The Four Vertex Theorem and its Converse

Clay Shonkwiler
DRL 3E3A

October 6, 2006

The Theorems

Theorem 1 (Four Vertex Theorem). *Every simple closed plane curve has either constant curvature or the curvature function has at least two local maxima and two local minima.*

Local extrema of the curvature are called “vertices”, so the above says that curves are either circles or have at least four vertices.

Theorem 2 (Converse to the Four Vertex Theorem). *Every continuous, real-valued function on the circle S^1 which has at least two local maxima and two local minima is the curvature function of a simple closed curve in the plane.*

The Four Vertex Theorem was proved in the positive curvature (convex) case by Syamadas Mukhopadhyaya in 1909 and in the general case by Adolf Kneser in 1912. The Converse to the Four Vertex Theorem was proved in the convex case by Herman Gluck in 1971 and in the general case by Björn Dahlberg in 1997 (appeared 2005).

Why is the Four Vertex Theorem True?

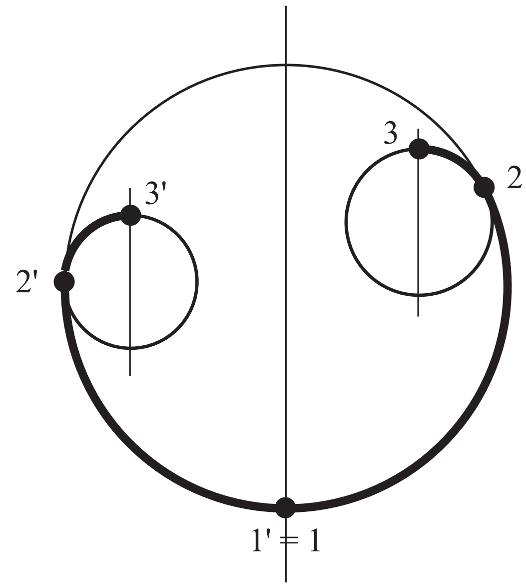

Figure 1: Trying to build a counter-example.

- A curve built from arcs of circles cannot have only two vertices and remain simple
- Using a limiting argument, this idea can be extended to a proof of the Four Vertex Theorem for convex curves

Why is simplicity a necessary assumption?

The following is an obvious example of a non-simple closed curve with only two vertices:

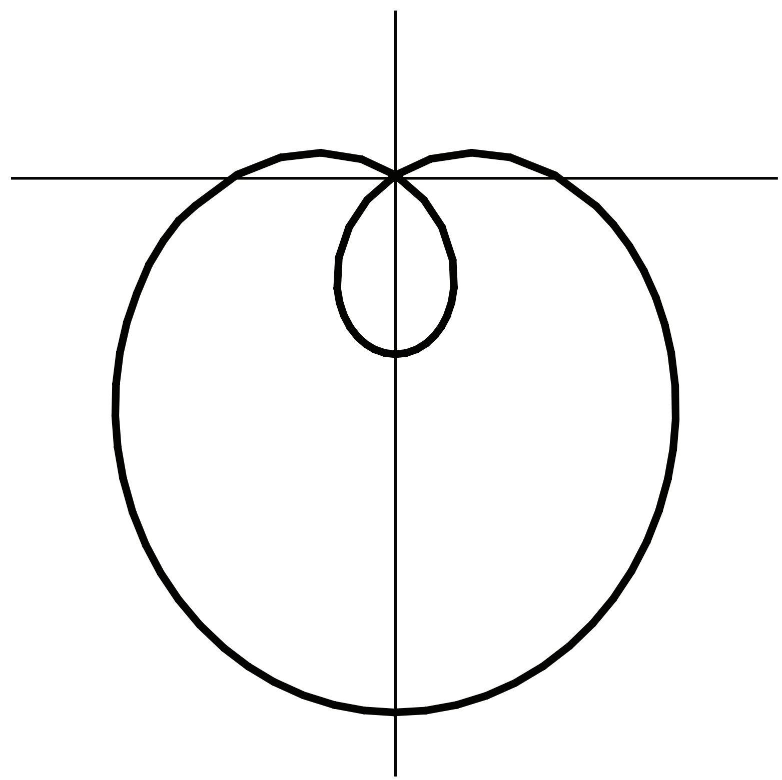

Figure 2: The curve $r = -1 - 2\sin\theta$

A simple proof of the Four Vertex Theorem

Robert Osserman's 1985 proof of the Four Vertex Theorem can be distilled into a single phrase: *consider the circumscribed circle*.

