

THE FOUR VERTEX THEOREM AND ITS CONVERSE

SHEA VICK

Everything presented here represents joint work with D. DeTurck, H. Gluck and D. Pomerleano.

WHAT IS IT AND WHY DO WE CARE?

Theorem 1 (Four Vertex Theorem). *A simple closed curve in the plane, other than a circle, must have at least four “vertices”, that is, at least four points where the curvature has a local maximum or local minimum.*

Theorem 2 (The Converse to the Four Vertex Theorem). *Let $\kappa : S^1 \rightarrow \mathbb{R}$ 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 $\alpha : S^1 \rightarrow \mathbb{R}^2$ whose curvature at the point $\alpha(t)$ is $\kappa(t)$ for all $t \in S^1$.*

The Four Vertex Theorem is one of the oldest results in global differential geometry.

- First proved in convex case by Mukhopadhyaya (1909).
- Kneser (1912) relaxed the convexity condition.
- Gluck (1971) proved the converse in the strictly convex case.
- Dahlberg (1997) proved the full converse, without the convexity restriction.
- Over the years it has been extended and generalized by many other mathematicians.

1. WHY IS THE FOUR VERTEX THEOREM TRUE?

Let’s not give a detailed proof of the Four Vertex Theorem, but we should at least convince ourselves that such a theorem is “inevitable”.

- A counter-example would be a simple closed curve in the plane whose curvature is nonconstant, has one minimum and one maximum, and is weakly monotonic on the two arcs between them.
 - Lets try to build such a curve from a few arcs of circles, fail, and see why.

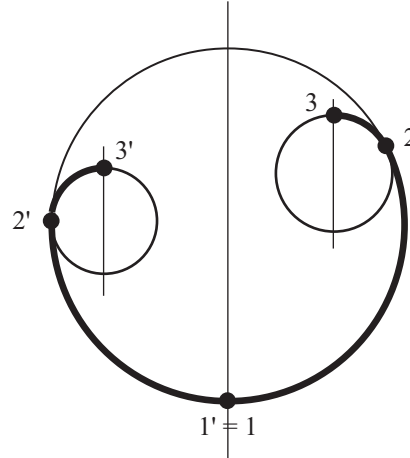
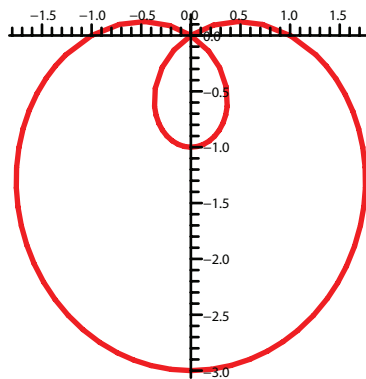


FIGURE 1. A failed counterexample.

- Think of the circles in Figure 1 as train tracks, switching from one track to another at the labeled junction points.
 - * Note that the vertical diameter of this smaller circle is displaced to the right (resp. left) of the vertical diameter of the original circle.
- To form a convex simple closed curve, then the top points 3 and 3' must coincide...but they don't. So the curve does not exist.
- Notice that no matter how many circles we use, we get the same contradiction.
- If we permit a single self-intersection, then it's easy to get just one maximum and one minimum for the curvature, as shown in Figure 2.

FIGURE 2. The (polar) graph of $r = -1 - 2 \sin \theta$.

PROVING THE CONVERSE IN THE CONVEX CASE.

- The basic idea is to use a winding number argument in the group of diffeomorphisms of the circle.
 - We begin with an arbitrary, continuous, strictly positive candidate curvature function $\kappa : S^1 \rightarrow \mathbb{R}$, satisfying the min-max-min-max condition imposed by the four vertex theorem.

* Since the curves we build will usually not close up, it is better to cut S^1 open into the interval $[0, 2\pi]$.

- There is a unique map $\alpha : [0, 2\pi] \rightarrow \mathbb{R}^2$ which begins at the origin and has unit tangent vector $(\cos \theta, \sin \theta)$ and curvature $\kappa(\theta)$ at the point $\alpha(\theta)$. In fact, if s is arc length along this curve, then the equations

$$\frac{d\alpha}{ds} = (\cos \theta, \sin \theta) \quad \text{and} \quad \frac{d\theta}{ds} = \kappa(s)$$

quickly lead to the explicit formula

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

- From this we can define an **Error Vector**, $E = \alpha(2\pi) - \alpha(0)$, which measures the failure of our curve to close up.

