WHAT IS... THE ZIGZAG THEOREM?

ANDREW NICOL

There are actually many different zigzag theorems. We will discuss two different zigzag theorems. We will begin with Isbell's zigzag theorem, which is a an algebraic result involving semigroups and dominions. Then we will also discuss zigzags between two circles, an ergodic result, which is a closure theorem, similar to Poncelet's porism and Steiner's porism. We will then mention a general theorem, which has all three closure theorems as special cases.

1. Isbell's ZigZag theorem

We begin by presenting Isbell's zigzag theorem.

First, recall that a *monoid* is a set G together with a binary operation $G \times G \to G$ that satisfies the law of associativity. Additionally, G has a unit element. We say that H is a *submonoid* of a monoid G if H is itself a monoid, H is a subset of G, and H has the same binary operation as G.

Now we define the *dominion* of U in S to be the set of elements $s \in S$ such that for all monoids T and for all homomorphisms $h_1, h_2 : S \to T$ with the property that $h_1\big|_U = h_2\big|_U$, we have $h_1(s) = h_2(s)$. We shall denote the dominion of U in S by $\mathrm{Dom}_S(U)$.

We now turn to [5] for some motivation to continue. Usually, an *epimorphism* is defined to be an onto homomorphism. In category theory, an *epimorphism* $\alpha : A \to B$ is defined by a cancellation property. That is, for all objections C in the category and for all morphisms $\beta, \gamma : B \to C$, if $\beta \circ \alpha = \gamma \circ \alpha$, then it follows that $\beta = \gamma$. For monoids, every onto morphism is an epimorphism. The converse holds in groups, but not necessarily in rings or monoids. Dominions are useful here because there is a result which states that the inclusion map $U \hookrightarrow S$ is an epimorphism if and only if $\mathrm{Dom}_S(U) = S$. Additionally, a map $\alpha : S \to T$ is an epimorphism if and only if $\mathrm{Dom}_T(\mathrm{im}\ \alpha) = T$.

Now that we have the motivation to understand dominions, we introduce the zigzag theorem, using the notation and a sketch of the proof used in [4].

Theorem 1. *Let U be a submonoid of S, a monoid with $s_* \in S$. $s_* \in Dom_S(U)$ if and only if either $s_* \in U$ or there exists a U-zigzag with value s_*. That is, we have*

$$s_* = s_1 u_1^0 = s_1 u_1^1 s^1 = s_2 u_2^1 s^1 = s_2 u_2^2 s^2 = \ldots = s_n u_n^n s^n = u_{n+1}^n s^n = s_*,$$

with $u_k^l \in U$, $s_k, s^l \in S$, $s_i u_i^i = s_{i+1} u_{i+1}^i$, $u_{i+1}^i s^i = u_{i+1}^{i+1} s^{i+1}$ for $i = 1, \ldots, n-1$, and $u_1^0 = u_1^1 s^1$, $s_n u_n^n = u_{n+1}^n$.

Another way of seeing this factorization of s_* can be seen in figure 1, taken from [4]. The U-zigzag with value s_* is equivalent to the commutativity of figure 1.

Proof. (sketch) The reverse direction is straightforward, and will left as an exercise.

For the forward direction, assume s_* is in the dominion of U in S. Let A be the set of all elements of S combined with a new element t, such that $t^2 = 1$ and

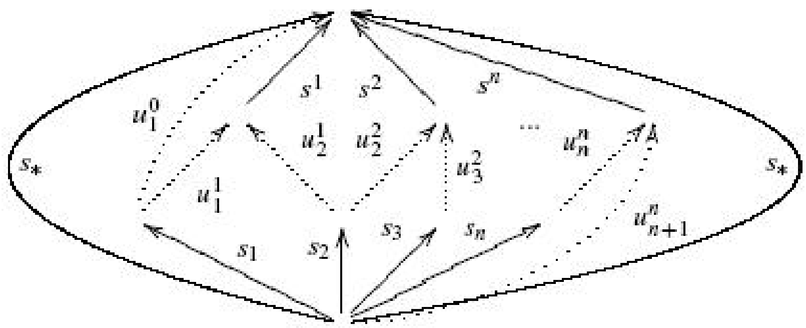

FIGURE 1. A U-zigzag with value s_*.