Theorem 3 (Osserman). *Let α be a simple closed curve of class C^2 in the plane, and C the circumscribed circle. If $\alpha \cap C$ has at least n components, then α has at least $2n$ vertices.*

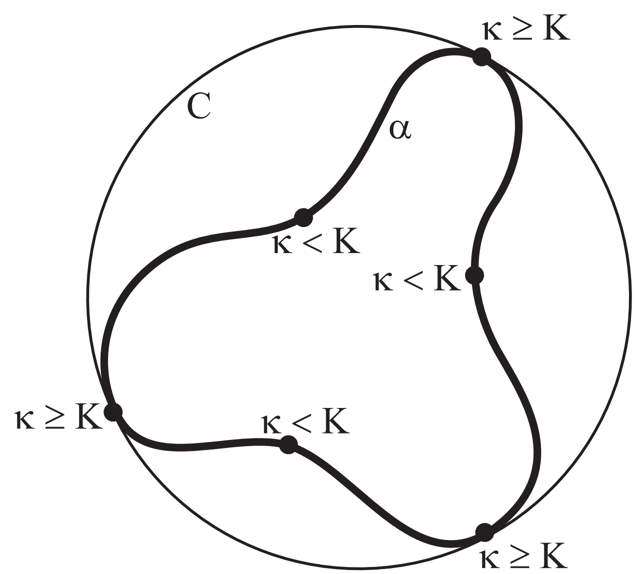

Figure 3: The curve α and its circumscribed circle.

Proof of Osserman's Theorem

To get points where $\kappa < K$, note there exists $Q \in \alpha_1$ to the "right" of the line from P_1 to P_2. Form the circle C' through P_1, Q and P_2. Then $K' < K$. Translate C' until it is tangent to α_1:

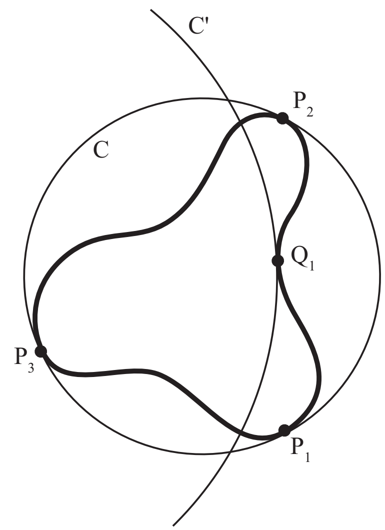

Figure 4: The circle C' translated.

Note: $\kappa(Q_1) \leq K' < K$.

This construction relies on the fact that we can pick P_1 and P_2 so that the arc between them is less than a semi-circle.

The Bonus Clause

If a component of $\alpha \cap C$ is an arc rather than a point, we are guaranteed two extra vertices, over and above the $2n$ given by Osserman's Theorem.

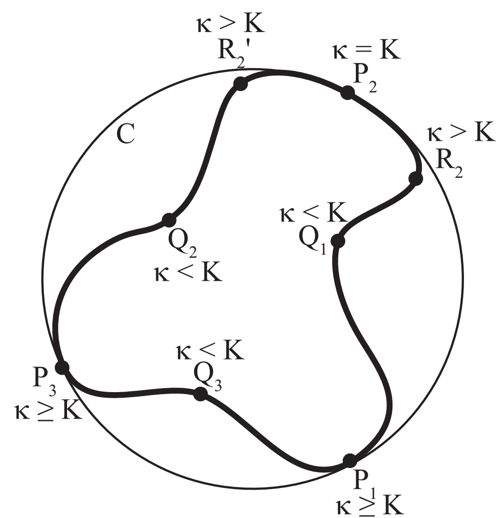

Figure 5: The bonus clause.

Here, the points R_2 and R_2' have curvature more than K, so P_2 is now a local *minimum*.

Completing the Proof: Osserman's Theorem plus the bonus clause yields the full Four Vertex Theorem because, if $\alpha \cap C$ has only one component, that component *must* be an arc (in fact, at least a semi-circle).

The Converse to the Four Vertex Theorem

Theorem 4. *Let* $\kappa : S^1 \to \mathbb{R}$ *be a continuous function which is either a nonzero constant or else has at least two local maxima and two local minima. Then there is an embedding* $\alpha : S^1 \to \mathbb{R}^2$ *whose curvature at the point* $\alpha(t)$ *is* $\kappa(t)$ *for all* $t \in S^1$*.*

Basic idea in the convex case

Let $\kappa : S^1 \to \mathbb{R}$ be continuous and strictly positive and think of the parameter on S^1 as the angle of inclination θ of the desired curve α.

It's easy to see that the curve $\alpha : [0, 2\pi] \to \mathbb{R}^2$ with curvature $\kappa(\theta)$ at the point $\alpha(\theta)$ is given by

$$\alpha(\theta) = \int_0^\theta \frac{(\cos\theta, \sin\theta)}{\kappa(\theta)} d\theta.$$

In general, α will not close up, so we define the **error vector** $E = \alpha(2\pi) - \alpha(0)$ which measures the failure of α to close up.

