

2 Winding Numbers

2.1 Let me not be pent up, sir; I will fast, being loose.

Fast and Loose is the name of a family of magic tricks (or con games) performed with ropes, chains, and belts that have been practiced since at least the 14th century; the con game is mentioned in three different Shakespeare plays. In one such trick, now sometimes called the *Endless Chain*, the con artist arranges a closed loop of chain into a doubled figure-8, and then asks the mark to put their finger on the table inside one of the loops. The con artist then pulls the chain along the table. If the chain catches on the mark's finger, then the chain is *fast* and the mark wins; if the con artist can pull the chain completely off the table, the chain is *loose* and the mark loses.

The con artist shows the mark that there are two different ways for the loops to fall. (Notice how the chain crosses itself in the lower corners.) Because the chain is bright and shiny and bumpy, it's impossible for the mark to tell which way the chain is actually arranged, but because these are the only possibilities, the mark should have a 50-50 chance of winning. Right? *Riiiiight?*

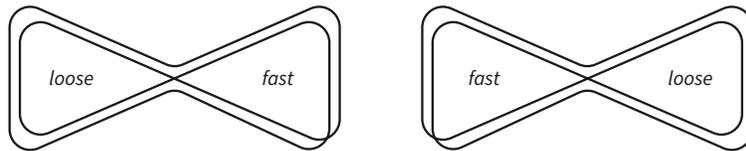


Figure 1: Two arrangements of the Endless Chain

Oh, you sweet summer child. Of course not! As soon as the mark places money *on the barrelhead*, the con artist wins every time. The con artist was lying; there is a third arrangement of the chain that is *always* loose, no matter where the mark puts their finger.¹

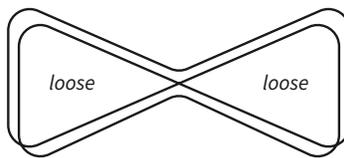


Figure 2: The actual arrangement of the Endless Chain

2.2 Shoelaces and Signed Areas

Swap this section and next? This is historical order, but the narrative is a bit clunky.

Before discussing the mathematical reasons you just lost all your money, let's consider a basic computational geometry problem: How quickly can we compute the area enclosed by a given polygon P ?

A particularly simple algorithm was described by Albrecht Meister in 1785. In principle, we can calculate the area of P by cutting P into disjoint triangles and then summing the triangle areas,

¹There is another completely different "Fast and Loose" con game, also known as called "Pricking the Garter", that's usually performed with a belt. It's less self-working, less mathematically interesting (despite *seeming* to invoke the Jordan curve theorem), and less *shiny* than the Endless Chain.

each of which can be computed in $O(1)$ time, but it would be more than a century before anyone knew how to cut polygons into triangles. Meister's insight was to consider the *signed* areas of *overlapping* oriented triangles.

The signed area of a triangle depends not only on its vertex coordinates, but on the orientation of its three vertices. By convention, counterclockwise triangles have positive signed area, and clockwise triangles have negative signed area. Recall that a triple of points (q, r, s) is oriented counterclockwise or clockwise if and only if the following determinant is positive or negative, respectively:

$$\Delta(q, r, s) = \det \begin{bmatrix} 1 & q.x & q.y \\ 1 & r.x & r.y \\ 1 & s.x & s.y \end{bmatrix} = (r.x - q.x)(s.y - q.y) - (r.y - q.y)(s.x - q.x).$$

The *signed area* of the triangle Δqrs is $\frac{1}{2}\Delta(q, r, s)$.

