2017

Interaction graphs derived from activation functions and their application to gene regulation

Simon Joyce
Binghamton University--SUNY

INTERACTION GRAPHS DERIVED FROM ACTIVATION FUNCTIONS AND THEIR APPLICATION TO GENE REGULATION

BY

SIMON JOYCE

BS, National University of Ireland Maynooth, 2006
MS, National University of Ireland Maynooth, 2007

DISSERTATION

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematical Sciences
in the Graduate School of
Binghamton University
State University of New York
2017

Accepted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematical Sciences
in the Graduate School of
Binghamton University
State University of New York
2017

November 27, 2017

Thomas Zaslavsky, Chair and Faculty Advisor
Department of Mathematical Sciences, Binghamton University

Laura Anderson, Member
Department of Mathematical Sciences, Binghamton University

Fernando Guzman, Member
Department of Mathematical Sciences, Binghamton University

Leslie Lander, Outside Examiner
Department of Computer Science, Binghamton University

Abstract

Interaction graphs are graphic representations of complex networks of mutually interacting components. Their main application is in the field of gene regulatory networks, where they are used to visualize how the expression levels of genes activate or inhibit the expression levels of other genes.

First we develop a natural transformation of activation functions and their derived interaction graphs, called conjugation, that is related to a natural transformation of signed digraphs called switching isomorphism. This is a useful tool for the analysis of interaction graphs used throughout the rest of the dissertation.

We then discuss the question of what restrictions, if any, apply to interaction graphs derived from activation functions. Within these restrictions, we then construct activation functions with any desired interaction graph. The specific case of threshold activation functions, a commonly used kind of activation function, is also considered.

We then conclude with some discussion, and new proofs of the conjectures of René Thomas, using the theory of conjugate activation functions. These conjectures relate feedback in the interaction graph to dynamic properties of multi-stationarity and periodic stability. We prove a more general form of Richard and Comet's version of René Thomas' first conjecture. Included is a new counterexample to a local version of Thomas' second conjecture, on only eight components. This is the smallest counterexample that I am currently aware of.

Acknowledgments

I would like to thank my advisor, Thomas Zaslavsky for introducing me to this subject, for the always stimulating math discussions, for asking questions I would never think to ask. I truly am a much better mathematician thanks to your mentoring.

I would also like to thank all of the faculty and staff at Binghamton University and NUI Maynooth who taught and aided me throughout my student years.

To my parents, without whom, in most ways this statement can be interpreted, I would not be here. Without your love, support and encouragement, I would not have succeeded.

And to my wife Amanda, my steadfast partner, and the love of my life. Through the tough times, you always believed in me. I could not have seen this through without your love and support.

And to all my other family and friends who've helped and supported me throughout the years. You know who you are.

Contents

List of Tables

List of Figures

Introduction

Signed digraphs are directed graphs where each arc is signed, + or -. They are often used to represent complex systems of multiple interacting components, such as gene regulatory networks. Positive or negative arcs between nodes denote positive or negative influence of one component on another. This dissertation is primarily concerned with interaction graphs, a kind of signed directed graph that is derived from an activation function. Interaction graphs, together with their activation function intend to model these kinds of complex systems of multiple interacting components.

One of the main purposes of this dissertation is to determine what, if any, restrictions there are on the structure of interaction graphs. More specifically, for a given signed digraph Δ, when can we find an activation function whose interaction graph is Δ? What can we say about Δ if no such activation function exists?

We also discuss the conjectures or René Thomas. These conjectures, loosely stated, say that positive and negative feedback in the interaction graph are necessary for multi-stationary and stable-periodic dynamics respectively. These phenomenon are of great interest in biology since multi-stationary and stable-periodic dynamics of gene expression levels correspond to processes of cell differentiation and homeostasis respectively. These kinds of dynamics are also of interest in the theory of chemical reaction networks and population models. The strongest versions of René Thomas' conjectures that I am aware of have been proved in [4] and [5]. Using my new techniques developed herein, we will prove the same, or slightly more general versions of these conjectures. We also present a smaller counter example to a local version of René Thomas' second conjecture than the first known counter example presented in [6].

The dissertation is arranged as follows.

Chapter 1 is a short development of relevant background theory on digraphs and signed digraphs that will be used throughout this dissertation.

In chapter 2, we begin with the development of my take on the theory of activation functions. We develop some theory on their associated interaction graphs and present new methods of transforming activation functions and their interaction graphs which aid greatly in their analysis. Then we tackle what, if any, restrictions there are on the structure of interaction graphs as well as how to construct activation functions that have a given interaction graph. We do this first for the general case, and then for the specific case of threshold activation functions.

In chapter 3 we introduce the state transition graph, a graph that represents how the states of an activation function can change over time. Then, using new techniques developed in this dissertation, we explore René Thomas' first and second conjectures, on multi-stationary and stable-periodic dynamics in the state transition graph respectively. Finally, we present a new counter-example to a local version of René Thomas' second conjecture. This counter-example has fewer components than previously known counterexamples.

Chapter 1

Signed Digraphs

1.1 Digraphs

A *directed graph*, or *digraph* is a pair $D = (V, A)$ where V the finite, non-empty *vertex set* of D and A is the finite *arc set* of D, a multiset whose elements are from $V \times V$. For an arc $a = (v, w)$, we say that a is *incident with* v and w. For two subsets of vertices, A and B, we say that there is an arc *between* A and B if there is an arc incident with a vertex in A and a vertex in B.

A digraph D is *connected* if D contains a single vertex, or for all non-empty $W \subset V$, there is an arc between W and $V \setminus W$. For us, $\subset$ means strict-subset, i.e., $W \neq V$.