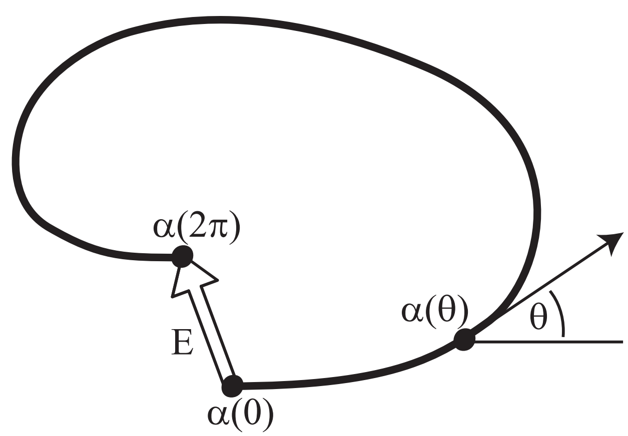

Figure 6: The error vector E.

So what does this get us?

If the curvature function κ has at least two local maxima and two local minima, we want to find a loop of diffeomorphisms h of the circle so that when we construct curves $\alpha_h : [0, 2\pi] \to \mathbb{R}^2$ whose curvature at the point $\alpha_h(\theta)$ is $\kappa \circ h(\theta)$, the corresponding error vectors $E(h)$ will wind once around the origin. Furthermore, we will do this so that the loop is contractible in the group $\mathrm{Diff}(S^1)$ of diffeomorphisms of the circle, and conclude that for some h "inside" this loop, the curve α_h will have error vector $E(h) = 0$, and hence close up to form a smooth simple closed curve.

Its curvature at the point $\alpha_h(\theta)$ is $\kappa \circ h(\theta)$, so if we write $\alpha_h(\theta) = \alpha_h \circ h^{-1} \circ h(\theta)$, and let $h(\theta) = t$, then its curvature at the point $\alpha_h \circ h^{-1}(t)$ is $\kappa(t)$. Therefore $\alpha = \alpha_h \circ h^{-1}$ is a reparametrization of the same curve, whose curvature at the point $\alpha(t)$ is $\kappa(t)$.

The proof in one picture

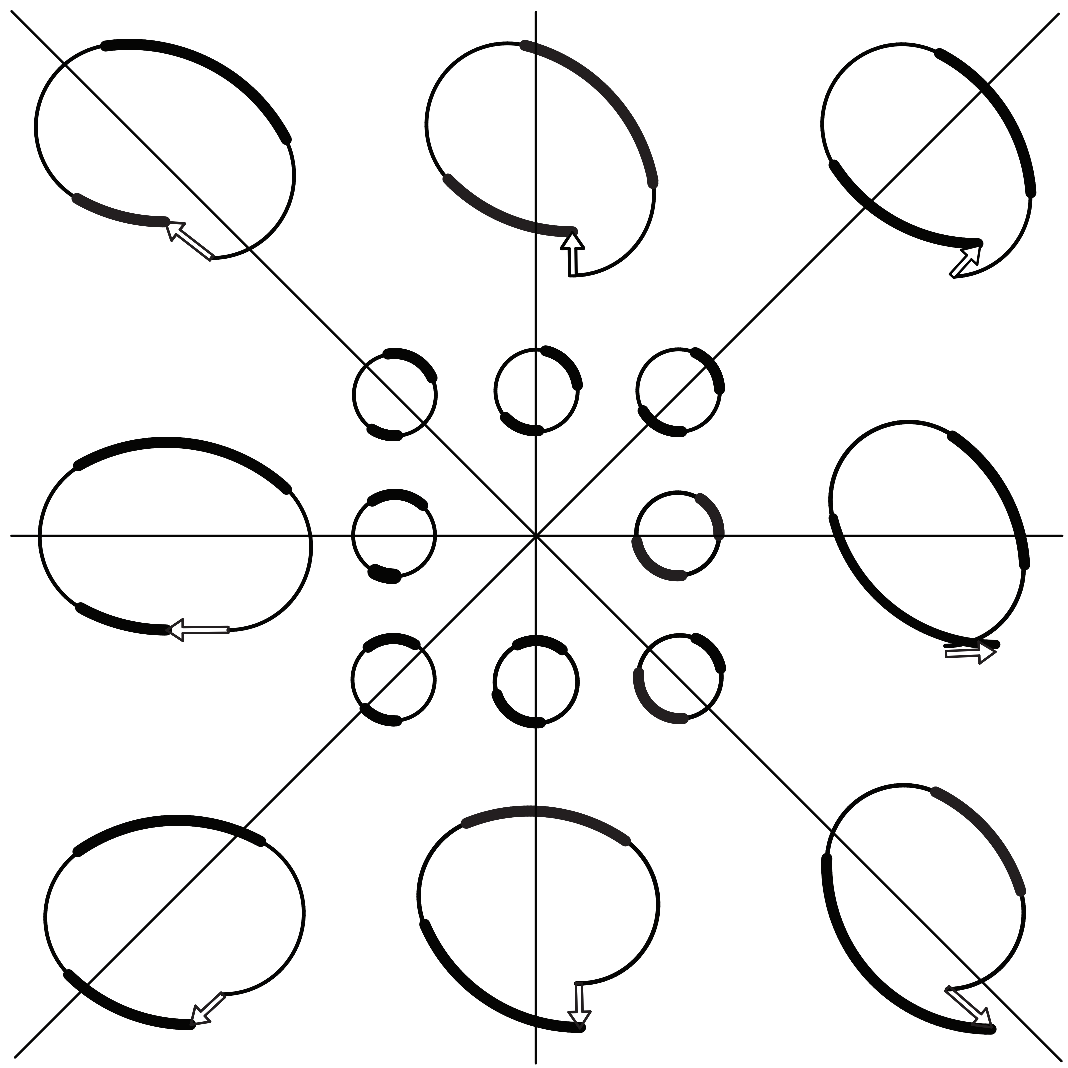

Figure 7: A curve tries unsuccessfully to close up.