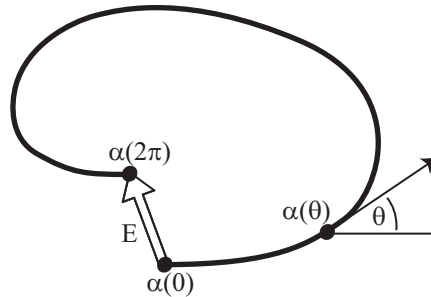


FIGURE 3. The error vector E .

- We find a contractible loop of diffeomorphisms h of the circle so that when we construct, as above, curves $\alpha_h : [0, 2\pi] \rightarrow \mathbb{R}^2$ 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.

* Contractibility implies 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.

- Finding the loop of diffeomorphisms.
 - Since κ is a strictly positive function with alternating local minima and maxima, we can find positive numbers $b > a > 0$ so that κ takes on the values a, b, a, b in succession.
 - We want to “replace” our curvature function κ with a simpler curvature function κ_0 .
 - * κ_0 will be the step function with values a, b, a, b on the arcs $(0, \pi/2)$, $(\pi/2, \pi)$, $(\pi, 3\pi/2)$, $(3\pi/2, 2\pi)$ respectively.
 - * To do this, we find a diffeomorphism h_1 of S^1 , so that $\kappa \circ h_1$ is “ ϵ -close in measure” to κ_0 .

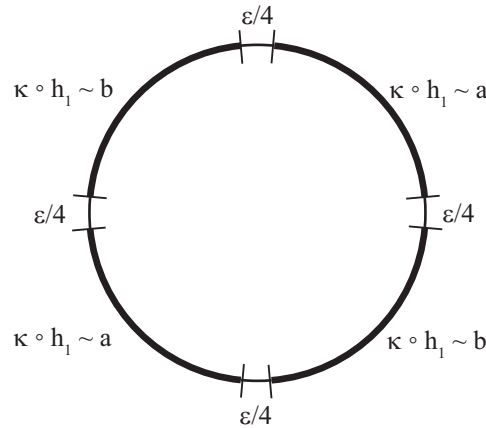


FIGURE 4. $\kappa \circ h_1$ is ϵ -close in measure to the step function κ_0 .

- * Since the curvature functions $\kappa \circ h_1$ and κ_0 are ϵ -close in measure, the corresponding curves are C^1 -close. This allows us to shift our attention to the simpler function κ_0 .
- * We find a contractible loop of diffeomorphisms h of the circle so that the error vectors $E(h)$ for the curvature step functions $\kappa_0 \circ h$ wind once around the origin.

• The proof in one picture.

- Since each of the curves in Figure 5 leaves the south pole fixed, this loop is contractible in $\text{Diff}(S^1)$. Thus for some h , we have that $E(h) = 0$.
- This is hardly news, for $h = \text{id}$, we have $E(h) = 0$.

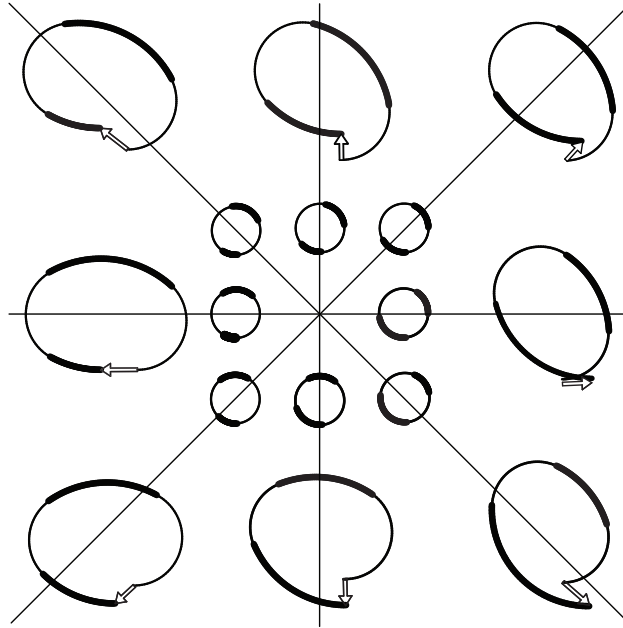


FIGURE 5. A curve tries unsuccessfully to close up.

