The Optimal Paper Moebius Band: A Friendly Account

Richard Evan Schwartz *

November 17, 2023

Abstract

This paper gives a friendly account of my proof of the optimal paper Moebius band conjecture. The material here parallels the material in the original paper but there is a lot more explanation here and also more pictures.

1 Introduction

My paper [**S**] solves the optimal paper Moebius band conjecture of Benjamin Halpern and Charles Weaver. They formulate this conjecture in their 1977 paper [**HW**]. I learned about this conjecture from the beautiful expository article written by Dmitry Fuchs and Sergei Tabachnikov, [**FT**, Chapter 14]. This is a chapter in their book, *The Mathematical Omnibus*.

This is probably the first paper I have written that would be of interest to people even in high school. The conjecture and its answer are pretty easy to state, so I thought it would be nice if someone without an extensive math background could follow the proof. The material here parallels the material in [**S**], but there is more commentary, more explanation, and fewer technical terms. My original paper [**S**] also has more references than this account.

You make a paper Moebius band by taking a strip of paper, giving it an odd number of twists, and then taping the ends together. If you want to appreciate this article, you should do this on your own! The question is, how

*Supported by N.S.F. Grant DMS-2102802 and also a Mercator Fellowship.

short of a strip can you use? So, make a paper Moebius band by taking a strip of paper, giving it <u>one</u> twist, and then taping the ends. Now I want you to do an experiment with this thing. Try to simulate a shorter strip by pinching off the strip you have. In the same way that it is difficult to describe how to tie your shoes, it is difficult to describe exactly what I mean by "pinching it off", but you should just play around. If you keep pinching this thing off, you might eventually see a kind of triangular pattern form.

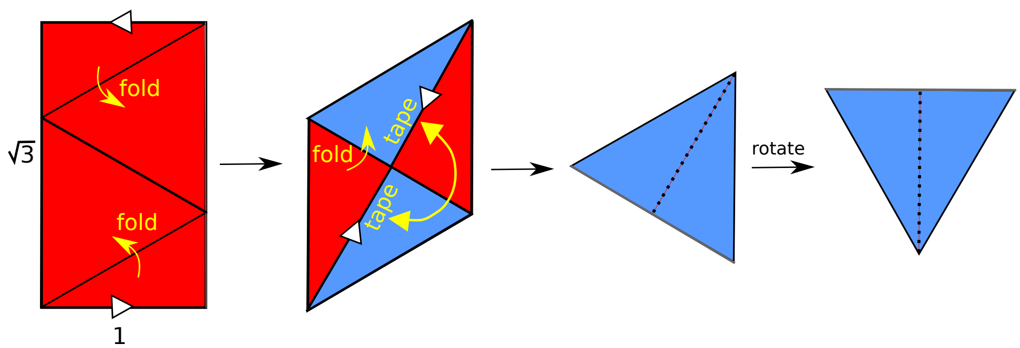

Figure 1.1: The triangular Moebius band

Whether or not your pinching experiment works, Figure 1.1 shows a famous example that is based on a $1 \times \sqrt{3}$ strip of paper. The strip in Figure 1.1 is colored red on one side and blue on the other. You are supposed to fold and then somehow tape this thing as indicated in Figure 1.1. The tape runs along the dotted line in the "inside" of the little triangular "wallet" you are making.

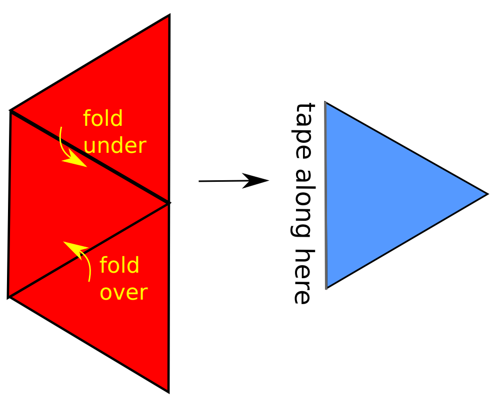

Figure 1.2: The triangular Moebius band: another view

This thing is difficult to tape. Figure 1.2 shows another way to make it where the taping is done "on the outside". This one looks a lot simpler. Why bother with Figure 1.1? Well, Figure 1.1 starts with a *rectangular* strip of paper, and this is nicer. Also, the rightmost image in Figure 1.1 shows a kind of "T-pattern" made from the top edge and the dotted line which will be important in the proof of the conjecture.

The result I proved is that if you can twist a $1 \times \lambda$ rectangle strip into a paper Moebius band, then $\lambda > \sqrt{3}$. In [**S**] this is called the Main Theorem. That was the conjecture of Halpern and Weaver. I also proved that if you do twist a $1 \times \lambda$ strip of paper into a paper Moebius band, and λ is very nearly equal to $\sqrt{3}$ (but, of course, larger) then the result looks very much like the triangular Moebius band depicted in Figures 1.1 and 1.2. The closer λ comes to $\sqrt{3}$ the closer the thing you made comes to the triangular Moebius band. In [**S**] this is called the Triangular Limit Theorem.

The first thing to note about these results is that they require some formal definition of what a paper Moebius band actually is. Let me give an example of what it is not. You could start with a square paper and fold it up like an accordion using an odd number of folds. This would give you a pretty thin strip. You could then take this strip, give it a twist, and take the ends together. Both ends make a kind of zig-zag pattern and they really do fit together. Figure 1.3 illustrates this.

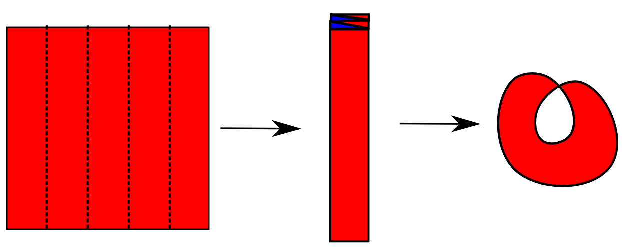

Figure 1.3: A weird example that doesn't count

There! You just turned a 1×1 strip of paper into a Moebius band. This seems to violate the Main Theorem. Also, Halpern and Weaver knew all about examples like this, so how could they have made the conjecture? The answer is that this weird example doesn't count as a paper Moebius band. To rule it out, we need a good definition. I'll motivate the definition first and then give it, somewhat informally.

Take a piece of paper and gently bend it in space. The thing is curved in some sense. But, at the same time, it still has lots of straight line segments on it. If you put your finger anywhere on your curved piece of paper, you will see that there is actually a direction you can move, in a straight line, that keeps you on the paper. To take a special case, try rolling the paper into a cylinder. You'll see that through every point there is a straight line segment that runs along the cylinder. The thing about these line segments is that they do not just *stop* in the middle of the paper. They run end-to-end.

If you gently twist up a paper strip into a Moebius band you will also see these line segments. In contrast, if you make the weird construction from Figure 1.3 you will not see these line segments running all the way across the Moebius band. So, let us say that a *paper Moebius band* is a thing you can make from a strip of paper that has, through each point, a straight line segment that cuts across the strip and has endpoints in the boundary. Figure 1.4 shows the kind of thing I have in mind.