How do we find this loop of diffeomorphisms?

We want to replace κ by a simpler function that is "close" to κ. The most naïve idea is to approximate κ by a step function.

Since κ is strictly positive and has at least two local maxima and two local minima, there exist $0 < a < b$ such that κ takes values a, b, a, b at four points in order around the circle.

Precompose κ by a diffeomorphism h_1 so that $\kappa \circ h_1$ is ϵ-close in measure to a step function κ_0 taking values a, b, a, b on arcs of length 2π.

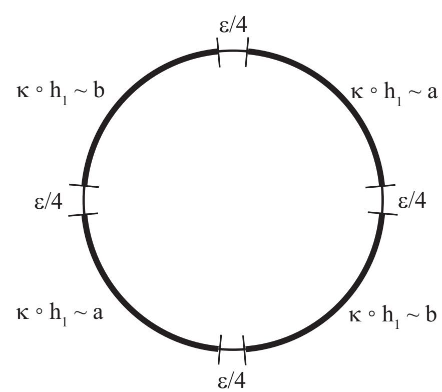

Figure 8: $\kappa \circ h_1$ is ϵ-close in measure to the step function κ_0.

Note: $\kappa \circ h_1$ and κ_0 are C^1-close

A snake tries to eat its tail

We consider a subset of diffeomorphisms of the circle, eight of which are depicted here:

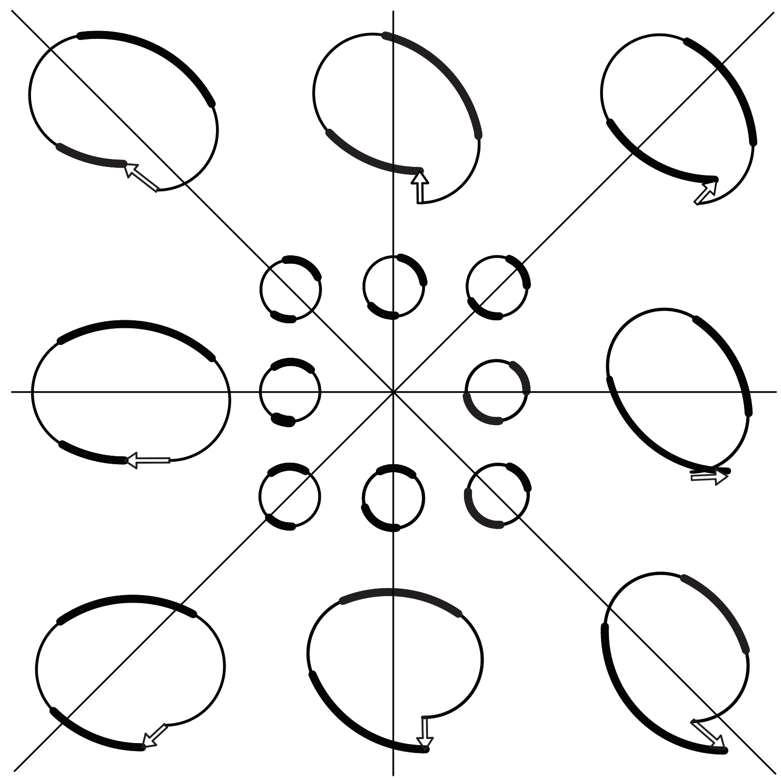

Figure 9: A curve tries unsuccessfully to close up.

Since each of these diffeomorphisms leaves the south pole fixed, this loop in $\mathrm{Diff}(S^1)$ is contractible.

Completing the convex case

Since this loop of diffeomorphisms is contractible in $\mathrm{Diff}(S^1)$, it follows that for some $h \in \mathrm{Diff}(S^1)$, the corresponding curve with curvature $\kappa_0 \circ h$ has error vector zero, so it closes up. Of course, this is obvious: if $h = \mathrm{Id}$, then the below curve with curvature κ_0 closes up.