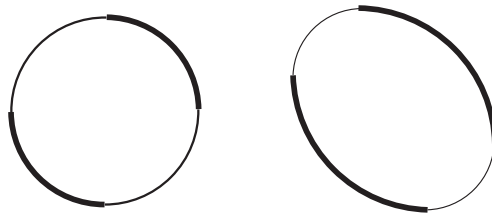


FIGURE 6. Preassign κ_0and get this “bicircle”.

- The point is that this argument is robust, and applies equally well to the curvature function $\kappa \circ h_1$, which is ϵ -close in measure to κ_0 .

PROVING THE CONVERSE IN THE NON-CONVEX CASE.

- The overall plan is to follow the winding number argument from the strictly positive curvature case, but being mindful of the need for the following changes.
 - A smooth simple closed curve with strictly positive curvature has a distinguished parametrization by the angle of inclination of its tangent line. This doesn't work in general, so another parametrization scheme is needed.

- A smooth closed curve with strictly positive curvature whose tangent line turns through an angle of 2π is automatically simple. This is false without the curvature assumption, and part of Dahlberg's Key Idea is to force the curve to be C^1 -close to a fixed convex curve in order to make it simple.

When asked to draw a simple closed curve in the plane with strictly positive curvature, and another one with mixed positive and negative curvature, a typical response might be as shown in Figure 7.

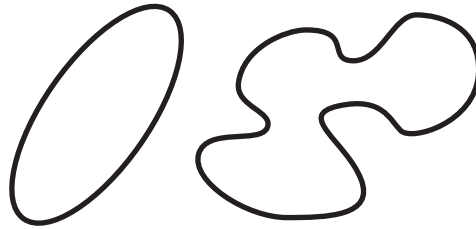


FIGURE 7. Typical Curves.

But for mixed positive and negative curvature, Dahlberg envisioned the curve shown in Figure 8.

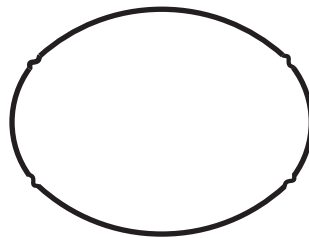


FIGURE 8. Dahlberg's Vision.

Its four major subarcs are almost circular. They are connected by four small wiggly arcs, each of which has an almost constant tangent direction, but largely varying curvature, including negative curvature.

Dahlberg's Key Idea. *You can construct such a curve with any preassigned curvature which has at least two local maxima and two local minima. You can use the winding number argument to get it to close up smoothly, and you can also make it C^1 -close to a fixed convex curve, which will imply that it is simple.*

- We would like to define the configuration space CS of ordered 4-tuples of distinct points on the unit circle to help visualize the arguments to come. This space is diffeomorphic to $S^1 \times \mathbb{R}^3$.

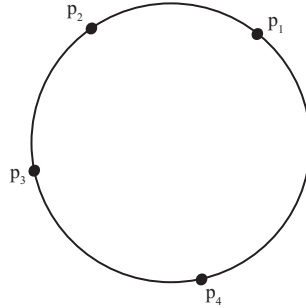


FIGURE 9. An element of the configuration space CS .

- As in the positive curvature case, there are positive numbers $0 < a < b$ so that, after possibly changing its sign, κ takes the values a, b, a, b at four points in order around the circle, and a preliminary diffeomorphism h_1 of the circle so that the function $\kappa \circ h_1$ is ϵ -close in measure to the step function κ_0 with values a, b, a, b on the arcs $(0, \pi/2)$, $(\pi/2, \pi)$, $(\pi, 3\pi/2)$, $(3\pi/2, 2\pi)$.
- We again focus on the step function κ_0 and its compositions $\kappa_0 \circ h$ as h ranges over $\text{Diff}(S^1)$. These are all step functions with the values a, b, a, b on the four arcs determined by some four points p_1, p_2, p_3, p_4 in order on S^1 .
- For each such step function, we construct a curve from circular arcs with curvatures proportional to a, b, a, b , scaled up or down to make the total curvature 2π . Let $E(p_1, p_2, p_3, p_4)$ denote the error vector for this curve.
- The above gives us an **error map** $E : CS \rightarrow \mathbb{R}^2$.
 - The vanishing of the error vector signals that our curve closes up. We call the set of such points the **Core**, and denote it by CS_0 .
 - We leave it as an exercise to show that a point $p = (p_1, p_2, p_3, p_4)$ is in the core if and only if p_1 and p_3 are antipodal, and p_2 and p_4 are antipodal.
 - The core CS_0 is diffeomorphic to $S^1 \times \mathbb{R}$.
- To further aid in visualization, we define the **Reduced Configuration Space** (RCS) to be the set of points in CS with $p_1 = 1$.