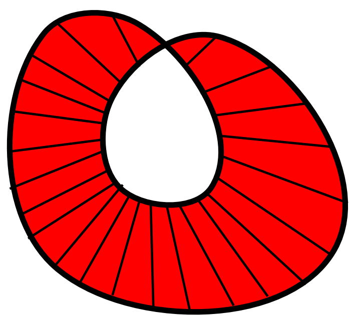

Figure 1.4: A paper Moebius band with the lines drawn in

The definition I give in [**S**] is more formal than this because in a math paper I don't want to use terms like *piece of paper*. But, these lines are the most important part of the definition and they are what I use to prove the conjecture. I call these lines the *bends*, because the Moebius band is somehow gently bending in space along these lines.

You might wonder how this definition fits with the triangular Moebius band shown in Figures 1.1 and 1.2. Figure 1.5 shows how these bends look on the triangular Mobiue band. I have drawn them on the first part of Figure 1.1 (and then copied the result into Figure 1.5.) So, Figure 1.5 shows the lines drawn on the initial strip of paper. I call these lines the *pre-bends* because they are drawn on the strip before you fold it up. So, we might say that you make a paper Moebius band by drawing the pre-bends on a strip of paper, then bending the paper in space along the pre-bends, thereby creating a paper Moebius band with bends.

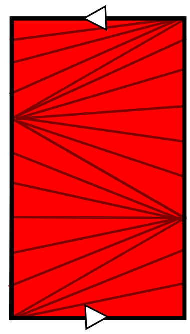

Figure 1.5: The pre-bends.

As in Figure 1.4, these line segments are supposed to go through every point, but of course I can only draw finitely many of them.

Another analogy, suggested to me by Ander Petri-Hidalgo, is that you could think of the Moebius band as something like a rope ladder. The bends are like the rungs of the ladder. The ladder is stiff in the direction of the rungs, and flexible along the direction of the ropes that hold the rungs together.

You can think about a paper Moebius band in terms of these bends. The bends are twisting around in space somehow and making a kind of loop. What I do is consider pairs of these bends. If you have two bends they might or might not lie in the same plane in space. Let's call two bends *coplanar* if they lie in a common plane. If the bends are not parallel, this means that the lines extending them intersect.

How can you visualize such a thing? Well, draw two line segments on a piece of paper that is sitting on your desk. These two line segments are in the same plane, namely the plane that extends your desk. Now imagine that these two line segments are part of a loop of such line segments which might rise above or below your desk and tilt in crazy ways. Most of the loop is floating around in space in a complicated way but at two special bends – the ones you drew first – it sits in the plane of your desk.

Given two coplanar bends, you can draw the lines that extend them, as shown in Figure 1.6. These lines might or might not be perpendicular.

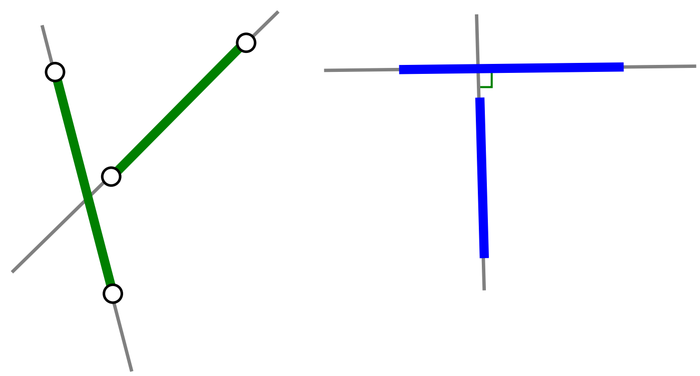

Figure 1.6: Coplanar bends. The right side is a T-pattern.

The green bends in Figure 1.6 do not lie on perpendicular lines and the blue bends do. Let us say that a *T-pattern* is a pair of coplanar bends that lie on perpendicular lines. The two bends comprising the T-pattern do not touch each other, but the lines extending them might touch each other. We will take up this issue more carefully in the next chapter.

The first half of my proof is contained in the following result:

Lemma 1.1 (T) *Every paper Moebius band has a T-pattern.*

That's right. No matter how you twist up a paper Moebius band and put it in space, you can always find a pair of bends that are coplanar and lie on perpendicular lines.

The second half of my proof uses the T-pattern to prove the result about $\lambda > \sqrt{3}$. I'll phrase this like this.

Lemma 1.2 (G) *Suppose that a $1 \times \lambda$ paper Moebius band has a T-pattern. Then $\lambda > \sqrt{3}$.*

In light of Lemma T, the phrasing of Lemma G might seem a bit odd. Since they all have T-patterns, why include the word *suppose*? The phrasing is designed to emphasize the idea that I will prove $\lambda > \sqrt{3}$ by using properties of the T-pattern. The idea is that I will first prove Lemma G, just using properties of the T-pattern, then I will prove lemma T which shows that these things always exist. Lemma G and Lemma T together obviously imply the Main Theorem. It turns out that they also imply the Triangular Limit Theorem.

Here is how the rest of this paper is organized. In §2 I will prove Lemma G. In §3 I will prove Lemma T. At the end of §2 I will make some remarks on how the Triangular Limit Theorem works. The proof here are kind of parallel to what I do in [**S**] but here I include a lot more detail and try to give a more intuitive account.

I got a lot of help thinking about paper Moebius bands. I would like to thank Brienne Brown, Matei Coiculescu, Robert Connelly, Dan Cristofaro-Gardiner, Dmitry Fuchs, Javier Gomez-Serrano, Jim Hourihan, Anton Izosimov, Jeremy Kahn, Stephen D. Miller, Ander Petri-Hidalgo, Neil Schwartz, and Sergei Tabachnikov for helpful discussions about this subject.

2 Proof of Lemma G

The goal of this chapter is to prove Lemma G. This result is mentioned at the end of the introduction. First I'll assenble some general geometric principles, then I'll talk a bit more about the structure of paper Moebius bands, and then I'll put it all together and give the final argument.

2.1 The Pythagorean Theorem

The Pythagorean Theorem says that for a right triangle having side lengths a, b, c, with c the longest side, we have $a^2 + b^2 = c^2$. One of the most famous proofs follows the diagram in Figure 2.1.

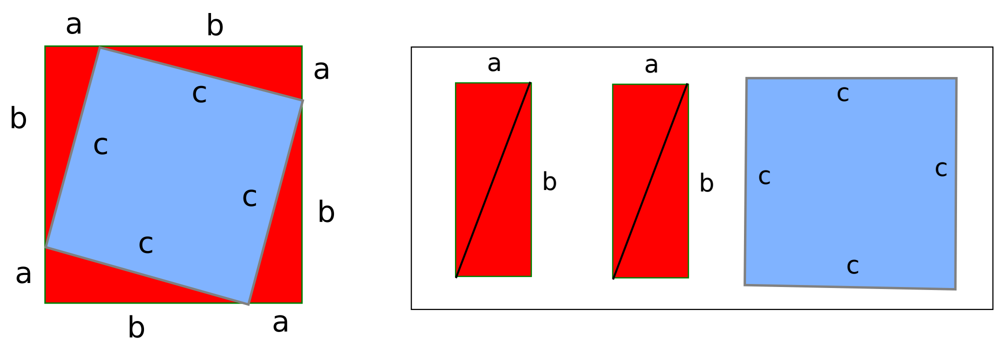

Figure 2.1: The diagram for the Pythagorean Theorem

Let's calculate the area of the red and blue square in two ways. First of all, the area is

$$(a + b)^2 = a^2 + b^2 + 2ab,$$

because the area of a square is just the square of the side length. At the same time, we can pull this thing apart and arrange the individual pieces on the right. The area calculated this way is

$$2ab + c^2.$$

But then

$$a^2 + b^2 + 2ab = 2ab + c^2.$$

Cancelling out the $2ab$ we get $a^2 + b^2 = c^2$, as desired.