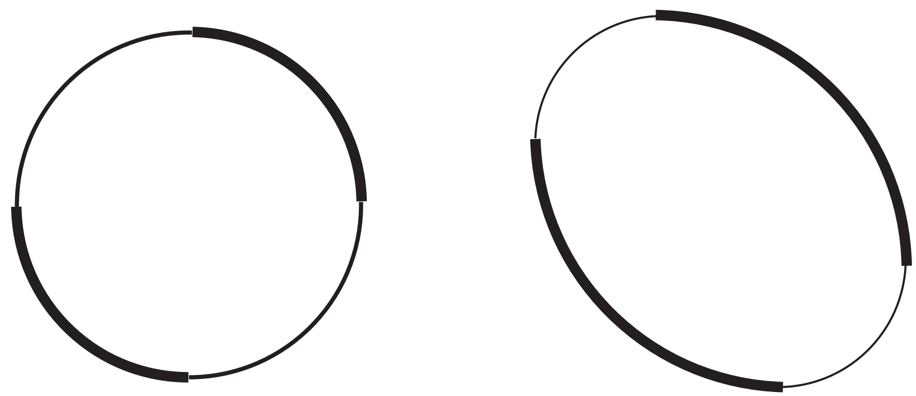

Figure 10: Preassign κ_0 and.....get the "bicircle".

The point, though, is that this argument is robust and, since the curves are C^1-close, applies equally well to the curvature function $\kappa \circ h_1$. Therefore, there is some diffeomorphism h of S^1 so that the curve with curvature $\kappa \circ h_1 \circ h$ is a strictly convex simple closed curve. Reparametrizing as discussed above shows that this curve has curvature κ, completing the proof in the convex case.

The full converse

When we envision arbitrary simple closed curves, it's natural to think of curves like these:

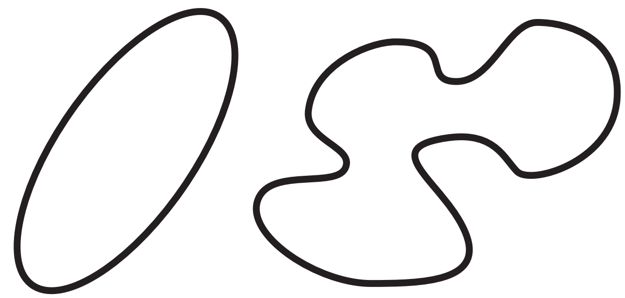

Figure 11: Two plane curves

Dahlberg's key idea, though, was that, by precomposing with a suitable diffeomorphism, an arbitrary curve can be made to look like this:

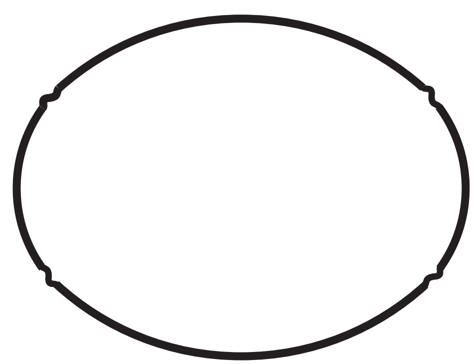

Figure 12: We should envision arbitrary curves like this

Dahlberg's proof strategy

The plan is to adapt the winding number argument from the convex case, especially the important role played by step functions.

To make the winding number argument work, Dahlberg exhibits a 2-cell $\mathcal{D} \subset \mathrm{Diff}(S^1)$ centered at the origin and satisfying a certain transversality condition which guarantees the winding number argument works on arbitrarily small loops in $\mathcal{D}$ about its center.

Configuration Space

We want to define a configuration space of four ordered points on a circle to help visualize the situation.

- Changing the sign of κ if necessary, the fact that κ has two local maxima and two local minima ensures that there exist $0 < a < b$ such that κ takes on values a, b, a, b at four points in order along the circle.
- Find a preliminary diffeomorphism h_1 of S^1 so that $\kappa \circ h_1$ is ϵ-close in measure to the same step function κ_0 we defined in the convex case.
- We focus on κ_0 and its compositions $\kappa_0 \circ h$ as h ranges over $\mathrm{Diff}(S^1)$. These are all step functions with values a, b, a, b on four arcs determined by some points p_1, p_2, p_3, p_4. Making scalings where appropriate to make the total curvature equal 2π, construct a curve from circular arcs with curvatures proportional to a, b, a, b and let $E(p_1, p_2, p_3, p_4)$ denote the error vector for this curve.
- Let CS denote the configuration space of order 4-tuples (p_1, p_2, p_3, p_4) of distinct points on S^1.

CS and its core

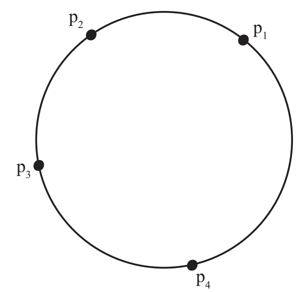

Figure 13: An element of the configuration space CS.

Note: CS is diffeomorphic to $S^1 \times \mathbb{R}^3$.

- The error vector $E(p_1, p_2, p_3, p_4)$ defines a map $E : CS \to \mathbb{R}^2$. When E vanishes, the corresponding curve closes up. We call the kernel of E the *core* of CS, denoted CS_0.

