it from some number larger that π to π. So, the deformation keeps us in $S_+(N, \lambda)$ and reduces us to showing that $S_+(N-1, \lambda)$ is connected. This is true by induction. ♠

8 References

[**M**] B. Matschke, *A survey on the Square Peg Problem*, Notices of the A.M.S. **Vol 61.4**, April 2014, pp 346-351.

[**M**] M. Meyerson *Equilateral Triangles and Continuous Curves*, Fundamenta Mathematicae, 110.1, 1980, pp 1-9.

[**N**] M. Neilson, *Triangles Inscribed in Simple Closed Curves*, Geometriae Dedicata, 1991

[**P**] I. Pak, *Lectures on Discrete and Polyhedral Geometry*, online book, www.math.ucla.edu/∼pak/geompol8.pdf, (2018)

[**S**] R. E. Schwartz, *A Trichotomy for Rectangles Inscribed in Jordan Loops*, preprint, 2018