Let o be an *arbitrary* point in the plane. Meister observed that the *signed area* of any *oriented* polygon P is the sum of the signed areas of the triangles determined by o and the edges of P :

$$\text{area}(P) = \frac{1}{2} \sum_{i=0}^{n-1} \Delta(o, p_i, p_{i+1})$$

(To simplify notation, I'll omit "mod n " from all index arithmetic.) In particular, if we take o to be the origin $(0, 0)$, we have

$$\text{area}(P) = \frac{1}{2} \sum_{i=0}^{n-1} (x_i \cdot y_{i+1} - y_i \cdot x_{i+1}) = \frac{1}{2} \sum_{i=0}^{n-1} \det \begin{bmatrix} x_i & y_i \\ x_{i+1} & y_{i+1} \end{bmatrix}$$

where $p_i = (x_i, y_i)$ for each index i . This expression of Meister's algorithm is commonly known as the *shoelace formula*, because the pattern of multiplications resembles the usual method for threading shoelaces:

$$\begin{bmatrix} x_0 & \times & x_1 & \times & x_2 & \times & x_3 & \times & \dots \\ y_0 & & y_1 & & y_2 & & y_3 & & \dots \end{bmatrix}$$

Proving that the shoelace algorithm correctly computes areas is straightforward. First, we can prove that the formula is correct for *triangles*, either by verifying² the algebraic identity

$$\Delta(q, r, s) = \Delta(o, q, r) + \Delta(o, r, s) + \Delta(o, s, q)$$

or by geometric case analysis, as suggested by the figure below. Areas outside the triangle are either counted once positively and once negatively, or not counted at all; areas inside the triangle are counted exactly once, with the correct sign.

Then to prove the shoelace formula correct for any larger polygon P , we can sum the signed areas of all triangles in any frugal triangulation of P and observe that the terms involving diagonals of the triangulation cancel out. As expected, the resulting signed area is positive if P is oriented counterclockwise (that is, with the interior on the left) and negative if P is oriented clockwise.

Meister actually used his shoelace formula to define the signed areas of an *arbitrary* polygon, even if that polygon is not simple. Here it's somewhat less clear that the formula is *correct* for

²Hint: cofactor expansion

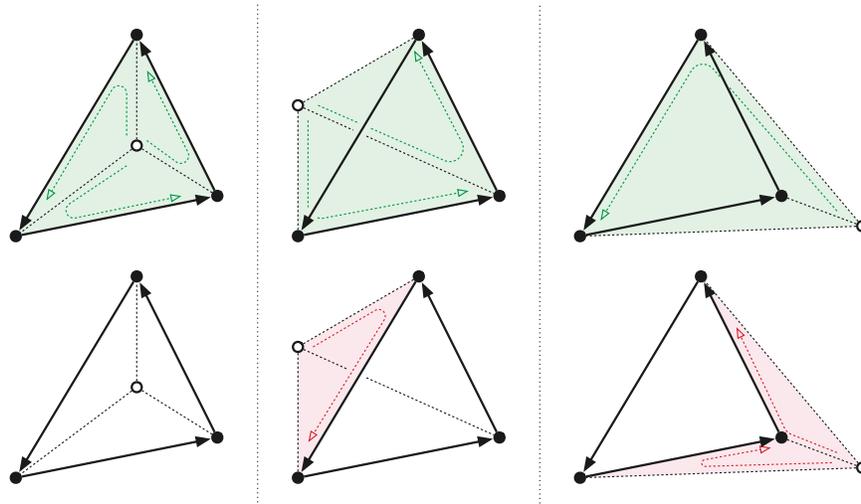


Figure 3: Lacing a triangle: Add the green (counterclockwise) triangles and subtract the pink (clockwise) triangles.

non-convex polygons, but we can verify two facts that suggest that it is at least *sensible*. First, the resulting signed area is still independent of the choice of point o . Second, the formula counts the area of each component of $\mathbb{R}^2 \setminus P$ some integer number of times; in particular, the (infinite) area of the unbounded region is not counted at all. The resulting assignment of integers to the components of $\mathbb{R}^2 \setminus P$ is called the *Alexander numbering* of P .