$tut = u$ for all $u \in U$. Let A^* be the set of all finite words over the alphabet A, with ϵ being the empty word.

Now, there are three types of reductions that can be done on the set of words from A^*. *Refactorization* occurs when $s_1 \dots s_n \leftrightarrow s'_1 \dots s'_k$ for $s_1 \dots s_n = s'_1 \dots s'_k$ in S and $s_i, s'_j \in S$. A *right/left shift* is when $tu \leftrightarrow ut$ for $u \in U$. Finally, a *creation/deletion* is the reduction $\epsilon \leftrightarrow tt$. Then $\leftrightarrow$ is almost an equivalence relation, though it lacks transitivity (the result of a creation followed by a shift cannot be accomplished by a single reduction). Thus, let $\leftrightarrow^+$ be the transitive closure, making it an equivalence relation, giving us a quotient module T.

Now, define $\mu, \nu : S \to T$ by $\mu(s) = s$ and $\nu(s) = tst$. By definition, for all $u \in U$, we have $\mu(u) = u = tut = \nu(u)$, so μ and ν agree on all of U. Thus, by definition of s_* being in the dominion of U in S, we have $\mu(s_*) = \nu(s_*)$. Therefore, $s_* \leftrightarrow ts_*t$, which is equivalent to $s_*t \leftrightarrow ts_*$.

This will give rise to a sequence of reductions $s_*t = w_1 \leftrightarrow \dots \leftrightarrow w_n = ts_*$. Using this, we can get a U-zigzag with value s_*. To do this, we need two observations. The first observation is that we can track specific occurrences of t through any reductions. The second is that if $w_i \leftrightarrow w_{i+1}$, for some t does not get deleted at this step, then $v_i \leftrightarrow v_{i+1}$, where v_i and v_{i+1} come from w_i and w_{i+1} by removing all other occurrences of t (other than the one not deleted).

Using this, we can create a sequence of new reductions that will only consist of a single occurrence of t without the creation/deletion reduction. □

2. zigzags Between Two Circles

Before we present the zigzag theorem between two curves, we begin with some similar results, beginning with Steiner's Porism.

Given two circles, one inside the other, draw a circle, γ_1 between the two, tangent to each. Then construct more (distinct!) circle, each one between the two original circles, and each one tangent to the one constructed right before it. If some γ_n is also tangent to γ_1, then the chain of circles is said to close and have period n.

Additionally, for any circle γ between the two original circles, repeating the same procedure, the nth circle will still be tangent to the first.

A picture of this process, when it closes, can be seen in figure 2.

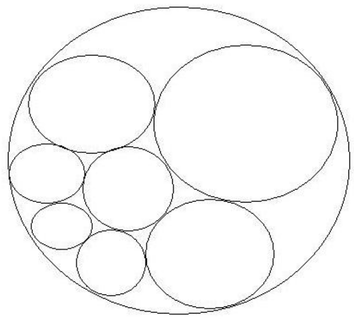

FIGURE 2. This is an example of the Steiner process that closes. There is a large circle, with a small circle inside of it. The other circles have been constructed between and tangent to the inner and outer circle.

A formalization of this process and the theorem which accompanies is can be found in [6], and is as follows.

"*Steiner Process.* Given circles α_0 and α_1 with α_1 inside of α_0, let $\mathscr{M}$ be the set of circles that are tangent to α_0 and α_1 and lie between them. For an arbitrary $\gamma_1 \in \mathscr{M}$ we take a circle $\gamma_2 \in \mathscr{M}$ tangent to γ_1, and then for any $k \geq 3$ we take a circle $\gamma_k \in \mathscr{M}$ tangent to γ_{k-1}, different from γ_{k-2}. This process has period $n \geq 2$ if $\gamma_{n+1} = \gamma_1$.

Theorem 2. *If the Steiner process is periodic for some initial circle γ_1, then is has the same period for any $\gamma \in \mathscr{M}$."*

We now take a look at Poncelet's Porism. To set this up, construct two ellipses, one inside of the other. Then, select a starting point, D_1, on the outer ellipse. From D_1, draw a line tangent to the inner ellipse and let D_2 be the point of intersection on the outer ellipse. Continue constructing new tangent lines and points of intersection on the outer ellipse. If we have $D_{n+1} = D_1$ for some minimal $n \in \mathbb{N}$, then the process closes and has period n. Poncelet's theorem states that for a fixed inner ellipse and a fixed outer ellipse, if the process closes with period n for some starting point D_1, then the process will close with period n for any starting point D. Figure 3 illustrates an example of this.