2.2 A Minimizing Property of Isosceles Triangles

Let's consider triangles which have a horizontal base b and height h. Figure 2.2 shows that there are many of these, depending on where the apex is in relation to the base. If the apex lies directly beneath the midpoint of the base, then the triangle is isosceles. We're going to show that the perimeter is always at least $b + \sqrt{b^2 + 4h^2}$. Call the perimeter P.

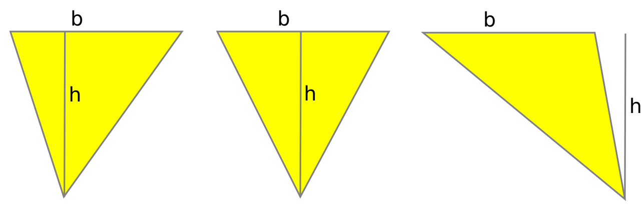

Figure 2.2: Several triangles with base b and height h.

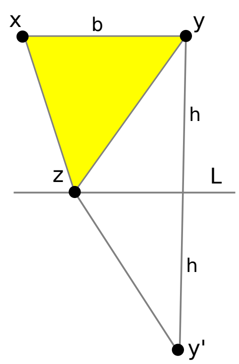

Figure 2.3: The diagram for the proof

The expression $d(a, b)$ denotes the distance from a point a to a point b. Call the vertices of the triangle x, y, z as shown in Figure 2.3. Let y' be the point you get when you reflect y in the horizontal line L through z. By symmetry, $d(z, y) = d(z, y')$. We have

$$P = d(x, y) + d(x, z) + d(z, y) = d(x, y) + d(x, z) + d(z, y') \geq$$

$$d(x, y) + d(x, y') = b + \sqrt{b^2 + 4h^2}.$$

The inequality has the following explanation. The quantity $d(x, z) + d(z, y')$ is the length of the possibly crooked path connecting x to y'. But this path is at least as long as the line segment connecting x to y', and this line segment has length $d(x, y')$. The last equality is the Pythagorean theorem applied to the right triangle with vertices x, y, y'.

2.3 Changing Variables

Let ∇ denote the kind of triangle we have considered in the previous section. Let $n(\nabla)$ denote the sum of the lengths of the non-horizontal sides of ∇ and let $p(\nabla)$ be the perimeter. From the previous section, we have

$$n(\nabla) \geq \sqrt{b^2 + 4h^2}, \qquad p(\nabla) \geq b + \sqrt{b^2 + 4h^2}.$$

We only worked out the second of these inequalities in the previous section but clearly $n(\nabla) = p(\nabla) - b$. Again, we stress that these bounds work for any triangle having base b and height h.

Let us transform this result several times. Suppose that ∇ has base b and height greater than 1. Then the expressions above decrease when we replace h by 1. So, we can conclude that

$$n(\nabla) > \sqrt{b^2 + 4}, \qquad p(\nabla) > b + \sqrt{b^2 + 4}.$$

Now we set $b = \sqrt{1+t^2}$. If ∇ has base $\sqrt{1+t^2}$ and height greater than 1 then

$$n(\nabla) > \sqrt{5+t^2}, \qquad p(\nabla) > \sqrt{1+t^2} + \sqrt{5+t^2}. \tag{1}$$

2.4 Loops and Arcs

Let $\ell(\gamma)$ denote the length of a curve γ. As in the previous section, suppose ∇ is a triangle of base $\sqrt{1+t^2}$ and height greater than 1. Suppose γ is a loop that contains all the vertices of ∇, as shown on the left side of Figure 2.4. What I mean is that γ is some kind of loop in space and then γ touches down into the plane containing the triangle ∇ and contains all the vertices of ∇.

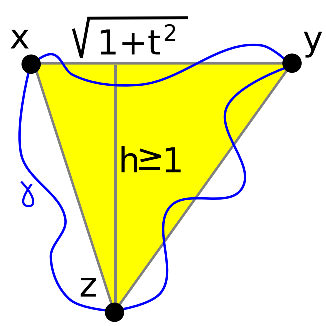

Figure 2.4: A loop containing all the vertices.

In this situation, the length of γ is at least as long as $p(\nabla)$. That is

$$\ell(\gamma) > \sqrt{1+t^2} + \sqrt{5+t^2}. \tag{2}$$

Relatedly, suppose γ is an arc which connects the two vertices of the base of ∇ and goes through the third vertex of ∇. This is shown in Figure 2.5. Once again, γ is an arc in space but it touches down and hits the three vertices of ∇.

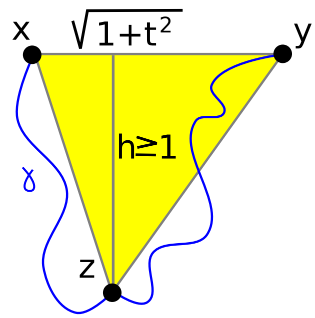

Figure 2.5: An arc connecting the base vertices.

In this situation, the length of γ is at least as long as $n(\nabla)$. That is

$$\ell(\gamma) > \sqrt{5+t^2}. \tag{3}$$

Equations 2 and 3 are two of our main ingredients in our proof of Lemma G. Now we turn to the Moebius band with its T pattern. Towards the end of the argument we will bring these two equations back into the picture.

2.5 Cutting Open A Paper Moebius Band

Call the Moebius band M. We are assuming that M has a T-pattern. Again, this means that there are two bends which are coplanar and lie on perpendicular lines. Look again at the right side of Figure 1.6. I will call these bends B and T. Here B is going to stand for "bottom" and T is going to stand for "top". All this means is that when I draw things B will be on the bottom and T will be on the top.

Let $\overleftrightarrow{B}$ be the line extending B and let $\overleftrightarrow{T}$ be the line extending T. We already know that B and T are disjoint from each other: they do not intersect. However, several things could happen with the extending lines. Here are the possibilities:

1. $\overleftrightarrow{B}$ is disjoint from T and $\overleftrightarrow{T}$ is disjoint from B.
2. $\overleftrightarrow{B}$ intersects T and $\overleftrightarrow{T}$ is disjoint from B.
3. $\overleftrightarrow{B}$ is disjoint from T and $\overleftrightarrow{T}$ intersects B.

The fourth possibility, a "double intersection" cannot arise because then T and B would not themselves be disjoint. If the third option occurs, we can rename our bends so that instead the second option occurs. This means that if we switch the names of our bends, if necessary, we can arrange that $\overleftrightarrow{T}$ does not intersect B. This is what we do.

Next, we rotate the plane containing these bends so that B is vertical and T is horizontal and B lies beneath T. Figure 2.6 shows the situation. Figure 2.6 also shows a triangle made from some of the vertices of these bends. Notice the similarity with the triangle we have been considering in the previous sections.

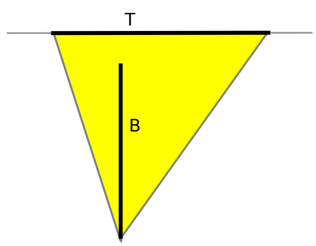

Figure 2.6: The T-pattern

It is worth mentioning that the bend B might be so far over to one side that $\overleftrightarrow{B}$ does not intersect T. But, we don't care about whether or not this happens. All we care about is that B lies totally beneath T, and this is guaranteed by the fact that $\overleftrightarrow{T}$ misses B.