- It's easy to see that such curves close up if and only if opposite arcs are equal in length, so $(p_1, p_2, p_3, p_4) \in CS$ lies in CS_0 if and only if p_1 and p_3 are antipodal and p_2 and p_4 are antipodal, which occurs precisely when $p_1 - p_2 + p_3 - p_4 = 0$ in the complex plane.

- CS_0 is diffeomorphic to $S^1 \times \mathbb{R}$.

The reduced configuration space

Let $RCS \subset CS$ be the subset where $p_1 = 1$. Then $RCS \simeq \mathbb{R}^3$ and we have a homeomorphism $S^1 \times RCS \to CS$ given by

$$\left(e^{i\theta}, (1, p, q, r)\right) \mapsto \left(e^{i\theta}, e^{i\theta}p, e^{i\theta}q, e^{i\theta}r\right)$$

Now, change coordinates by writing

$$p = e^{2\pi i x}, \qquad q = e^{2\pi i y}, \qquad r = e^{2\pi i z}.$$

Then RCS can be represented as an open solid tetrahedron:

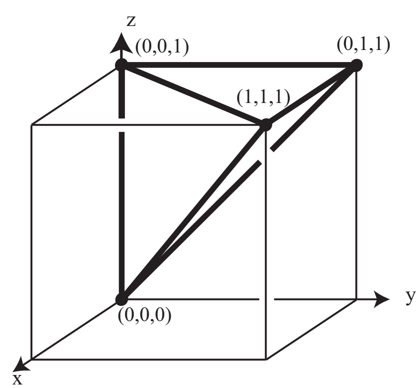

Figure 14: The reduced configuration space RCS

The core

A point $(1, p, q, r)$ in RCS is in the core if and only if 1 and q are antipodal and p and r are antipodal. Hence, $RCS_0 := RCS \cap CS_0$ is given by

$$0 < x < y = \frac{1}{2} < z = x + \frac{1}{2} < 1.$$

RCS_0 is pictured below:

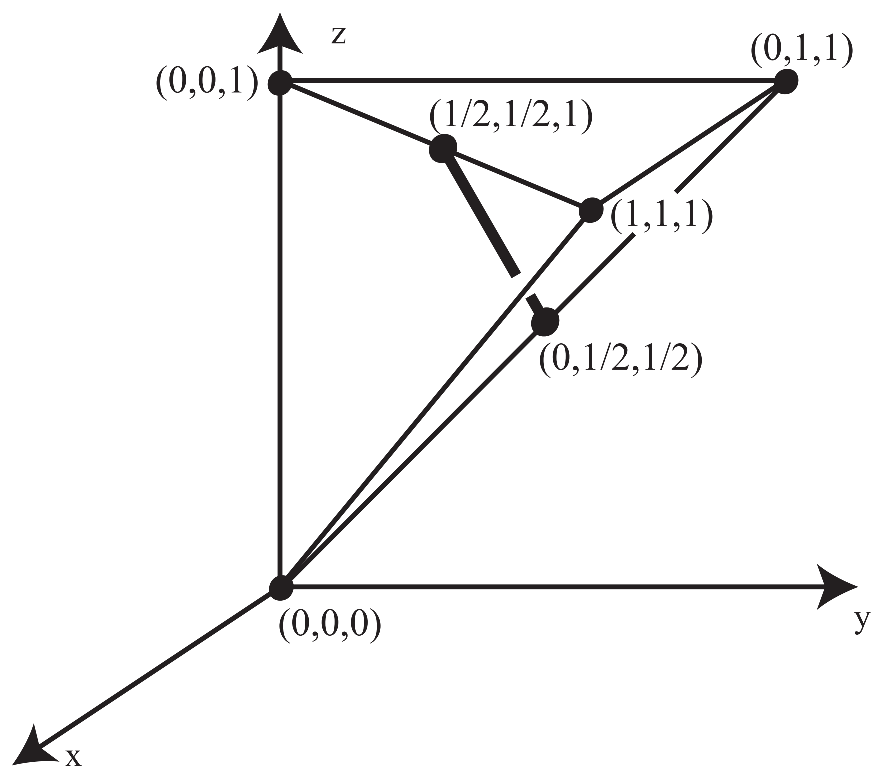

Figure 15: The core of the reduced configuration space.

Topology of the error map

Let λ be a loop in $CS - CS_0$ which bounds a disc in CS and links CS_0 once (pictured below in RCS). Then we have the following two propositions about λ:

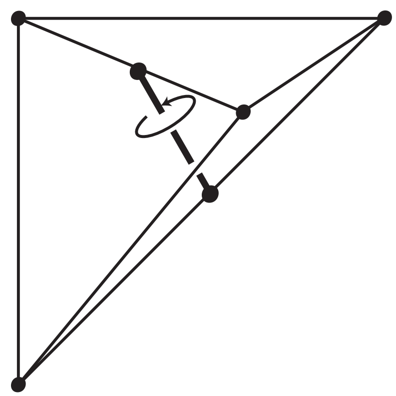

Figure 16: A loop in the reduced configuration space.