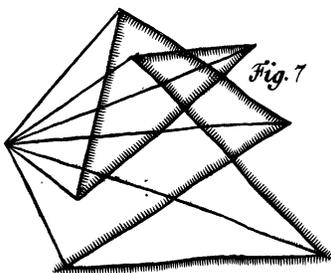


Figure 4: Computing the signed area of a polygon, from Meister (1785)

2.3 Winding numbers

The *winding number* of a polygon P around a point o is intuitively (and not surprisingly) the number of times that P winds counterclockwise around o . For example, if P is a *simple* polygon, its winding number around any exterior point is zero, and its winding number around any interior point is either $+1$ or -1 , depending on how the polygon is oriented. If the polygon winds *clockwise* around o , the winding number is negative. Crucially, the winding number is only well defined if the polygon does not *contain* the point o .

We can define the winding number more formally as follows. Let p_0, p_1, \dots, p_{n-1} denote the vertices of P in order. For each index i , let θ_i denote the interior angle at o in the triangle $\Delta p_i o p_{i+1}$, with positive sign if (o, p_i, p_{i+1}) is oriented counterclockwise, and with negative sign if (o, p_i, p_{i+1}) is oriented clockwise. Assuming angles are measured in *circles*,³ the winding number

³This is, of course, the only *correct* way to measure angles, as opposed to radians or degrees or some other heresy.

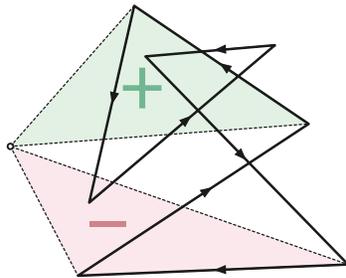


Figure 5: One positive triangle and one negative triangle for Meister's polygon

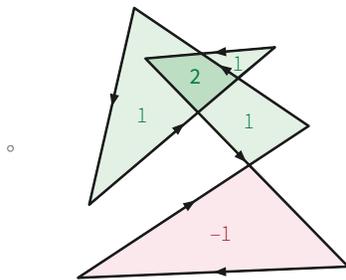


Figure 6: The Alexander numbering of Meister's polygon

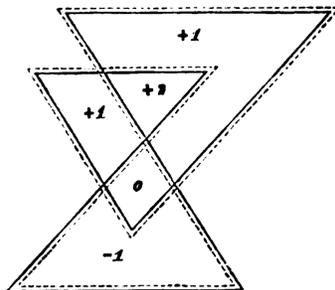


Fig. 5.

Figure 7: Another Alexander numbering, from Möbius (1865)

of P around o is the sum $\sum_i \theta_i$.

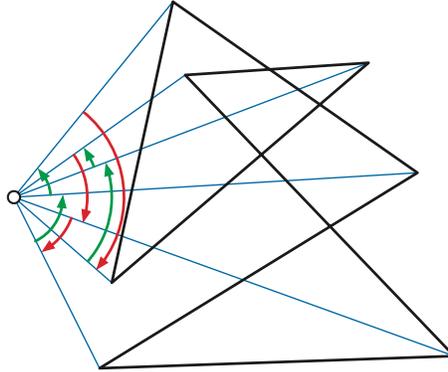


Figure 8: Winding number as a sum of angles, after Meister

Actually computing the winding number according to this definition requires inverse trigonometric functions, square roots, and other numerical madness. Fortunately, there is an equivalent definition that builds on our ray-shooting test from the previous lecture. Let R be a vertical ray shooting upward from o . We distinguish two types of crossings between the R and the polygon, depending on the orientation of the crossed edges. Specifically, if the crossed edge is directed from right to left, we have a *positive* crossing; otherwise, we have a *negative* crossing. Equivalently, when R crosses an edge $p_i p_{i+1}$, the sign of the crossing is the sign of the determinant $\Delta(o, p_i, p_{i+1})$.