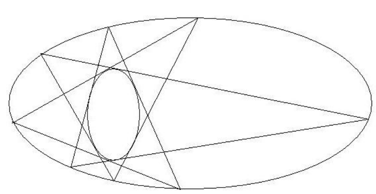

FIGURE 3. An example of the Poncelet process closing for these ellipses.

Again, we seek a formalization of this process. After applying a projective transformation, we can reduce to the case where the ellipses are circles. So now we turn to [6], which states the following.

"*Poncelet Process.* Given two circles α and δ, we draw through some point $D_1 \in \delta$ a line tangent to α that intersects δ a second times at a point D_2. We draw the second tangent to α through D_2; it meets δ a second time at a point D_3, and so on. This process is periodic if $D_{n+1} = D_1$.

Theorem 3. *If the Poncelet process has period $n \geq 3$ and the points $D_1, \ldots, D_n$ are different, then the process has the same period for any initial point $D_1 \in \delta$ through which one can draw a tangent to α."*

We are now ready to move on to the theorem about zigzags between two circles. Using [1] as a guide, Just as before, begin with two circles, Γ and $\tilde{\Gamma}$, subject to the requirement that there exists some number x such that for every point on either circle, there are exactly two points of the other circle distance x away. Then, starting with some initial point z_1 from one of the circles, say Γ. Then pick $\tilde{z_1} \in \tilde{\Gamma}$ such that $|\tilde{z_1} - z_1| = x$. In a similar manner, pick $z_2 \in \Gamma$ and so on, with that addition restriction that $z_k \neq z_{k-1}$ and $\tilde{z_k} \neq \tilde{z}_{k-1}$. If there exists some n such that $z_{n+1} = z_n$, then the process closes. The zigzag theorem states that the zigzag could have started at any point and it would still close. Figure 4 shows an example of this occurring.

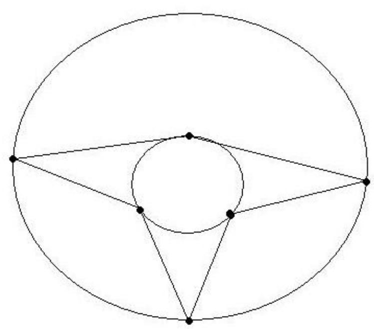

Figure 4. Two circles with a zigzag between them closing after only a couple steps.

Taking a step back, we can see some obvious similarities and some difference between this zigzag theorem and Steiner's porism and Poncelet's porism. First of all, they all deal with some sort of process that gets iterated many times. In all cases, if that iteration is periodic, the starting position does not affect the periodicity. Also, one method of proving this zigzag theorem uses a technique which works for Poncelet's Porism as well. The idea, from [1] is the find how much z_{n+1} moves when z_1 is moved. Recall that z_1 is our initial point and $z_{n+1}(= z_1)$ was our $n+1$st point, both in Γ. It can be shown that z_{n+1} moves at the same rate as z_1, ie, $dz_{n+1} = dz_1$, so wherever z_1 moves to, z_{n+1} will also move to that same location, guaranteeing that the process will close after the same number of steps. The full proof, using differential equations, can be found in [1].

The main difference that separates the zigzag theorem from Poncelet's Porism and Steiner's Porism is the fact that the zigzag theorem holds for any two circles in

$\mathbb{R}^3$, still subject to its restriction. If the two circles are coplanar, this requirement that there exist some x such that every point on either circle is distance x away from exactly two points of the other circle can be simplified. The equivalent requirement for the coplanar case is that the smaller circle must enclose the center of the larger circle [1].

We now turn to [6] for a formalization. In this paper, however, there is a difference in the statement. [6] formalizes the zigzag theorem as follows.

