Simple Loop Parity

02C

1 Basics and Notation

Simple loop parity deductions are a nontrivial concept to define. They clearly involve considering a checkerboard coloring of the grid, but exactly what the types of parity deduction are, is not fully clear. For our purposes, we'll consider two types of parity deduction. Basically all parity deductions I know of that immediately give the status of some edge are of one of these types. There are deductions that are not immediately reducible to one of these types, such as loop theory, but those aren't parity.

Both types of parity deduction involve a parity region. This is a subset of the cells in the simple loop. They also involve the parity of cells, which is just based on a checkerboard coloring. We'll say cells are either black (B) or white (W). Further, they both involve the net parity of the parity region. For us, this will be the number of black cells in it minus the number of white cells; we could have also defined it to be the reverse. A simple loop will have some shaded cells which aren't used in the loop; we don't count those in computing net parity. We'll use $B(A)$ and $W(A)$ to refer to the sets of black and white cells in A, so the net parity of A is $|B(A)| - |W(A)|$.

For both types of parity deduction, the boundary of the parity region is also relevant. There are, at a given stage of solving, two important properties of each edge around the boundary. The first is whether the side of the edge in the region is on a black cell, or a white cell. The second is whether the edge is known to be used by the loop, known to not be used, or unknown. This gives six types of edge. We'll represent "whether the side of the edge in the region is on a black cell, or a white cell" by B or W respectively, and we'll represent "used/unused/unknown" by +, -, and ? respectively. This gives us shorthand whereby we can talk about, for example, $[B+]$ edges on the boundary. We'll also use ∂A to represent the boundary of a region A. We'll use $[B+](\partial A)$ to represent the set of $[B+]$ edges on the boundary of A, letting us use $|[B+](\partial A)|$ to represent the number of such edges.

One very useful thing is that for any region A, considering the completed simple loop solution (so that all edges are known),

$$2(|B(A)| - |W(A)|) = |[B+](\partial A)| - |[W+](\partial A)|.$$

That is, the "net parity" of the exits is twice the net parity of the region.

We can now define the two types of parity.

2 First Type of Parity

The first type of parity is: if

$$2(|B(A)| - |W(A)|) = |[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)|,$$

then all the edges in $[B?](\partial A)$ are used and all the edges in $[W?](\partial A)$ are not used. Further, if

$$2(|B(A)| - |W(A)|) > |[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)|,$$

there's a contradiction.

In this type of parity, the only way to get the needed parity is to use all the $[B?](\partial A)$ and none of the $[W?](\partial A)$ edges, to maximize $|[B+](\partial A)| - |[W+](\partial A)|$. This type of parity has the nice property that after it, you know the whole perimeter of A.

The above first type of parity of course also holds with B and W swapped. However, one gets this "for free" by replacing A by the complement A^c consisting of all the other cells. Indeed, in any simple loop, the net parity of the whole grid is 0, so $|B(A)| - |W(A)| = |W(A^c)| - |B(A^c)|$. Also, $[B+](\partial A) = [W+](\partial A^c)$ and so on, because if the black side of a boundary edge is in A, the white side is in A^c.

3 Second Type of Parity

The second type of parity is: if $|B(A)| - |W(A)| = 0$, $|[B+](\partial A)| = 0$, and $|[B?](\partial A)| = 1$, then the one edge in $[B?](\partial A)$ is used. Furthermore, if $|B(A)| - |W(A)| = 0$, $|[B+](\partial A)| = 0$, $|[B?](\partial A)| = 0$, and A is neither empty nor all the (unshaded) cells, there's a contradiction.

Note that here, A is a zero-parity region. A zero-parity region A, in the solution, has the same number of exits in $[B+](\partial A)$ as in $[W+](\partial A)$. If A is neither empty nor all the unshaded cells, A has to connect to the rest of the loop, so there must be at least one edge in $[B+](\partial A)$. So if there's only one edge between $[B+](\partial A)$ and $[B?](\partial A)$, it has to be used, and if there are none, we reach a contradiction.

The condition that A is neither empty nor all the cells is not needed in the first part of this type of parity, because in those trivial cases, A could have no edges between the outside and the inside (one or the other would be empty) and so $|[B?](\partial A)| = 0$ (all edges in ∂A are known not to be used from the start). So $|[B?](\partial A)| = 1$ implies that A is not one of those trivial cases.

As with the first type of parity, we can swap B and W but this is the same as replacing A by the complement A^c.