At this point we are going to cut open M along B. You can try this yourself. Take your paper Moebius band and cut it open with a scissors, making a straight-line cut straight across. If you do this, and then you unwrap the thing, you will get a trapezoid with bilateral symmetry. Figure 2.7 shows this.

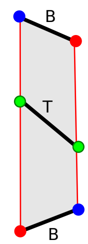

Figure 2.7: The Moebius band cut open along B.

The colors of the dots indicate how the re-taping would work if you were to reassemble this thing. The taping would match up the red dots and it would simultaneously match up the blue dots. You should really try this experiment a few times to see what I am talking about here.

2.6 The First Bound

Let's put together Figures 2.6 and 2.7 and add some more detail. Figure 2.8 shows the result. The most important thing to note here is the number t. We define this so that $-t$ is the slope of the T segment on the left when the cut-open Moebius band is rotated so that the red boundary is vertical. The segment B goes over 1 and down t, so its length is $\sqrt{1+t^2}$. This is an application of the Pythagorean Theorem.

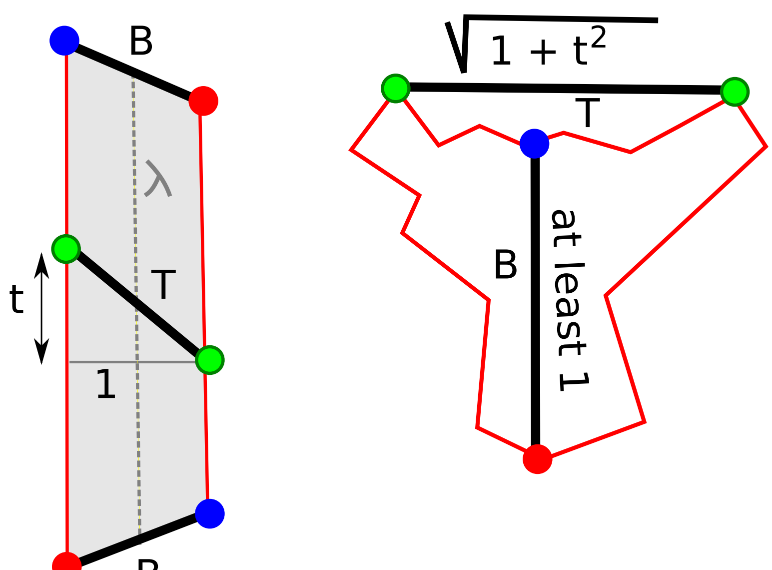

Figure 2.8: The trapezoid and the T-pattern side by side.

The red portion of the boundary of the trapezoid on the left is the whole of the boundary of M. When you tape M together along B all that is left in the boundary is the red part. When we remember that this boundary is floating around in space, we note that it makes a loop that hits all the vertices of our T-pattern. By Equation 2, the red loop on the right has length greater than

$$\sqrt{1+t^2}+\sqrt{5+t^2}.$$

The point here is that the base of the relevant triangle is $\sqrt{1+t^2}$ and height is greater than 1.

Notice that this red loop is what you get after you tape up the trapezoid on the left and look at the vertical sides. The sum of these sides is just 2λ, because λ is the length of the centerline of the trapezoid, shown as a dotted line. Let me explain this in a bit more detail. The length of the grey centerline is the average of the length of the left red side and the red right side. So, the sum of the left red length and the right red length is twice the length of the grey centerline.

In short we have proved that

$$2\lambda > \sqrt{1+t^2}+\sqrt{5+t^2}. \tag{4}$$

It is worth pointing out that this bound is already not bad. No matter what value of t we put in we always have $2\lambda > 1+\sqrt{5}$. In other words, λ exceeds the golden ratio $(1+\sqrt{5})/2$. This is already an improvement over the bound in the 1977 paper by Halpern and Weaver.

Another thing to notice is that when $t \geq 1/\sqrt{3}$ the right hand side of Equation 4 is $2\sqrt{3}$. This gives us $\lambda > \sqrt{3}$ whenever $t \geq 1/\sqrt{3}$. But, of course, we do not know that $t \geq 1/\sqrt{3}$. We are going to get a second bound which takes care of the possibility that $t < 1/\sqrt{3}$.

2.7 The Second Bound

Now we are going to get a different bound by looking more carefully at this red boundary. Let us recolor some of the red boundary magenta. Figure 2.9 shows this. Let H denote the total length of the magenta boundary and let D denote the total length of the red boundary. In terms of the picture $H = H_1 + H_2$ and $D = D_1 + D_2$.

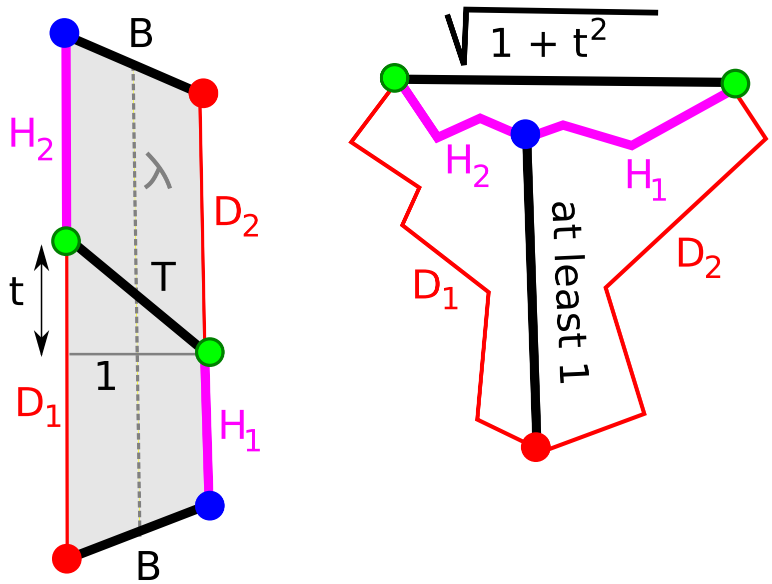

Figure 2.9: The trapezoid and the T-pattern side by side: recolored

Why these names? Well, the red boundary on the right kind of moves diagonally and that is why I call the sum of the lengths D. You could think of D as measuring the length of the "diagonal part" of the boundary. The magenta part is sort of moving horizontally, so I call the sum of these lengths H. You might not love this mnemonic, but it helped me get everything straight.

I claim that

$$H = D - 2t. \tag{5}$$

Why is this? Well, suppose that $t = 0$. In this case we would have $H = D$ by symmetry. Suppose we think of T rotating about its midpoint so as to have slope $-t$. Then on each side the magenta parts lose a total of $t/2$ and the rad parts gain a total of $t/2$. So, the difference $D - H$ increases by $4 \times (t/2) = 2t$.

Now as before $2\lambda = H + D$. At the same time, by Equation 3, the red arc on the left has length greater than $\sqrt{5 + t^2}$. In other words

$$D > \sqrt{5 + t^2}.$$

Now we put it all together

$$2\lambda = H + D = 2D - 2t > 2\sqrt{5 + t^2} - 2t.$$

In short

$$2\lambda > 2\sqrt{5 + t^2} - 2t. \tag{6}$$

2.8 The End of the Proof

We have established 2 facts:

$$2\lambda > f(t) := \sqrt{1+t^2} + \sqrt{5+t^2}, \qquad 2\lambda > g(t) := 2\sqrt{5+t^2} - 2t. \tag{7}$$

We don't know what t is, but regardless of the value of t we get when we cut our Moebius band open, we have these two bounds.