" *Zigzag process.* Given two circles σ and δ, neither of which contains the center of the other, and given a number $\rho > 0$, choose some $D_1 \in \delta$ and $S_1 \in \sigma$ such that $D_1S_1 = \rho$. Take a point $D_2 \in \delta$ such that $D_2S_1 = \rho$ and $D_2 \neq D_1$ (if it does not exist, then $D_2 = D_1$); then take a point $S_2 \in \sigma$ such that $S_2D_2 = \rho$ and $S_2 \neq S_1$ (otherwise $S_2 = S_1$), and so on. The zigzag has period n if $D_{n+1} = D_1$.

Theorem 4. *If the zigzag has period $n \geq 3$ and all intermediate points are different, then it has the same period for any initial points $D_1 \in \delta$ from which one can make the first step."*

We can easily see the main difference between the previous presentation of the zigzag theorem and this presentation. An example of this version of the zigzag theorem can be seen in figure 5.

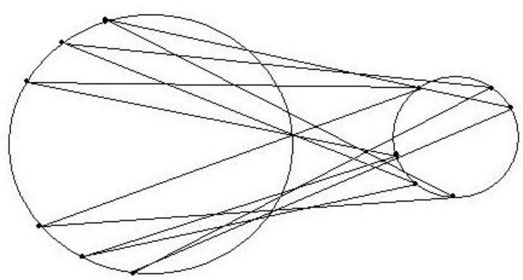

FIGURE 5. Two circles, none of which contain the center of the other. A zigzag between the circles is shown.

A general theorem can be found in [6]. This theorem has Steiner's Porism, Poncelet's Porism, and the zigzag theorem as special cases. This generalized theorem uses the notation and format as presented in the processes and theorem included above.

Csikós and Hraskó, in their paper, [2], they extend some results relating to the zigzag theorem. Among these extensions, they discuss the zigzag theorem when the two circles are not contained in 3 dimensions, they show that the theorem holds in hyperbolic space, and they show the theorem holds when the curves are " arbitrary uniform curves of the space" as opposed to necessarily being circles. These are some results which, as mentioned earlier, can be found in their paper, [2]. Here, we shall discuss other results of their's, some results in the case when dealing with two coplanar circles.

Csikós and Hraskó discuss how a zigzag can be described by four parameters: x_R, x_r, the radii of the large and small circles, x_d, the distance between the centers of the two circles, and x_ρ, the length of the zig/zag. We represent this as a four-tuple, (x_R, x_r, x_d, x_ρ). For these results, assume that (x_R, x_r, x_d, x_ρ) forms a configuration such that the zigzags close after n steps. Then (x_R, x_r, x_ρ, x_d) will also close after n steps. That is, if we interchange the distance between the circles and the length of the zig/zag, we will again have a situation where the zigzag closes after n steps.

Another result presented in [2] is that interchanging the small radius with the length of the zig/zag, and switching the values of the larger radius and the distance between the centers of the circles, the new arrangement will still close after some number of steps. That is, (x_d, x_ρ, x_R, x_r) will still close. If n is even, the new configuration, it will close in n steps, while it will close in m steps, $m|n$, if n is odd.

They also present the result that in the situation (x_a, x_a, x_d, x_a), that is, if the two circles have the same radius, a, and the length of the zig/zag is also a, then any zigzag will close in 3 steps.

These results and more can be found, with more detail, in [2]. This zigzag theorem, and other similar results presented above, are just a few of these closure related results.

References

[1] H. C. H. Black W. L. and B. Howland, *A theorem about zig-zags between two circles*, The American Mathematical Monthly, 81 (1974), pp. 754–757.

[2] B. Csikós and A. Hraskó, *Remarks on the zig-zag theorem*, Periodica Methematica Hungarica, 39 (1999), pp. 201–211.

[3] P. M. Higgins, *A short proof of isbell's zigzag theorem*, Pacific Journal of Mathematics, 144 (1990), pp. 47–50.

[4] P. Hoffman, *A proof of isbell's zigzag theorem*, Journal of the Australian Mathematical Society, 84 (2008), pp. 229–232.

[5] J. M. Howie, *Isbell's zigzag theorem and its consequences*, Semigroup Theory and its Applications, MLS 231 (1996), pp. 81–91.

[6] V. Y. Protasov, *A generalization of poncelet's theorem*, RUSS MATH SURV, 61 (2006), pp. 1180–1182.