4 Interlude

It's worth noticing that, with the exception of simple loops from NP-completeness proofs and one specially created simple loop using the same construction as the NP-completeness proofs (but in a perhaps slightly simpler way), these two types of parity are enough to solve all simple loops I've seen, as far as I know (and I've tested a lot, including every simple loop that looks interesting). We'll cover a way to automatically check these types of parity fully, but it's not quick enough to handle some big simple loops for me and it's also not needed; special cases of parity deductions are usually enough, with a bit of manual intervention for some tricky parity regions. However, parity as defined here can't get the fact that there's no edge between Row 3 Column 3 and Row 3 Column 4 in `https://puzz.link/p?simpleloop/6/4/00036` so it partially does seem to depend on the simple loop being unique, somehow.

5 Implementation

Note: The key ideas here come from Lebossle.
Given a region A, both these types of parity are very easy to check. However, what if we want to find parity deductions without brute-forcing every A? It turns out both these types of parity can be transformed, roughly, into finding zero-sum cycles in a weighted directed graph with negative weights, which is polynomial-time. (Basically using Bellman-Ford.)

The first step is to define our graph. We'll have the vertices of our graph be the points between cells, not the cells themselves. The edges will connect adjacent points between cells, corresponding one-to-one with edges between adjacent cells. Also, the edges should be directed. There is a mess involving clockwise vs counterclockwise directions of paths, but it's fairly easy to see that we can compute things like $|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)|$ edge-by-edge if we know the used/unused/unknown status of each edge. The hard part then becomes computing the net parity of A given just the edges of ∂A. However, this is very doable; we just assign vertical edges the weight 0 and assign horizontal edges the net parity of the cells above them (or the negation of that, if they go left). This means that the net parity of the area between two vertical edges is the sum of their weights.

This makes it possible to assign weights to edges to find cycles for which

$$2(|B(A)| - |W(A)|) = |[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)|$$

(zero-weight cycles) or

$$2(|B(A)| - |W(A)|) > |[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)|$$

(negative-weight cycles, if we set things up properly). However, to avoid trivial cycles that just take a path and then go back on it, the only method I currently know (though it's inefficient) is to pick an edge and then find paths from its end to its start that don't include it. (Of course, this needs to be done for every edge, or, depending on what you're doing, at least any unknown edge.) It's messy, but it avoids the trivial-cycle issue.

This is the first type of parity. The second is slightly more complicated. The $|[B+](\partial A)|$ and $|[B?](\partial A)|$ conditions can be ensured by just removing all the edges of type $[B+]$ or $[B?]$ from the graph. This at first seems nonsensical, since whether an edge is type $[B+]$ or $[W+]$ depends on the region A. However, the fact that edges are directed in fact lets us say one direction of an edge is $[B+]$ and the other is $[W+]$ (we can define this via, for example, the cell on the left side of the edge). For the $|[B?](\partial A)| = 1$ case, we need to include one $[B?]$ edge. This, fortunately, fits very well with our approach of finding paths from the end to the start of an edge that don't include that edge. We just consider each $[B?]$ edge and then look for paths like that using no $[B+]$ or $[B?]$ edges. For the $|[B?](\partial A)| = 0$ case, we instead consider each edge that's not a $[B+]$ or $[B?]$ edge, and do the same end-to-start-path thing for each of them.

6 Notes on First Type of Parity

Note: Much of this is likely already known; see `https://en.wikipedia.org/wiki/Vertex_cycle_cover`
The approach here works for bipartite graphs in general (i.e. those where the vertices can be divided into two subsets and all edges are between the subsets, which I guess is just parity). I'm not sure if results as strong as those here exist for other types of graph.

Let's define another genre Simple Loops. Simple Loops is like Simple Loop, except that there can be any number of loops in the solution (but every cell still have to be in one of the loops). The deductions from the second type of parity don't necessarily apply in Simple Loops. However, the deductions from the first type of parity do apply in Simple Loops, because they didn't use the fact that there's only one loop. Furthermore, it turns out that the deductions from the first type of parity are exactly those you can make in a Simple Loops puzzle. This applies even if the puzzle is partially solved. Even further, if there's no contradiction from the first type of parity, there's a solution, even if the puzzle starts out partially solved. Lastly, when we say 'the deductions from the first type of parity', this is single-step deductions from defining some region A; chains of deductions of this type don't get us more information.

Let's prove all this. We'll start with two parity regions, A and B. We'll assume that we apply the first type of parity to A, and then apply it to B. We'll try to show that anything we get by applying it to B (deductions, or a contradiction), we could've gotten immediately. We can extend this to a supposed chain of three or more parity deductions by just applying it to each step, to turn any chain of parity deductions (of the first type) into a large list of individual, independent, parity deductions.