There are two cases to consider. Suppose first that $t \geq 1/\sqrt{3}$. We have already discussed this above, but let's say it again. You can check that $f(1/\sqrt{3}) = 2\sqrt{3}$. Also, as t increases past $1/\sqrt{3}$ the function f increases. Therefore, if $t \geq 1/\sqrt{3}$ we have $2\lambda > 2\sqrt{3}$, or $\lambda > \sqrt{3}$.

Now we consider the other case. Suppose that $t \leq 1/\sqrt{3}$. You can check that $g(1/\sqrt{3}) = 2\sqrt{3}$. Also, the function g increases as t decreases from $1/\sqrt{3}$. You can see this by making a computer plot and if you really want a proof, you could check that the derivative of g is negative. This shows that $2\lambda > 2\sqrt{3}$ in this case as well.

So, no matter what the value of t we have, we get $\lambda > \sqrt{3}$. This proves Lemma G.

2.9 The Triangular Limit Theorem

Suppose you have 2 points in space that are L units apart. The shortest path between these points is a line segment. Call this segment S. Suppose that γ is another path connecting these same two points. If γ has length exactly L, then $\gamma = S$. If γ has length only slightly longer than L, then γ must stay near S. This is the basic principle behind the Triangular Limit Theorem.

Let us spice up this principle and also make it more particular. Suppose that ∇ is a triangle of perimeter $2\sqrt{3}$. Suppose that γ is a loop of length 2λ that contains all the vertices of ∇. Then, as we have remarked above, the $\lambda \geq \sqrt{3}$. Moreover, if λ is only slightly larger than $\sqrt{3}$, the path γ very nearly equals the boundary of ∇. Call this the *tautness principle*.

Now we turn directly to the proof of the Triangular Limit Theorem. Suppose that we have a paper Moebius band made from a $1 \times \lambda$ strip where λ is close to $\sqrt{3}$. In this case, our two bounds $2\lambda > f(t)$ and $2\lambda > g(t)$ from the proof of Lemma G force t to be almost equal to $1/\sqrt{3}$. Also, the triangle ∇ must be very close to isosceles and having height 1. This means that the bend B is almost horizontal in Figure 2.9.

So, the triangle ∇ made from the green vertices and the red vertex on the right side of Figure 2.9 is just about the equilateral triangle with perimeter $2\sqrt{3}$. Since the red and magenta loop has length 2λ, and 2λ is very close to $2\sqrt{3}$, the red and magenta loop must almost exactly follow the boundary of ∇. Here we are using the tautness principle. This is to say that our paper Moebius band looks almost exactly like the triangular paper Moebius band.

In [**S**] I am more careful about saying what "looks almost exactly like" means, but the description above is the heart of the matter. The rest of the proof is just bookkeeping.

3 Proof of Lemma T

3.1 Vectors

You need to know something about vectors to read this chapter. We'll always talk about vectors in three dimensional space. A *vector* is a triple (x, y, z). We usually denote them like this: $\overrightarrow{v} = (x, y, z)$. The first thing you might wonder is: How is a vector different than a *point*? Honestly, I'd have to say that they are the exactly the same thing. When I was a student first learning about vectors this fact irritated me quite a bit. Why bother with vectors then?

If you are new to vectors, the best I can offer is to say that there is a wider class of objects, called *manifolds*, in which both the concept of *point* and *vector* make sense and are different from each other. Also, there is a psychological difference between a vector and a point. One usually thinks of a point as just frozen in one location in space. One thinks of a vector as a kind of movable arrow. The vector (x, y, z) points from $(0, 0, 0)$ to (x, y, z), but you could also move it and imagine that it is based at the point (say) $(1, 2, 5)$ and points to $(x + 1, y + 2, z + 5)$. As you carry this arrow with you around space, it does not change its identity.

We use vectors to record the direction and magnitude of something. More precisely, a line segment u (such as a bend) defines two vectors: namely $u_2 - u_1$ and $u_1 - u_2$. Here u_1 and u_2 are the endpoints of u. One of the vectors is the negation of the other. Here is an example: If u is the segment connecting $(2, 1, 3)$ to $(-3, 7, 1)$ then the two associated vectors are $\pm(5, -6, 2)$. The vectors represent the bends. Each vector is parallel to the bend it represents. The vector "remembers" the direction and magnitude but forgets the position.

The nice things about this convention is that it lets you apply some operations, usually reserved for vectors, to segments. For instance, to say that two segments are perpendicular in space – even though they might lie on non-intersecting lines – is to say that the vectors representing them are perpendicular. Even though there is a sign ambiguity in making this definition, no matter how we choose the signs we would get the same answer, yes or no, to the perpendicularity question.

Now let me describe the two basic operations on vectors, the dot product and the cross product. These operations are sometimes taught in high school, but more usually you see them in a multivariable calculus class.

The *dot product* of two vectors $\overrightarrow{u} = (u_1, u_2, u_3)$ and $\overrightarrow{v} = (v_1, v_2, v_3)$ is defined to be

$$\overrightarrow{u} \cdot \overrightarrow{v} = u_1 v_1 + u_2 v_2 + u_3 v_3. \tag{8}$$

This is a number. Here are some properties of this operation.

1. $\overrightarrow{u} \cdot \overrightarrow{u}$ is a positive number, equal to the square of the length of u.

2. $\overrightarrow{u} \cdot \overrightarrow{v} = 0$ if and only if these vectors are perpendicular.

3. $\overrightarrow{u} \cdot \overrightarrow{v} = \overrightarrow{v} \cdot \overrightarrow{u} = (-\overrightarrow{u}) \cdot (-\overrightarrow{v})$.

The *cross product* of the same vectors $\overrightarrow{u}$ and $\overrightarrow{v}$ is is defined to be

$$\overrightarrow{u} \times \overrightarrow{v} = (u_2 v_3 - u_3 v_2, u_3 v_1 - u_1 v_3, u_1 v_2 - u_2 v_1). \tag{9}$$

This is another vector rather than a number. Here are some properties of this operation.

1. $\overrightarrow{u} \times \overrightarrow{v} = 0$ if and only if $\overrightarrow{u}$ and $\overrightarrow{v}$ are parallel.

2. When $\overrightarrow{u} \times \overrightarrow{v} \neq 0$, it is perpendicular to both $\overrightarrow{u}$ and $\overrightarrow{v}$.

3. $\overrightarrow{u} \times \overrightarrow{v} = -\overrightarrow{v} \times \overrightarrow{u} = (-\overrightarrow{u}) \times (-\overrightarrow{v})$.

Another basic property of both these operations is that they are continuous functions of the coordinates of $\overrightarrow{u}$ and $\overrightarrow{v}$. Intuitively, if you vary the coordinates of the vectors just a little bit, you also just change the coordinates of the dot and cross product by a little bit. I'll say a bit more about this in the next section.