- RCS is diffeomorphic to \mathbb{R}^3 , and we use the group structure on S^1 to define the diffeomorphism $S^1 \times RCS \rightarrow CS$ by

$$(e^{i\theta}, (1, p, q, r)) \rightarrow (e^{i\theta}, e^{i\theta}p, e^{i\theta}q, e^{i\theta}r).$$

- For the purpose of drawing pictures, we change coordinates by writing

$$p = e^{2\pi ix}, \quad q = e^{2\pi iy} \quad \text{and} \quad r = e^{2\pi iz}.$$

Then $RCS \cong \{(x, y, z) : 0 < x < y < z < 1\}$.

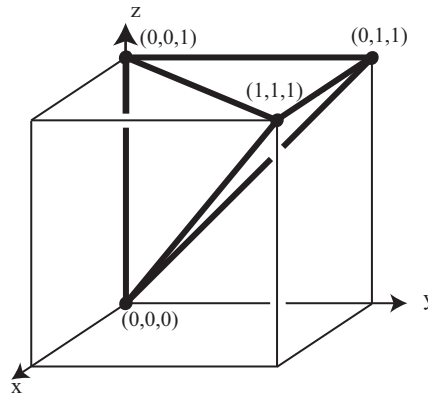


FIGURE 10. The reduced configuration space RCS .

- The core intersects RCS in the open line segment connecting $(0, 1/2, 1/2)$ to $(1/2, 1/2, 1)$.

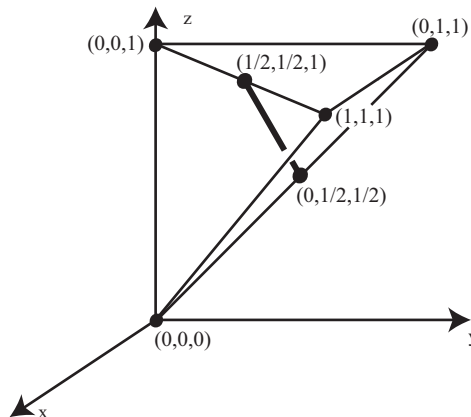


FIGURE 11. The core of the reduced configuration space.

THE TOPOLOGY OF THE ERROR MAP.

Let λ be a loop in $CS - CS_0$ which is null-homologous in CS but links the core CS_0 once. For convenience, we go down one dimension and show this loop in $RCS - RCS_0$.

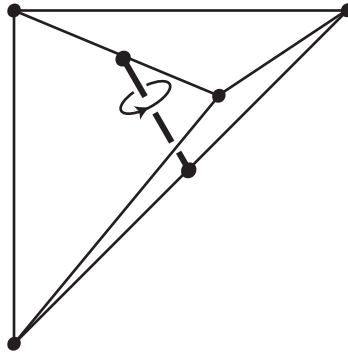


FIGURE 12. A loop in the reduced configuration space.

We have the following two propositions about λ .

Proposition 3. *The image of the loop λ under the error map E has winding number ± 1 about the origin in \mathbb{R}^2 .*

Proposition 4. *The differential of the error map $E : RCS \rightarrow \mathbb{R}^2$ is surjective at each point of the core.*

DAHLBERG'S DISK \mathcal{D} .

Dahlberg's choice of 2-cell $\mathcal{D} \subset \text{Diff}(S^1)$ consists of the *special Möbius transformations*

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

where $|\beta| < 1$ and $\bar{\beta}$ is the complex conjugate of β .

THE SECOND TRANSVERSALITY RESULT.