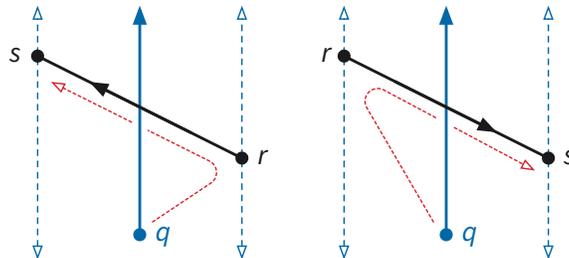


Figure 9: A positive crossing (left) and a negative crossing (right)

I'll leave the equivalence of these two definitions as an exercise. (Hint: prove equivalence for triangles, and then look at Meister's figure again!)

Here is the ray-shooting algorithm in (pseudo)Python. Any similarities with the point-in-polygon algorithm from the previous lecture are purely intentional.

```
def windingNumber(P, o):
    wind = 0
    n = size(P)
    for i in range(n):
        p = P[i]
        q = P[(i+1)%n]
        Delta = (p.x - o.x)*(q.y - o.y) - (p.y - o.y)*(q.x - o.x)
        if p.x <= o.x < q.x && Delta > 0:
            wind += 1
        elif q.x <= o.x < p.x && Delta < 0:
```

```

wind -= 1
return wind

```

Winding numbers has a third useful interpretation, which we've already seen in this lecture. Case analysis similar to our proof of Lemma ≤ 2 from the previous lecture implies that if o and o' are two points in the same component of $\mathbb{R}^2 \setminus P$, then P has the same winding number around both points. Moreover, if o and o' are in components of $\mathbb{R}^2 \setminus P$ that share a segment of some polygon edge e on their boundary, then the winding numbers around o and o' differ by 1, with the higher winding number to the left of e .

This assignment of winding numbers to the components of $\mathbb{R}^2 \setminus P$ is identical to the Alexander numbering of P that we defined earlier. That is, the winding number of P around any point $q \notin P$ is precisely the number of times that the area around q is counted by the shoelace formula. Thus, the signed area of any polygon P can expressed in terms of winding numbers as

$$\text{area}(P) = \sum_c \text{wind}(P, c) \cdot \text{area}(c)$$

where the sum is over the components of $\mathbb{R}^2 \setminus P$.

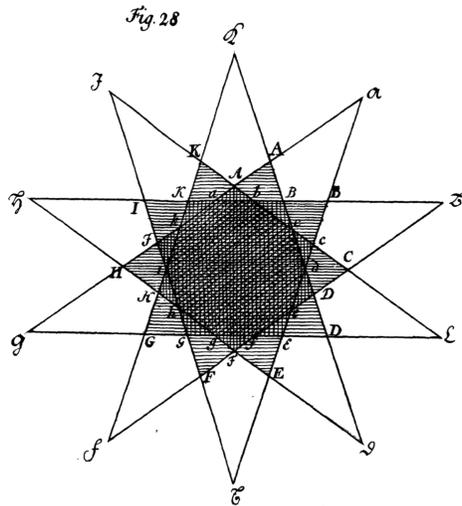


Figure 10: Another non-simple polygon from Meister, with winding numbers indicated by shading (1785)

2.4 Homotopy

Now let's go back to the Endless Chain. A bit of case analysis should convince you—or should at least *strongly suggest*—that in all three configurations, the chain is loose around your finger if and only if the winding number of the Chain around your finger is zero.

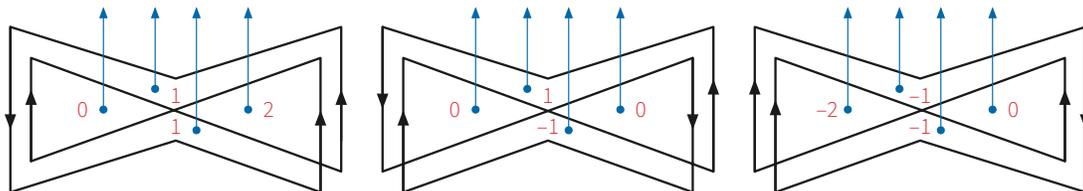


Figure 11: Winding numbers of the Endless Chain around various points

But in the actual game, we aren't dealing with a single fixed closed curve. The con man grabs one side of the chain and pulls, causing the chain to move continuously across the barrelhead and around your finger, until it either gets stuck Fast or pulls Loose. Instead of thinking about how a fixed polygon wraps with a changing point, we need a way to reason about how a *continuously changing* curve wraps around a *fixed* point.