3.2 Continuous Functions

The proof of Lemma T relies on continuity properties of functions in a big way. I'll give both an informal definition and a formal definition. A function f is a gadget that assigns an output number to each input number. This is written as $f(x)$, or perhaps f_x, where x is the input and then $f(x)$ or f_x is the output. I'll stick with the first notation for now. Informally f is defined to be continuous if a small change in the input produces only a small change in the output.

Here is a more formal definition: f is defined to be *continuous at* x_0 if the following is true: For all $\epsilon > 0$ there is a $\delta > 0$ such that if $|x - x_0| < \delta$

then $|f(x) - f(x_0)| < \epsilon$. What this is saying is that if we want the output to change by less than ϵ we can find a δ so that if we vary the input away from x_0 by less than δ then the output does change by less than ϵ. Following this definition, f is defined by be *continuous* if it is continuous at all points where it is defined. Typically f is defined on an interval, or on all real numbers, or on a circle, or something similar. I remember that when I first learned this definition in high school it took me a long time to wrap my brain around it.

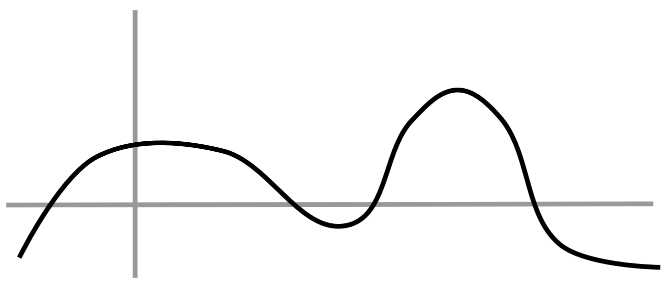

Figure 3.1: The grap of a typical continuous function

Another way to think about a continuous function is that when you draw its graph you can do so without picking up your pen. The graph has no sudden jumps in it.

You can use the definition of a continuous function to say what it means for a line segment to vary continuously as a function of a parameter. There are 6 functions which describe the line segment: namely the coordinates of each of the endpoints. What we mean by saying that the line segment varies continuously is that these 6 functions are all continuous. You can also use the definition to say what it means for a plane to vary continuously in space, and so on. In all cases, you describe the situation by a bunch of functions and then say that all these functions are continuous.

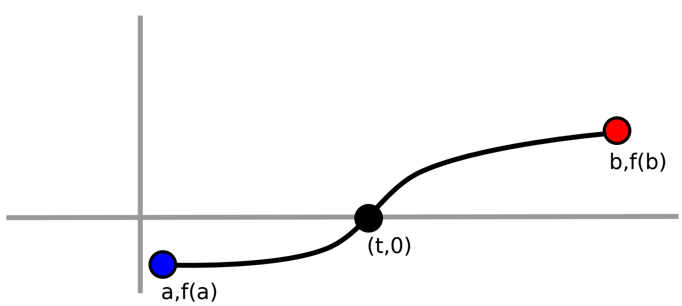

Figure 3.2: The Intermediate Value Theorem

There is one very important property of continuous functions we will use, namely the Intermediate Value Theorem. Suppose that f is defined on the interval from a to b and $f(a) < 0$ and $f(b) > 0$. Then there is some value t with $a < t < b$ such that $f(t) = 0$. I won't prove this here, but let me just say intuitively that f must change from being negative to positive at some instant, and at that instant it is 0. The same conclusion holds if $f(a) > 0$ and $f(b) < 0$.

3.3 Winding Numbers

In this section I will explain another property of continuity that has a more "topological" feel. For reasons that will have to do with the argument we give later, we say that an *avatar* is a curve in the plane that joins a point on the positive X-axis to a point on the negative X-axis and does not contain the origin. Figure 3.3 shows 3 examples of avatars.

We can assign a half-integer to each avatar. What we do is stand at the origin and watch the avatar travel from the positive X-axis to the negative X-axis. When we are done, our head has turned some number of times. Figure 3.3 shows how we assign these numbers.

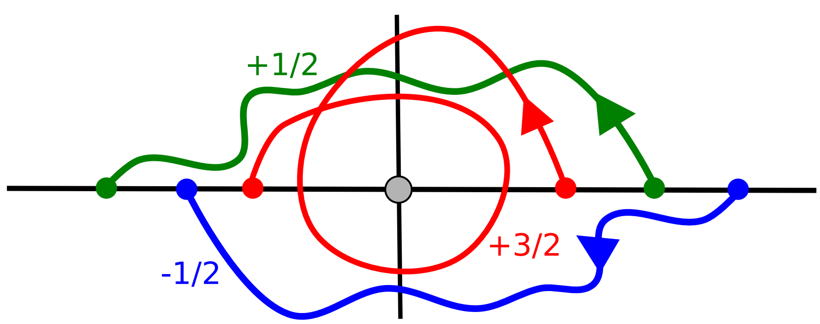

Figure 3.3: Winding numbers of avatars

These objects satisfy two basic principles:

1. If we have a continuously varying family of avatars, then their winding numbers are all the same. The only way that the avatar could change its winding number is if it crosses through the origin. But we are assuming that this does not happen.

2. Suppose γ is an avatar. Then we get a new avatar by considering $-\gamma$ and reversing the direction. Equivalently, $-\gamma$ is obtained by rotating γ

180 degrees around the origin and then reversing the direction so that it again goes from positive to negative. We call $-\gamma$ the *opposite* of γ. The blue and green avatars in Figure 3.3 are opposites. The fact is that when we replace an avatar by its opposite, we switch the sign of its winding number.

We can combine these two principles: There is no continuous motion of avatars which starts with an avatar and ends with the opposite avatar. The problem is that the winding number would have to switch sign at some point and this is impossible. If we try to make such a continuous motion, one of the curves will contain the origin and fail to be an avatar. When we put it like this, it sounds a lot like the Intermediate Value Theorem.

3.4 The Circle of Bends

On any paper Moebius band we have the bends. If you think about it, there is a circle's worth of bends. Take a look at Figure 3.4.

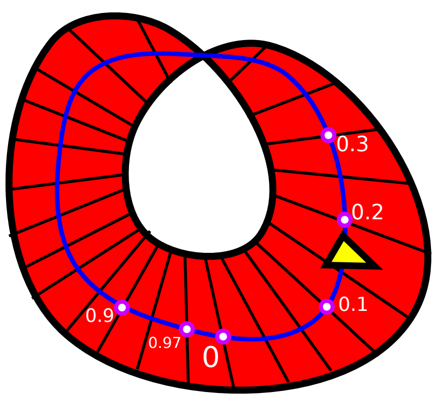

Figure 3.4: The circle of bends

You can see that the bends go around in a circle. The "pinch point" at the top of the picture is a bit hard to draw. I am drawing a 3D picture in the plane. This pinch point is really a whole band, and the blue circle is running through the middle of it. If you make a paper Moebius band out of a strip of paper and then draw a blue circle right down the middle, then you will see what I mean about the blue circle in Figure 3.4.

Let me express the circle idea differently. We can write the bends as u_t, where t is a parameter that varies from 0 to 1. The bend u_0 is some initial bend. Then, as t increases, u_t starts moving away in a continuous way and then comes back at $t = 1$ to rejoin u_0.

We don't have to start at 0. Suppose we start at the bend $u_{1/2}$. Then we can think of the t parameter as increasing to 1 and then jumping back to 0 and increasing again to 1/2. When the parameter t jumps from 1 to 0 it does not mean that the bends experience discontinuous motion. The reason is that $u_1 = u_0$. We are just starting out loop at a different place on the Moebius band. It is better to think that the parameter interval has been picked up off the number line and bent so that 1 joins together with 0. Then the parameter just endlessly goes around in a loop.