The main tactic in our approach will be to use $A \cup B$ and $A \cap B$. We can almost say that, preserving direction, $\partial A + \partial B = \partial(A \cup B) + \partial(A \cap B)$. This is an equation with multiplicity, rather than a set equation, because ∂A and ∂B may have parts in common. However, there may be part of ∂A which is part of ∂B but in the other direction; call this P. To make this definition clear, P is exactly the edges with A (but not B) on one side and B (but not A) on the other. We can say

$$(\partial A) \setminus P + (\partial B) \setminus P = \partial(A \cup B) + \partial(A \cap B)P.$$

The point of having an equation like

$$\partial A + \partial B = \partial(A \cup B) + \partial(A \cap B) + P - P$$

is to handle perimeter terms in the first kinda of parity. We can immediately say that

$$2(|B(A)| - |W(A)|) + 2(|B(B)| - |W(B)|) = 2(|B(A \cup B)| - |W(A \cup B)|) + 2(|B(A \cap B)| - |W(A \cap B)|).$$

However, making the analogous statement for the perimeter terms is harder. And to make matters worse, some of the edges of B may have been altered by applying the first type of parity to A. We'll thus use, for example, $\{W+\}(\partial B)$ to talk about the edges of type $W+$ in ∂B after parity was applied to A. This nicely also solves the problem of how we talk about a certain type of edge on P, since a $B+$ edge from A's perspective is a $W+$ edge from B's perspective; we can say $[B+](P)$ or $\{W+\}(P)$ and since we're only using $\{\}$ for B, this is unambiguous.

We know that

$$|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)| - 2(|B(A)| - |W(A)|) = 0,$$

since if it were greater than 0 we couldn't have done parity on A and if it were less than 0 we would have reached a contradiction and thus would be done. We also know that

$$|\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| - 2(|B(B)| - |W(B)|) \leq 0,$$

since if it were greater than 0 we couldn't have done parity on B. We won't actually need these equations for a while; we're just stating them because their left sides are the things we care about.

We want to show that

$$\begin{aligned}
&|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)| - 2(|B(A)| - |W(A)|) + \\
&\quad |\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| - 2(|B(B)| - |W(B)|) = \\
&|[B+](\partial(A \cup B))| + |[B?](\partial(A \cup B))| - |[W+](\partial(A \cup B))| - 2(|B(A \cup B)| - |W(A \cup B)|) + \\
&\quad |[B+](\partial(A \cap B))| + |[B?](\partial(A \cap B))| - |[W+](\partial(A \cap B))| - 2(|B(A \cap B)| - |W(A \cap B)|).
\end{aligned}$$

Notably, we have

$$2(|B(A)| - |W(A)|) + 2(|B(B)| - |W(B)|) = 2(|B(A \cup B)| - |W(A \cup B)|) + 2(|B(A \cap B)| - |W(A \cap B)|).$$

This lets us simplify to wanting to show

$$\begin{aligned}
&|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)| + |\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| = \\
&\quad |[B+](\partial(A \cup B))| + |[B?](\partial(A \cup B))| - |[W+](\partial(A \cup B))| + \\
&\quad\quad |[B+](\partial(A \cap B))| + |[B?](\partial(A \cap B))| - |[W+](\partial(A \cap B))|.
\end{aligned}$$

From

$$(\partial A) \setminus P + (\partial B) \setminus P = \partial(A \cup B) + \partial(A \cap B)P,$$

we have

$$\begin{aligned}
&|[B+]((\partial A) \setminus P)| + |[B?]((\partial A) \setminus P)| - |[W+]((\partial A) \setminus P)| + \\
&|[B+]((\partial B) \setminus P)| + |[B?]((\partial B) \setminus P)| - |[W+]((\partial B) \setminus P)| = |[B+](\partial(A \cup B))| + |[B?](\partial(A \cup B))| - |[W+](\partial(A \cup B))| + \\
&\quad |[B+](\partial(A \cap B))| + |[B?](\partial(A \cap B))| - |[W+](\partial(A \cap B))|.
\end{aligned}$$

So we want to show

$$\begin{aligned}
&|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)| + |\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| = \\
&\quad |[B+]((\partial A) \setminus P)| + |[B?]((\partial A) \setminus P)| - |[W+]((\partial A) \setminus P)| + \\
&\quad\quad |[B+]((\partial B) \setminus P)| + |[B?]((\partial B) \setminus P)| - |[W+]((\partial B) \setminus P)|.
\end{aligned}$$

Removing some common terms, this is equivalent to

$$\begin{aligned}
&|[B+](P)| + |[B?](P)| - |[W+](P)| + |\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| = \\
&\quad |[B+]((\partial B) \setminus P)| + |[B?]((\partial B) \setminus P)| - |[W+]((\partial B) \setminus P)|.
\end{aligned}$$