A *homotopy* between two closed curves is a continuous deformation—a morph—from one curve to the other. Homotopies can be defined between curves in any topological space, but for purposes of illustration, let's restrict ourselves to curves in the punctured plane $\mathbb{R}^2 \setminus o$, where o is an arbitrary point called the *obstacle*. (In Fast and Loose, the obstacle is your finger.) Without loss of generality, I will assume that o is actually the origin $(0, 0)$.

Formally, a *homotopy* between two closed curves in $\mathbb{R}^2 \setminus o$ is a continuous function $h: [0, 1] \times S^1 \rightarrow \mathbb{R}^2 \setminus o$, such that $h(0, \cdot)$ and $h(1, \cdot)$ are the initial and final closed curves, respectively. For each $0 < t < 1$, the function $h(t, \cdot)$ is the intermediate closed curve at “time” t . Crucially, none of these intermediate curves touches the obstacle point o .

Two closed curves in $\mathbb{R}^2 \setminus o$ are *homotopic*, or in the same *homotopy class*, if there is a homotopy from one to the other in $\mathbb{R}^2 \setminus o$. Homotopy is an equivalence relation.

A closed curve is *contractible* in $\mathbb{R}^2 \setminus o$ if it is homotopic to a single point (or more formally, to a constant curve).

2.5 Vertex moves

Similar to the definition of “connected”, the definition of “homotopy” allows intermediate curves to be arbitrarily wild closed curves even if the initial and final curves are polygons.

Fortunately, there is a general principle that allows us to “tame” homotopies between tame curves like polygons, by decomposing them into a sequence of elementary *moves*. (This principle is similar to the observation that any closed curve can be approximated by a sequence of line segments, otherwise known as a polygon.)

Let P be any polygon. A *vertex move* translates exactly one point p of P along a straight line from its current location to a new location p' , yielding a new polygon P' . As the point p moves, the edges incident to p pivot around their other endpoints. Typically the moving point p is a vertex of the initial polygon P and the final point p' is a vertex of the final polygon P' , but neither of these restrictions is required by the definition. We are allowed to freely introduce new vertices in the middle of edges, or freely delete “flat” vertices between two collinear edges.

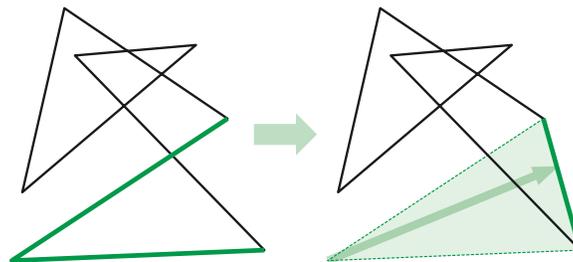


Figure 12: A vertex move.

Now suppose the polygon P lives in the punctured place $\mathbb{R}^2 \setminus o$. Let p, q, r be three consecutive vertices of P . The vertex move $q \mapsto q'$ is *safe* if neither of the triangles $\Delta pqq'$ or $\Delta qq'r$ contains

the obstacle point o . Equivalently, during a safe vertex move, the continuously changing polygon never touches o .