One thing I want to say here is that the description above hides an implicit choice we made. We have chosen to go around this circle of bends in a particular direction. You could imagine that we are floating in one of those swimming loops you sometimes see at water parks. The current takes us around the loop in one direction. If we reverse the current we would go around in the other direction. We make a choice once and for all. If we made the other choice the overall proof would be the same.

Recall that each bend u_s has two vectors associated to it. We call these vectors *orientation vectors*. Suppose $\overrightarrow{u}_s$ is one of them. If $s \neq t$ think of increasing the parameter (and possibly wrapping around) until we reach t. Once we have chosen $\overrightarrow{u}_s$ we can continuously propagate this choice until we reach u_t. This automatically makes a choice of $\overrightarrow{u}_t$. In other words, the orientation we choose for u_s forces the orientation for u_t. We write this as $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$. The way I think of this is that the orientation chosen for u_s spreads out and aligns (or infects) the orientations of nearby bends ahead of it, and this infection propagates until it reaches u_t.

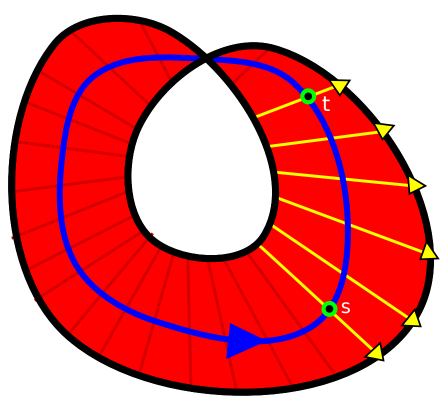

Figure 3.5: The operation $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$.

The operation $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$ has several properties.

1. $-\overrightarrow{u}_s \rightsquigarrow -\overrightarrow{u}_t$
2. $\overrightarrow{u}_t \rightsquigarrow -\overrightarrow{u}_s$

The second property comes from the following basic fact about Moebius bands: If we propagate the orientation all the way around, then it comes back reversed. We might write this as $\overrightarrow{u}_s \rightsquigarrow -\overrightarrow{u}_s$, but we are abusing the notation just to illustrate a point. Whenever we use this notation below, we have the understanding that we go less than all the way around.

3.5 A Special Case of Lemma T

Now we turn to Lemma T. Let me first give an easy argument under a simplifying assumption. We will assume that each bend u on the Moebius band M has a unique partner u' that varies continuously with u. We write the partner bend as u'_t. So, the pair (u_t, u'_t) is a continuously varying family of partner bends.

Let $m(u)$ denote the midpoint of u. Now we introduce the following function

$$h(t) = (m(u_t) - m(u'_t)) \cdot (\overrightarrow{u}_t \times \overrightarrow{u}'_t). \tag{10}$$

This equation takes a lot to unpack. The orientations are chosen so that $\overrightarrow{u}_t \rightsquigarrow \overrightarrow{u}'_t$. It seems at first like this definition depends on our choice of orientation for u_t. What if we had chosen $-\overrightarrow{u}_t$ instead? Well then we remember that $-\overrightarrow{u}_t \rightsquigarrow -\overrightarrow{u}'_t$ and also

$$\overrightarrow{u}_t \times \overrightarrow{u}'_t = (-\overrightarrow{u}_t) \times (-\overrightarrow{u}'_t).$$

So either choice of orientation leads to the same result. This is to say that the function h really is a well-defined function. It makes sense.

The fact that the bends and their partners vary continuously and that the dot and cross products are continuous functions of their inputs means that h is a continuous function. Suppose $h(t) = 0$. Lets set

$$\overrightarrow{n}_t = \overrightarrow{u}_t \times \overrightarrow{u}'_t.$$

When $h(t) = 0$ it means that the midpoints of u_t and u'_t lie in the same plane perpendicular to $\overrightarrow{n}$. But notice also that $\overrightarrow{n}$ is perpendicular to both $\overrightarrow{u}_t$ and $\overrightarrow{u}'_t$. This means that u_t and u'_t also lie in planes perpendicular to $\overrightarrow{n}$.

Consider the contraption you get by taking u_t and u'_t and the line segment joining $m(u_t)$ to $m(u'_t)$. This thing is a connected whole, and each piece is contained in a plane perpendicular to $\overrightarrow{n}$. But then all the pieces have to lie in the same plane perpendicular to $\overrightarrow{n}$. In short, when $h(t) = 0$ the bends u_t and u'_t make a T-pattern.

If $h(0) = 0$ we are done. Otherwise, we note that there is some parameter s such that $u_s = u'_0$. In other words u_s is the partner of u. But then the pair (u_s, u'_s) is just the same as the pair (u_0, u'_0) but with the two bends swapped. We compute

$$h(s) = (m(u_s) - m(u'_s)) \cdot (\overrightarrow{u}_s \times \overrightarrow{u}'_s) =$$
$$(m(u'_0) - m(u_0)) \cdot (\overrightarrow{u}'_0 \times (-\overrightarrow{u}_0)) =$$
$$(m(u'_0) - m(u_0)) \cdot (\overrightarrow{u}_0 \times \overrightarrow{u}'_0) =$$
$$-(m(u_0) - m(u'_0)) \cdot (\overrightarrow{u}_0 \times \overrightarrow{u}'_0) = -h(0).$$

In short, $h(s) = -h(0)$. So, by the Intermediate Value Theorem there is some parameter t such that $h(t) = 0$. This gives us our T-pattern.

3.6 Setting up the General Case

Let us first extend our function h so that it is more general. For any paper Moebius band and any two unequal parameters s, t we define

$$h(s,t) = (m_{u_s} - m_{u_t}) \cdot (\overrightarrow{u}_s \times \overrightarrow{u}_t). \tag{11}$$

Here $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$ as above. Also m_u is the midpoint of u. For the same reasons as above, h is a well defined function of 2 unequal variables. Also, if $h(s,t) = 0$ and $\overrightarrow{u}_s$ is not parallel to $\overrightarrow{u}_t$ (so that the cross product is nonzero) then u_s and u_t are coplanar.

Now we introduce a second function:

$$g(s,t) = \overrightarrow{u}_s \cdot \overrightarrow{u}_t. \tag{12}$$

Note that g is also a well-defined function, and $g(s,t) = 0$ if and only if $\overrightarrow{u}_s$ and $\overrightarrow{u}_t$ lie on perpendicular lines.

Finally, suppose that $g(s,t) = 0$ and $h(s,t) = 0$. Then u_s and u_t are both coplanar and lie on perpendicular lines. In other words u_s and u_t form a T-pattern. So, in order to prove Lemma T we just have to find unequal parameters s, t such that $g(s,t) = 0$ and $h(s,t) = 0$. This is kind of like a 2-dimensional version of the Intermediate Value Theorem.