Notably, the edges in $[B+](P)$ and $[B?](P)$ are exactly those B edges that become used after applying parity to A, and the $[W+](P)$ are the W edges that become used after applying parity to A. No edges in P remain unused after applying parity to A. Thus, taking into account that a B edge in P for A is a W edge for B, we get $[B+](P) \cup [B?](P) = \{W+\}(P)$, $[W+](\partial A) = \{B+\}(P)$, and $\{B?\}(P)$ is empty. Thus the equation we want to show becomes

$$\begin{aligned}
&|\{W+\}(P)| - |\{B+\}(P)| - |\{B?\}(P)| + |\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| = \\
&\quad |[B+]((\partial B) \setminus P)| + |[B?]((\partial B) \setminus P)| - |[W+]((\partial B) \setminus P)|.
\end{aligned}$$

or equivalently

$$\begin{aligned}
&|\{B+\}((\partial B) \setminus P)| + |\{B?\}((\partial B) \setminus P)| - |\{W+\}((\partial B) \setminus P)| = \\
&\quad |[B+]((\partial B) \setminus P)| + |[B?]((\partial B) \setminus P)| - |[W+]((\partial B) \setminus P)|.
\end{aligned}$$

This is in fact true; some edges where A and B are on the same side may change from $B?$ to $B+$, thus contributing to $[B?]((\partial B) \setminus P)$ and to $\{B+\}((\partial B) \setminus P)$, but these are both terms with the same sign. Similarly, some edges may change from $W?$ to $W-$, thus contributing to $[W?]((\partial B) \setminus P)$ and to $\{W-\}((\partial B) \setminus P)$, but neither of these terms appears so it's not an issue. So we've shown the original equation.

What does this mean? If

$$|\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| - 2(|B(B)| - |W(B)|) = 0,$$

then either one of

$$|[B+](\partial(A \cup B))| + |[B?](\partial(A \cup B))| - |[W+](\partial(A \cup B))| - 2(|B(A \cup B)| - |W(A \cup B)|)$$

or

$$|[B+](\partial(A \cap B))| + |[B?](\partial(A \cap B))| - |[W+](\partial(A \cap B))| - 2(|B(A \cap B)| - |W(A \cap B)|)$$

is negative, or both are zero. In the first case, we can get an immediate contradiction via parity. In the second case, we can use parity on both $A \cup B$ and $A \cap B$ immediately, and combined with parity on A this gives us the status of all edges of B, meaning parity on B itself is not needed. In no case do we need to chain parity deductions.

If, on the other hand,

$$|\{B+\}(\partial B)| + |\{B?\}(\partial B)| - |\{W+\}(\partial B)| - 2(|B(B)| - |W(B)|) < 0,$$

then necessarily one of

$$|[B+](\partial(A \cup B))| + |[B?](\partial(A \cup B))| - |[W+](\partial(A \cup B))| - 2(|B(A \cup B)| - |W(A \cup B)|)$$

or

$$|[B+](\partial(A \cap B))| + |[B?](\partial(A \cap B))| - |[W+](\partial(A \cap B))| - 2(|B(A \cap B)| - |W(A \cap B)|)$$

is negative. This means that we'd get an immediate contradiction without chaining parity deductions.

Ok, so what does this all mean for us? Considering parity for every possible region, as in the Implementations section, gives us all parity deductions of the first type. Furthermore, it is easy to see that setting one unknown edge may give new parity deductions, but if there are no unmade parity deductions it cannot lead immediately to a parity contradiction (because it can only change

$$|[B+](\partial A)| + |[B?](\partial A)| - |[W+](\partial A)| - 2(|B(A)| - |W(A)|)$$

by one, not letting it go from positive to negative). So, we get the following easy method for solving a Simple Loops puzzle: make all parity deductions of the first type, arbitrarily set some undetermined edge if any, and repeat. This can't lead to a contradiction, and it'll lead to a solution after a polynomial number of steps (each step determines one edge). So as long as there isn't an initial parity contradiction, there's a solution. Indeed, by choosing the edge we arbitrarily set and choosing what to set it to, we can say that for any edge not determined by initial parity deductions, we can set it to whatever we want and we will still have a solution.

7 Notes on Second Type of Parity

The second type of parity is as far as I know is just a mess. Remember, it couldn't show that an edge didn't exist when that edge would separate a 2x2 region from the rest of the grid.

However, the second type of parity does have the same property that applying it twice in a row doesn't give additional information. Indeed, deductions from the second type of parity only arise from edges going from unknown to unused; edges going from unknown to used doesn't give new deductions of the second type, or lead to new contradictions of the second type. But edges going from unknown to used is all the second type can do. So applying the second type doesn't help directly in applying the second type again.

One can show that the second type of parity has the same properties as the first for the genre Non-Simple Non-Loop (instead of Simple Loops), which we define to have only the rule "Any region with zero net parity must have at least one exit of each parity." But this genre is clearly based on the second type of parity's definition, and it's also clearly a very boring genre because you can just fill in every possible edge and get a solution.