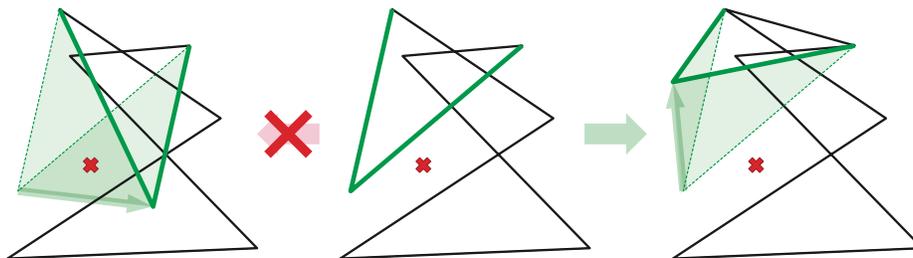


Figure 13: An unsafe vertex move and a safe vertex move.

It follows that every safe vertex move is a homotopy in $\mathbb{R}^2 \setminus o$. We can build up more complex homotopies by concatenating several safe vertex moves. In fact, *any* sequence of safe vertex moves describes a homotopy in $\mathbb{R}^2 \setminus o$.

2.6 Polygon homotopies are sequences of vertex moves

Unfortunately, the converse of this observation is false; not every homotopy is a sequence of vertex moves. Consider, for example, a simple translation or rotation of the entire polygon! Nevertheless, every homotopy can be *approximated* by a sequence of safe vertex moves.

Lemma: If two polygons in $\mathbb{R}^2 \setminus o$ are homotopic, then they are homotopic by a sequence of safe vertex moves.

Proof: Fix a homotopy $h: [0, 1] \times S^1 \rightarrow \mathbb{R}^2 \setminus o$ between two polygons $P_0 = h(0, \cdot)$ and $P_1 = h(1, \cdot)$.

For any parameters t and θ , let $d(t, \theta)$ be the Euclidean distance from $h(t, \theta)$ to the origin o , and let $\varepsilon = \min_{t, \theta} d(t, \theta)$. Because the cylinder $[0, 1] \times S^1$ is compact, this minimum is well-defined and positive.

We subdivide the cylinder $[0, 1] \times S^1$ into triangles as follows. First, cut the cylinder into a grid of $\delta \times \delta$ squares $\square(i, j) = [i\delta, (i+1)\delta] \times [j\delta, (j+1)\delta \bmod 1]$, where $\delta > 0$ is chosen so that the diameter of $h(\square(i, j))$ is at most $\varepsilon/2$. (The existence of δ is guaranteed by continuity.) Then further subdivide each grid square into two right isosceles triangles, as shown in the figure below. Without loss of generality, assume each vertex of P_0 and P_1 the image of some vertex on the boundary of the resulting triangle mesh Δ .

The homotopy h maps any cycle in this mesh to a closed curve, which consists of $O(1/\delta)$ curve segments, each with diameter at most $\varepsilon/2$, and each with distance at least ε from o . Define a new homotopy $h': [0, 1] \times S^1 \rightarrow \mathbb{R}^2 \setminus o$ that agrees with h at every grid vertex and linearly interpolates within each grid triangle. Changing from h to h' changes the image of any grid cycle by replacing each short curve segment with a straight line segment.

We can easily construct a sequence of $1 + 2/\delta^2$ cycles in Δ that starts with one boundary $0 \times S^1$ and ends with the other boundary $1 \times S^1$, such that the symmetric difference between two adjacent cycles is the boundary of one triangle in Δ . Two adjacent cycles in this sequence are shown on the right in the following figure.

The piecewise-linear homotopy h' maps any two adjacent cycles in this sequence to a pair of polygons P_t and P_{t+1} that differ by a single vertex move. Every vertex of each