Proposition 5. *$E(\lambda)$ has winding number ± 1 about the origin in $\mathbb{R}^2$.*

This proposition says that plane curves built as described above are capable of exhibiting the winding number phenomenon which makes the proof work. It follows directly from:

Proposition 6. *The differential of $E : RCS \to \mathbb{R}^2$ is surjective at each point of RCS_0.*

Dahlberg's disk $\mathcal{D}$

We restrict our attention to diffeomorphisms lying in the disk $\mathcal{D} \subset \mathrm{Diff}(S^1)$ consisting of the special Möbius transformations

$$g_\beta(z) = \frac{z - \beta}{1 - \bar{\beta} z}$$

where $|\beta| < 1$.

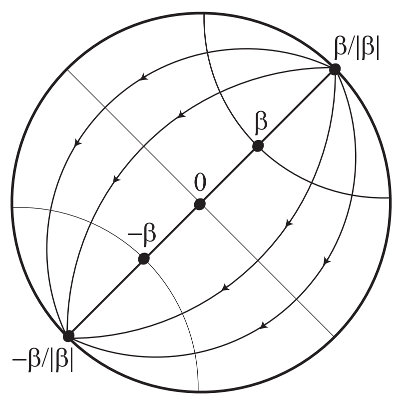

Figure 17: The action of g_β on the unit disk in the complex plane.

These maps are all isometries of the disk model of the hyperbolic plane. g_0 is the identity and, for $\beta \neq 0$, g_β is a hyperbolic translation along the line through 0 and β taking β to 0. The points $\frac{\beta}{|\beta|}$ and $-\frac{\beta}{|\beta|}$ on the circle at infinity are the only fixed points of g_β.

Another transversality result

If $g_\beta \in \mathcal{D}$ and $P = (p_1, p_2, p_3, p_4) \in CS$, then we define

$$g_\beta(P) = (g_\beta(p_1), g_\beta(p_2), g_\beta(p_3), g_\beta(p_4))$$

so that $\mathcal{D}$ acts on CS.

Proposition 7. *The evaluation map* $CS_0 \times \mathcal{D} \to CS$ *given by* $(P, g_\beta) \mapsto g_\beta(P)$ *is a diffeomorphism.*

Corollary 8. *For each fixed point* $P \in CS_0$*, the evaluation map* $g_\beta \mapsto g_\beta(P)$ *is a smooth embedding of Dahlberg's disk* $\mathcal{D}$ *into* CS *which meets the core transversally at* P *and nowhere else.*

The image of $\mathcal{D}$ in RCS

If we fix the point $P = (1, i, -1, -i) \in CS$ and compose the map $\mathcal{D} \to CS$ with the projection $CS \to RCS$, the following are two pictures of the image of $\mathcal{D}$ in RCS:

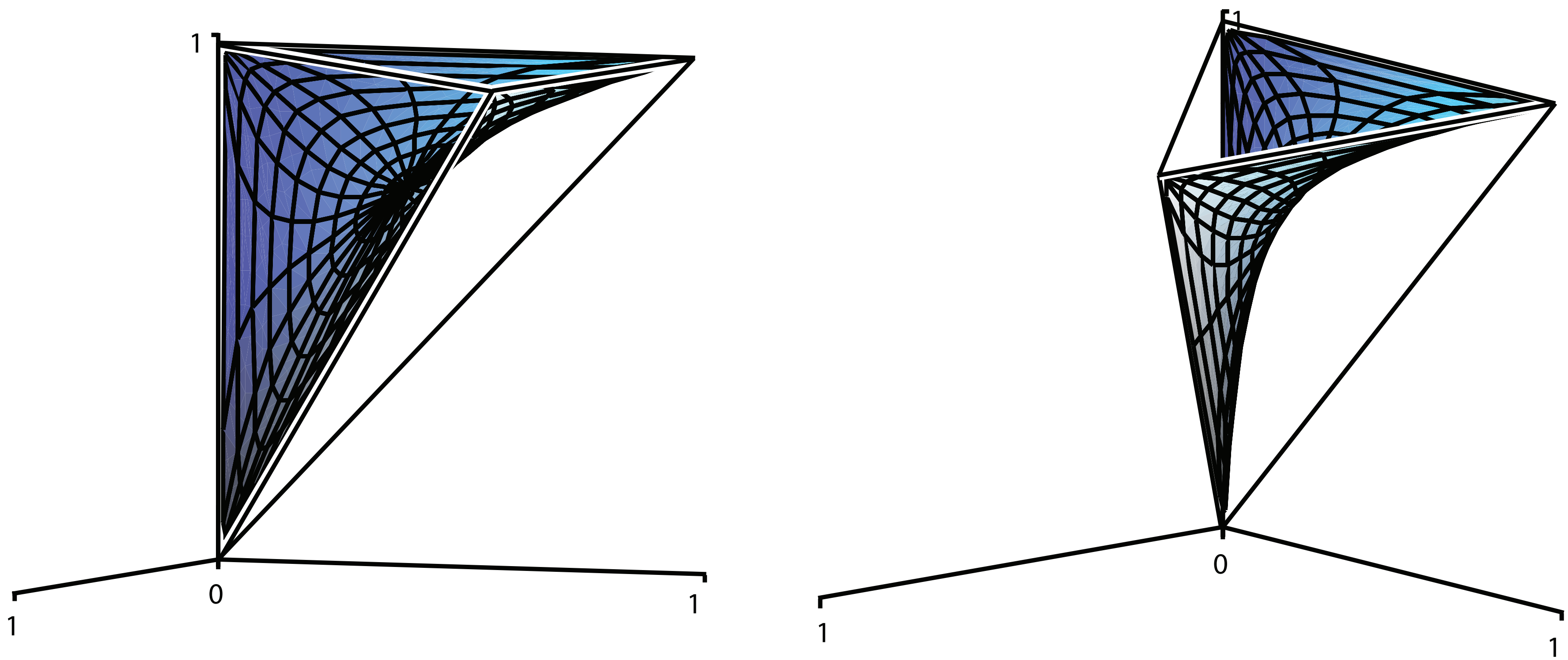

Figure 18: Two views of Dahlberg's disk.

We can see that the image of $\mathcal{D}$ really does meet the core transversally.

Grand finale

Start with a continuous, preassigned curvature function $\kappa : S^1 \to \mathbb{R}$ which has at least two local maxima and two local minima. We want to find an embedding $\alpha : S^1 \to \mathbb{R}^2$ with curvature $\kappa(t)$ at each point $\alpha(t)$.

Changing the sign of κ if necessary, there are real numbers $0 < a < b$ and four points on S^1 in counterclockwise order where κ takes the values a, b, a, b in succession.

The points $1, i, -1, -i$ divide S^1 into four equal arcs of length $\pi/2$. Let κ_0 be the step function taking values a, b, a, b on these arcs.

Given $\epsilon > 0$, we find a preliminary diffeomorphism h_1 of S^1 such that $\kappa \circ h_1$ is ϵ close in measure to κ_0. Rescale both to achieve total curvature 2π and they will still be ϵ close in measure for a new small ϵ.

Apply the winding number argument to κ_0

Consider $P_0 = (1, i, -1, -i) \in CS_0$ and map $\mathcal{D}$ to CS by sending $g_\beta \mapsto g_\beta(P_0)$. By Corollary 8, this map is a smooth embedding which meets CS_0 transversally at P_0 and nowhere else.

By Proposition 5, each loop $|\beta| =$ constant in $\mathcal{D}$ is sent to a loop in $\mathbb{R}^2 - \{0\}$ with winding number ± 1 about the origin.

More concretely, let $c(\beta)\kappa_0 \circ g_\beta$ be the rescaling of the curvature step function which has total curvature 2π and let $\alpha(\beta) : [0, 2\pi] \to \mathbb{R}^2$ be the corresponding arc-length parametrized curve with this curvature function. As β circles once around the origin, the corresponding loop of error vectors $E(\alpha(\beta))$ has winding number ± 1 about the origin.

Transfer this winding number argument to κ

Let $c(h_1, \beta)\kappa \circ h_1 \circ g_\beta$ be the rescaling of the curvature function $\kappa \circ h_1 \circ g_\beta$ which has total curvature 2π. Let $\alpha(h_1, \beta) : [0, 2\pi] \to \mathbb{R}^2$ be the corresponding arc-length parametrized curve with this curvature function.

Fixing $|\beta|$, we can choose ϵ sufficiently small so that $\alpha(h_1, \beta)$ is C^1-close to $\alpha(\beta)$, the curve corresponding to the step function κ_0. Then as β circles around the origin, the loop of error vectors $E(\alpha(h_1, \beta))$ will also have winding number ± 1 about the origin.

Therefore, there exists a diffeomorphism $g_{\beta'}$ with $|\beta'| \leq |\beta|$ so that $E(\alpha(h_1, \beta')) = 0$, so the curve $\alpha(h_1, \beta')$ closes up smoothly. If $|\beta|$ sufficiently small, then $\alpha(h_1, \beta')$ will be as close as we like to the fixed bicircle with curvature $c_0 \kappa_0$ and thus will be simple.

The simple closed curve $\alpha(h_1, \beta')$ realizes the curvature function $c(h_1, \beta')\kappa \circ h_1 \circ g_{\beta'}$. Rescaling realizes the curvature function $\kappa \circ h_1 \circ g_{\beta'}$ and thus, after reparametrization, it realizes the curvature function κ, completing the proof of the converse of the Four Vertex Theorem.

Further Reading

Content and images shamelessly stolen from "The Four Vertex Theorem and its Converse", by Dennis DeTurck, Herman Gluck, Daniel Pomerleano and Shea Vick, available on the arXiv as math.DG/0609268.

Thanks!