Let g_β be a point in Dahlberg's disk \mathcal{D} , and $P = (p_1, p_2, p_3, p_4)$ a point in the configuration space CS . We define

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

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

Proposition 5. *The evaluation map $CS_0 \times \mathcal{D} \rightarrow CS$ defined by $(P, g_\beta) \rightarrow g_\beta(P)$ is a diffeomorphism.*

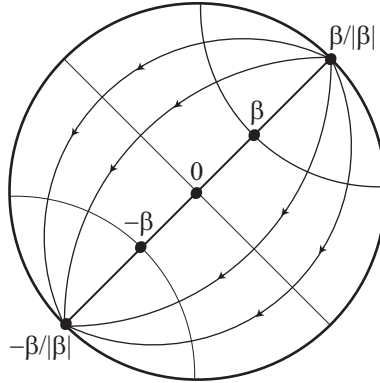


FIGURE 13. The action of g_β on the unit disk in the complex plane.

Corollary 6. *For each fixed point P in the core CS_0 , the evaluation map $g_\beta \rightarrow g_\beta(P)$ is a smooth embedding of Dahlberg's disk \mathcal{D} into CS which meets the core transversally at the point P and nowhere else.*

THE IMAGE OF DAHLBERG'S DISK IN THE REDUCED CONFIGURATION SPACE.

If we start with the point $P_0 = (1, i, -1, -i)$, and then follow the above map by the projection of CS to RCS , we get the correspondence

$$g_\beta \rightarrow (1, g_\beta(1)^{-1}g_\beta(i), g_\beta(1)^{-1}g_\beta(-1), g_\beta(1)^{-1}g_\beta(-i)).$$

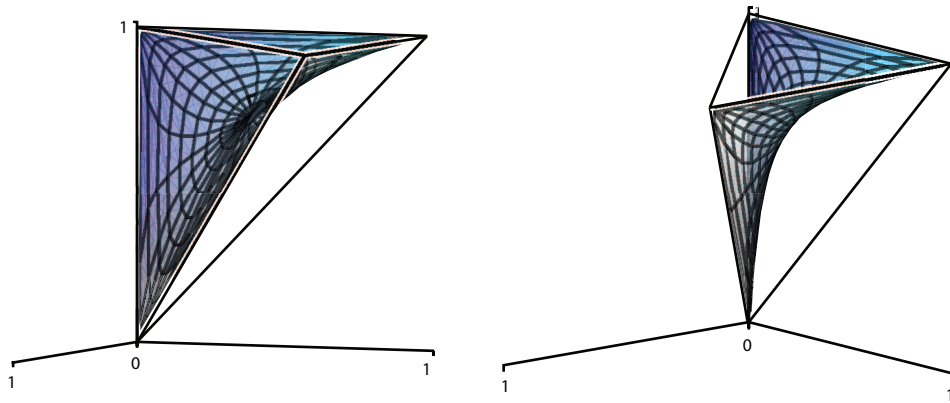


FIGURE 14. Dahlberg's disk \mathcal{D} mapped into the tetrahedron - two views.

PUTTING EVERYTHING TOGETHER.

- We apply a preliminary diffeomorphism h_1 and temporarily replace κ by a curvature step function κ_0 , which is ϵ -close in measure to $\kappa \circ h_1$.
 - κ_0 takes the values a, b, a, b ($b > a > 0$) along the four quarter-circles with end points at $1, i, -1, -i$.
- We apply the winding number argument to the curvature step function κ_0 .
 - Consider the point $P_0 = (1, i, -1, -i)$ in the core CS_0 of the configuration space CS , and map Dahlberg's disk \mathcal{D} into CS by sending $g_\beta \rightarrow g_\beta(P_0)$.
 - * By Corollary 6, this evaluation map is a smooth embedding of \mathcal{D} into CS which meets the core CS_0 transversally at the point P_0 and nowhere else.
 - * By Proposition 3, each loop $|\beta| = \text{constant}$ in \mathcal{D} is sent by the composition

$$\mathcal{D} \xrightarrow{\text{evaluation map}} CS \xrightarrow{\text{error map}} \mathbb{R}^2$$
 into a loop in $\mathbb{R}^2 - \{\text{origin}\}$ with winding number ± 1 about the origin.
- Since $\kappa \circ h_1$ was ϵ -close in measure to κ_0 , we can conclude that for a sufficiently small loop $|\beta|$ in Dahlberg's disk, the corresponding error vector circles once around the origin.
 - For $|\beta|$ and ϵ sufficiently small, the corresponding curve will be C^1 -close to a bicircle, and will therefore be simple.
- Reparametrizing and rescaling, we find a curve α that realizes the curvature function κ .