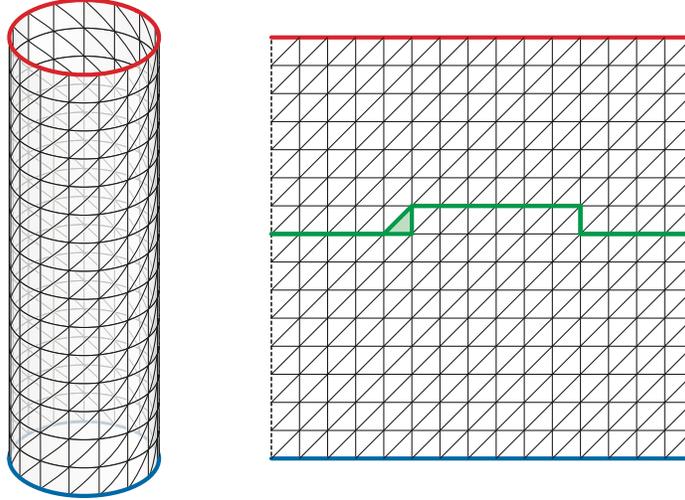


Figure 14: A grid on the unit cylinder.

intermediate polygon has distance at least ε from the origin, each edge has length at most $\varepsilon/2$, and each vertex move translates its vertex a distance of at most $\varepsilon/2$. It follows that every vertex move in this sequence is safe.

We conclude that P_0 can be transformed into P_1 by a sequence of $1 + 2/\delta^2$ safe vertex moves.

This lemma is a special case of a more general *simplicial approximation theorem*, which intuitively states that any continuous map between *nice* topological spaces (formally, geometric simplicial complexes) can be approximated by a *nice* continuous map (formally, simplicial maps between finite subdivisions of the original complexes); moreover, the original map can be continuously deformed to its approximation.

2.7 Homotopy Invariant

Winding numbers are our first example of a *topological invariant*, and specifically a *complete homotopy* invariant. A topological invariant is any property of objects or spaces that is unchanged by some form of topological equivalence. One simple example is the number of components of a subset of the plane; another standard example is the *genus* of a surface. A *homotopy* invariant is any property that is preserved by homotopy; a homotopy invariant is *complete* if it takes on different values for two objects that are not homotopic.

Theorem: Two polygons are homotopic in $\mathbb{R}^2 \setminus o$ if and only if they have the same winding number around the origin o . Thus, winding number is a complete homotopy invariant for polygons in $\mathbb{R}^2 \setminus o$.

Proof: Fix two polygons P_0 and P_1 in $\mathbb{R}^2 \setminus o$. If these two polygons are homotopic, then by the previous lemma, they are connected by a sequence of safe triangle moves. A safe triangle move does not change the winding number of a polygon around the origin. Thus, by induction, P_0 and P_1 have the same winding number.

To prove the converse, I'll describe a sequence of safe triangle moves that transforms any polygon P into a *canonical* polygon \diamond^w with the same winding number w around the origin.

(The notation \diamond^w will make sense later, honest.) Thus, if P_0 and P_1 have the same winding number w , we can deform P_0 into P_1 by concatenating the move sequence that takes P_0 to \diamond^w and the reverse of the move sequence that takes P_1 to \diamond^w .

Our homotopy consists of several stages. First let's consider the case where the winding number of P around 0 is not zero.

- Let p_i be any vertex of P , and let p_{i-1} and p_{i+1} be the next and previous vertices. We call q *redundant* if the triangle $\triangle p_{i-1}p_i p_{i+1}$ does not contain the origin. In particular, if the triples (o, p_{i-1}, p_i) and (o, p_i, p_{i+1}) have opposite orientations, one clockwise and the other counterclockwise, then p_i is redundant. In the first phase of our homotopy, we repeatedly remove redundant vertices, by moving each redundant vertex q to one of its neighbors, until none are left. The resulting polygon P' is *angularly monotone*: every triple (o, p_i, p_{i+1}) has the same orientation.

