A research proposal for the computational complexity of minimizing reducible sofic shifts

Justin Cai

October 23, 2019

Abstract

In symbolic dynamics, a certain class of objects called sofic shifts are sets of bi-infinite sequences that come from labeled graphs, such that each letter of the sequence is the label of an edge in a bi-infinite walk around the graph. If one wanted to use sofic shifts to model something, it would be desirable to have a graph that is as small as it can be while still presenting the same sequences. While the problem of how hard it is to compute this minimal graph for shifts with a certain property called irreducibility is known, the hardness of computing the same problem for shifts that do not have this property is unknown.

1 Background

A *full shift* is the set of all bi-infinite sequences over a finite alphabet $\mathcal{A}$. A *graph* G is a finite set of *verticies* $\mathcal{V} = \mathcal{V}(G)$ and a finite set of edges $\mathcal{E} = \mathcal{E}(G)$ with each edge e starting at a vertex $i(e) \in \mathcal{V}$ and terminating at a vertex $t(e) \in \mathcal{V}$. A bi-infinite walk on G is a bi-infinite sequence of edges such that the terminating vertex of each edge is the inital vertex of the next edge. The set of all bi-infinite walks on G is called the *edge shift* X_G. A *labeled graph* $\mathcal{G}$ is a graph G equipped with a *labeling* $\mathcal{L} : \mathcal{E}(G) \to \mathcal{A}$, which assigns each edge e from G a label $\mathcal{L}(e)$ from a finite alphabet $\mathcal{A}$. If x is a bi-infinite walk on G, then the *label of the walk* $\mathcal{L}_\infty(x)$ is the bi-infinite sequence of the labels of x. The set of all labels of bi-infinite walks is denoted $\mathsf{X}_\mathcal{G}$. A subset X of a full shift is a *sofic shift* if $X = \mathsf{X}_\mathcal{G}$ for some labeled graph $\mathcal{G}$. A labeled graph $\mathcal{G}$ is a *presentation* of X if $X = \mathsf{X}_\mathcal{G}$.

For example, let X be the set of bi-infinite sequences over $\{0, 1\}$ such that there is an even number of 0's between any two 1's. Then $X = \mathsf{X}_\mathcal{G}$, where $\mathcal{G}$ is any labeled graph in Figure 1. This shift is known as the *even shift*.

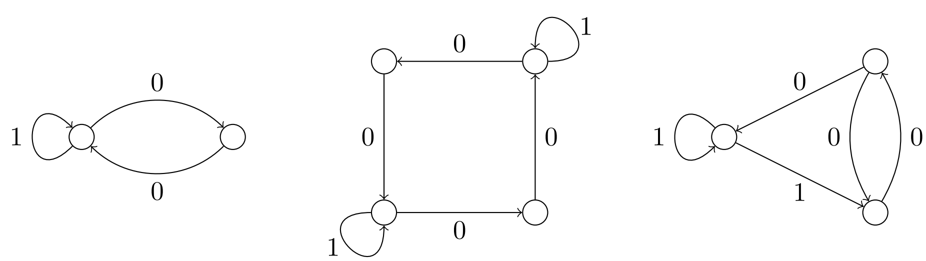

Figure 1: Presentations of the even shift

A *block* is a finite sequence of symbols over an alphabet. Let x be a point from a sofic shift. We way the a block w *occurs* in x if there exists integers i, j such that $x_i x_{i+1} \dots x_j = w$. The *language* of a sofic shift $\mathcal{B}(X)$ is the collection of blocks that occur in any point in X. A sofic shift is *irreducible* if for any pair of blocks $u, v \in \mathcal{B}(X)$, there exists another block $w \in \mathcal{B}(X)$ such that $uwv \in \mathcal{B}(X)$.

A presentation is *right-resolving* if for each vertex in the presentation, the labels of the outgoing edges of that vertex are all distinct. For example, the left and middle graphs in Figure 1 are right-resolving while the right graph is not right-resolving. A graph is *irreducible* if for each pair of vertices, there exists a path in the graph from the first vertex to the second vertex and path from the second vertex to the first vertex. Each graph in Figure 1 is irreducible.

For a labled graph, the *follower set* is the set of labels of the paths that

A *minimal right-resolving* presentation of a sofic shift X is a right-resolving presentation of X having the fewest vertices among all right-resolving presentations of X.

2 Minimization of reducible presentations

From [Lin+95] corollary (3.3.20), from an irreducible right-resolving presentation, we can find the minimal right-resolving presentation by merging vertices in the presentation that have the same follower set, creating a follower-seperated presentation. However, this does not work for reducible graphs, as being follower-separated does not imply minimality. We can see this with this presentation of the even shift:

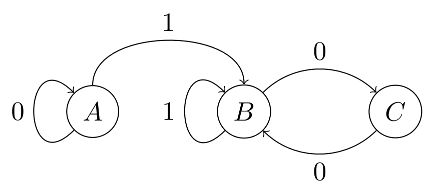

Figure 2: A reducible, follower-separated presentation of the even shift

The block 01 does not appear in the follower set of B, but does for the follower set of A, so $F_{\mathcal{G}}(A) \neq F_{\mathcal{G}}(B)$. The follower set of C is distinct from the follower sets of A and B, as any word that starts with 1 that appears in $F_{\mathcal{G}}(A)$ and $F_{\mathcal{G}}(B)$ does not appear in $F_{\mathcal{G}}(C)$, so $F_{\mathcal{G}}(A) \neq F_{\mathcal{G}}(C)$ and $F_{\mathcal{G}}(B) \neq F_{\mathcal{G}}(C)$. Hence, the graph is follower-separated. This graph also presents the even shift. The label of any bi-infinite walk that visits A has the left-infinite sequence of an infinite number of 0's, followed by a 1, and then followed by a right-infinite sequence from the even shift. Since an infinite number of 0's followed by a 1 is a left-infinite sequence for the even shift, a walk visiting A is in the even shift. Any walk only visiting B and C is a walk in the minimal right-resolving presentation of the even shift (left graph of Figure 1), so such walk would be in the even shift.

References

[Lin+95] Douglas Lind et al. *An introduction to symbolic dynamics and coding.* Cambridge university press, 1995.