Before getting to our proof, let us gather some more information about our functions g and h. Let me recall that $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$ means that $\overrightarrow{u}_t \rightsquigarrow -\overrightarrow{u}_s$. This means that

$$g(t,s) = \overrightarrow{u}_t \cdot (-\overrightarrow{u}_s) = -\overrightarrow{u}_s \cdot \overrightarrow{u}_t = -g(s,t).$$

Likewise,

$$h(t,s) = (m_{u_t} - m_{u_s} \cdot (\overrightarrow{u}_t \times (-\overrightarrow{u}_s)) = (m_{u_t} - m_{u_s}) \cdot (\overrightarrow{u}_s \times \overrightarrow{u}_t) = -h(s,t).$$

In summary,

$$g(t,s) = -g(s,t), \qquad h(t,s) = -h(s,t). \tag{13}$$

There are two more pieces of information we note.

1. Suppose the path $s \to t$ in the circle is very short. That is, t is just ahead of s. Then $g(s,t)$ is large and positive and $h(s,t)$ is nearly 0.
2. Suppose that the path from $s \to t$ in the circle is very long, going almost all the way around. That is, t is just behind s. Then $g(s,t)$ is large and negative and $h(s,t)$ is nearly 0.

3.7 The Annulus

Let's think more about the parameters (s,t) considered above. For each choice of s, there is an interval's worth of values for t. The t parameter starts just ahead of s and then can wrap all the way around and come back just behind s. So each s gives rise to an interval of possible values for t. At the same time, there is a circle's worth of choices for s. So, all in all, we can picture the *domain* for g and h as a "circle's worth of intervals". This is also known as a annulus. I've added in some of these intervals so you can see how they sit inside the annulus. The left side of Figure 3.6 shows a picture of this annulus with some points picked out. The right side of Figure 3.6 shows how these points correspond to pairs (s,t).

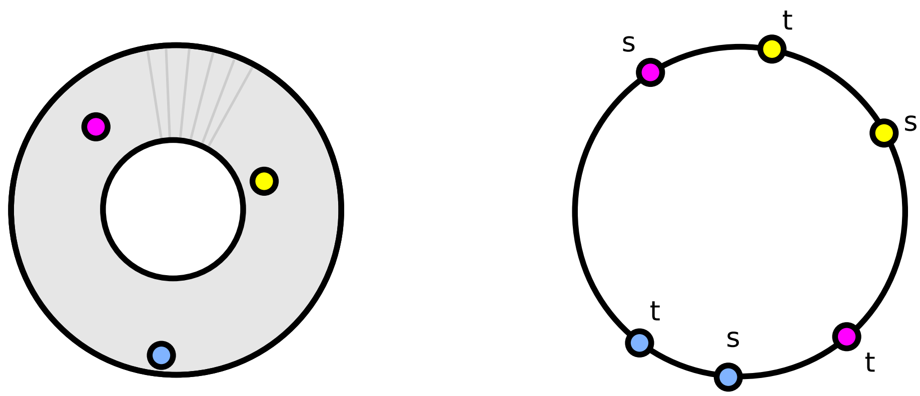

Figure 3.6: The operation $\overrightarrow{u}_s \rightsquigarrow \overrightarrow{u}_t$.

Very near the inner boundary of the annulus, (yellow) points correspond to pairs (s,t) where t is just ahead of s. Right in the middle of the annulus (magenta) points corredspond to pairs (s,t) which are diametrically opposed on the circle. Very near the outer boundary of the annulus, (blue) points correspond to pairs (s,t) where t is just behind s.

The left side of Figure 3.7 shows the same annulus. This time we have decorated the boundaries to indicate the behavior of g and h. We can think of the pair (g,h) as a *mapping* from the annulus to the plane. If we put our finger at a point (s,t) in the annulus, then $(g(s,t),h(s,t))$ is a point in the plane.

We are trying to show that there is some point (s,t) in the annulus where $g(s,t)=h(s,t)=0$. Let us assume that no such point exists. We will then derive a contradiction, and the only way out of the contradiction is that the point does exist. Here we go.

On the left hand side of Figure 3.7 we have also drawn a few curves that join the boundary components of the annulus. We call such curves *crosscuts*. Corresponding to these crosscuts there are paths in the plane. Given the way g and h behave on the boundary of the annulus, **and the nonvanishing assumption** the image of a crosscut is an avatar, as we have defined it above.

Let me say a bit about the name. You could picture the annulus as a kind of heavenly domain, or a spirit world. The mapping $(s,t)\to(g(s,t),h(s,t))$ is something like a representation of the crosscut in the heavenly world in the earthly plane, and that is what an avatar is.

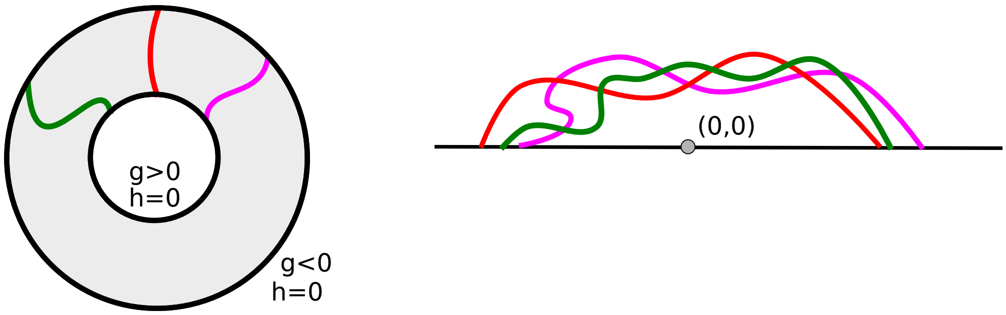

Figure 3.7: Crosscuts (left) and their avatars (right).

Our construction produces a bunch of avatars which connect the positive X-axis to the negative X-axis and avoid the origin.

3.8 The End of the Proof

Let α be any crosscut. We can write α as $\alpha(x) = (s(x), t(x))$. Here x is a parameter that varies from 0 to 1. The endpoint $\alpha(0)$ is on one boundary component and the endpoint $\alpha(1)$ is on the other. The formula for the avatar is:

$$\widehat{\alpha}(x) = (f(s(x), t(x)), g(s(x), t(x))).$$

This avatar has some winding number. Call it W_α.

Consider the new crosscut

$$\beta(x) = (t(x), s(x)).$$

We are just switching the coordinates. Since α and β are both crosscuts, we can find a continuous family of crosscuts which interpolates from α to β. We just gently slide the one into the other. This gives us a continuous family of avatars that starts at $\widehat{\alpha}$ and ends at $\widehat{\beta}$. What is wrong with this? Well, recall that $g(t, s) = -g(s, t)$ and $h(t, s) = -h(s, t)$. This means that we get the curve $\widehat{\beta}$ by negating every point of $\widehat{\alpha}$. In other words $\widehat{\alpha}$ and $\widehat{\beta}$ are opposite avatars. But we explained above that there is no continuous family of avatars that connects opposite avatars. We have a contradiction.

How do we escape the contradiction. What went wrong is that we assumed that g and h never vanish simultaneously in the annulus. So, in fact there are points (s, t) such that $g(s, t) = 0$ and $h(s, t) = 0$. Again this means that the bends (u_s, u_t) make a T-pattern.

This completes the proof of Lemma T.

4 References

[**FT**], D. Fuchs, S. Tabachnikov, *Mathematical Omnibus: Thirty Lectures on Classic Mathematics*, AMS 2007

[**HW**], B. Halpern and C. Weaver, *Inverting a cylinder through isometric immersions and embeddings*, Trans. Am. Math. Soc **230**, pp 41–70 (1977)

[**S**] R. E. Schwartz, *The optimal paper Moebius band*, arXiv 2308.12641