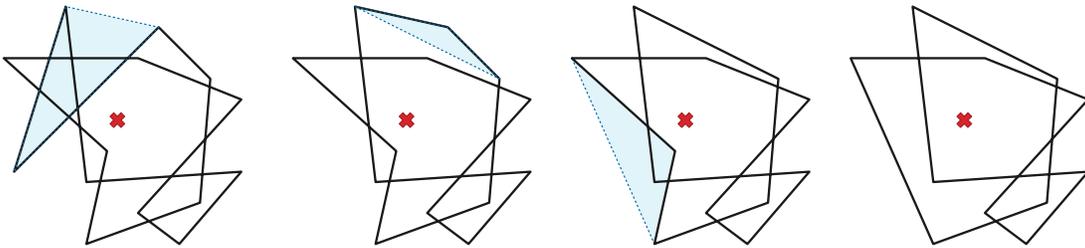


Figure 15: Removing three redundant vertices

- Next, we subdivide P' by adding vertices at its intersections with rays pointing up, down, left, and right from the origin o . After this subdivision, any vertex that is *not* on one of these rays is redundant. So in the second phase of the homotopy, we remove all non-ray vertices using safe vertex moves. The resulting polygon P'' is still angularly monotone.
- Finally, we move each vertex so that its distance from the origin is 1; each of these vertex moves is safe. The resulting polygon \diamond^w has vertices only at the points $(0, 1)$, $(1, 0)$, $(0, -1)$, and $(-1, 0)$; the polygon winds around this diamond $|w|$ times, counterclockwise if $w > 0$ and clockwise if $w < 0$.

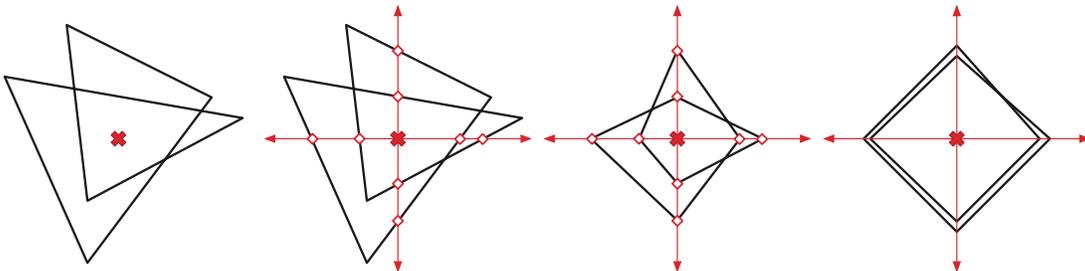


Figure 16: Making an angularly monotone polygon canonical

The special case where P has winding number 0 is even simpler. The first phase (removing redundant vertices) actually reduces P to a single point; we can then translate this point to $\diamond^0 = (1, 0)$ using one more safe vertex move.

This theorem immediately implies a linear-time algorithm to decide if two polygons are homotopic

in the punctured plane: Count positive and negative crossings between each polygon and an arbitrary ray from the origin.

2.8 ... and the Aptly Named Sir Not Appearing in This Film

- rotation number = total turning angle = smiles – frowns
- regular homotopy = vertex moves without spurs
- rotation number is a regular homotopy invariant
- complex root finding / fundamental theorem of algebra
- signed volumes of self-intersecting polyhedra (hic utres unilaterales nascuntur)

2.9 References

1. James W. Alexander. Topological invariants of knots and links. *Trans. Amer. Math. Soc.* 30(2):275–306, 1928. The reason we call it “Alexander numbering”.
2. Brian Brushwood. Fast and Loose. YouTube, October 19, 2015.
3. Brian Brushwood. Pricking the Garter. YouTube, June 26, 2017.
4. Albrecht Ludwig Friedrich Meister. Generalia de genesi figurarum planarum, et independentibus earum affectionibus. *Novi Commentarii Soc. Reg. Scient. Gott.* 1:144–180 + 9 plates, 1769/1770. Presented January 6, 1770. The shoelace formula.
5. August F. Möbius. Über der Bestimmung des Inhaltes eines Polyäders. *Ber. Sächs. Akad. Wiss. Leipzig, Math.-Phys. Kl.* 17:31–68, 1865. *Gesammelte Werke* 2:473–512, Leipzig, 1886. An earlier appearance of Alexander numbering.