The *in-degree* (*out-degree*) of v in D is $|\{w \mid (w, v) \in D\}|$ ($|\{w \mid (v, w) \in D\}|$). The *degree* of v is the sum of the in-degree and out-degree of v.

The degree of the vertices and the number of arcs in a digraph are related in a nice way.

Proposition 1.1.1. *Let d_v be the degree of the vertex $v \in D$. Then $\sum_{v \in D} d_v = 2|A|$.*

Proof. Any arc (a, b) will contribute 1 to the sum in d_a and 1 to the sum in d_b. So this arc is counted twice in the sum $\sum_{v \in D} d_v$. Therefore adding the degree of every vertex counts each arc twice. $\square$

Let the bijection $\phi : V \to \phi V$. The digraph $\phi D = (\phi V, \phi A)$ where $\phi A = \{(\phi v, \phi w) \mid (v, w) \in D\}$. Two digraphs D and D' are called *isomorphic* if there exists a bijection ϕ such that $\phi D = D'$.

1.1.1 Subgraphs, Paths, Trees, Circles and Cycles

For the remainder of this section, let the digraphs $D = (V, A)$, $D_1 = (V_1, A_1)$ and $D_2 = (V_2, A_2)$. D_1 is a *sub-digraph* of D_2 if $V_1 \subseteq V_2$ and $A_1 \subseteq A_2$. Often we will call sub-digraphs just *subgraphs*. The union of two digraphs is just the union of their vertex and arc sets respectively, or $D_1 \cup D_2 = (V_1 \cup V_2, A_1 \cup A_2)$.

If $V_1 \subseteq V_2$ and $A_1 = \{(i, j) \in A_2 \mid i, j \in V_1\}$, then we call D_1 *the subgraph of D_2 induced by V_1*. Subgraphs of this form are called *induced subgraphs*.

If D is connected and $|A| = |V| - 1$, then D is called a *tree*. Combining this definition with proposition 1.1.1,

$$\sum_{v \in D} d_v = 2|A| = 2|V| - 2. \tag{1.1}$$

I have much use for trees, so their theory demands development.

Proposition 1.1.2. *A tree T contains a vertex of degree zero if and only if the tree consists of a single vertex.*

Proof. Suppose that T contains a vertex v of degree zero. If $|V| > 1$, then there is an arc to or from v to $V \setminus v$ since T is connected. So the degree of v is at least 1, a contradiction. Therefore $|V| = 1$. Using equation (1.1), $\sum_{v \in V} d_v = 2|V| - 2 = 0 = 2|A|$. Therefore $|A| = 0$, so T contains no arcs and so consists of a single vertex. Conversely, if T consists of a single vertex v, then clearly the degree of v is zero. □

Proposition 1.1.3. *A tree T that contains more than one vertex has at least two vertices of degree-1.*

Proof. By proposition 1.1.2, T has no degree-0 vertex, since T contains more than one vertex. Suppose all but one vertex in T has degree at least two. Then $\sum_{v \in T} d_v \geq 2|V| - 1$. Combining this with equation (1.1), we see that $2|V| - 2 \geq 2|V| - 1$, a contradiction. Therefore T contains at least two degree-1 vertices. □

If the tree T is a subgraph of the digraph D and T contains every vertex of D, then T is called a *spanning-tree* of D.

Proposition 1.1.4. *Every connected digraph $D = (V, A)$ contains a spanning-tree.*

Proof. The proof is an iterative construction.

Any subgraph of D consisting of a single vertex is a tree, so any digraph D contains tree-subgraphs.

So let $T_k = (V_k, A_k)$ be a tree-subgraph of D. If T_k is a spanning-tree, we are done. If not, then there is an arc a_k between V_k and $V \setminus V_k$ since D is connected. Let v_k be the vertex in $V \setminus V_k$ incident with a_k. Let $V_{k+1} = V_k \cup v_k$ and $A_{k+1} = A_k \cup a_k$. I claim the subgraph $T_{k+1} = (V_{k+1}, A_{k+1})$ is also a tree. To justify this, we have to show T_{k+1} is connected and has one fewer arcs than vertices.

Let us first show that T_k one fewer arcs than vertices. Since T_k is a tree, $|A_k| = |V_k| - 1$. Therefore $|A_k \cup a_k| = |V_k \cup v_k| - 1$, i.e., $|A_{k+1}| = |V_{k+1}| - 1$.

To see T_{k+1} is connected, let $W \subset V_{k+1}$ be non-empty. We may assume that W contains v_k since the following argument can be just as easily applied to the compliment of W. If W contains only v_k, then a_k is an arc between W and $V_{k+1} \setminus v_k = V_k$. If W contains vertices other than v_k, then there is an arc between $W \setminus v_k$ and $V_k \setminus W$ in T_k since T_k is connected. Therefore the same arc is between W and $V_{k+1} \setminus W$ in T_{k+1}. Therefore T_{k+1} is connected.

Now if T_{k+1} is a spanning tree, we are done. If not, then we can iterate this process $|V| - |V_k|$ times to get a tree contained in D whose vertex set is V. This tree is then by definition a spanning tree of D. □

A tree that contains no vertex of degree 3 or more is called a *path*.

Proposition 1.1.5. *A path is either a single vertex, or it contains exactly two vertices of degree-1 and all other vertices have degree 2.*

Proof. Let $P = (V, A)$ be the path. If $|V| = 1$, then it P consists of a single vertex by proposition 1.1.2.

What if P contains more than one vertex? By proposition 1.1.3, P contains at least two degree-1 vertices. Since the degree of all other vertices is less than three, $\sum_{v \in V} d_v = |V| - k$, where k is the number of degree-1 vertices. Combining this with equation (1.1), $|V| - 2 = |V| - k$, i.e., $k = 2$. So P contains exactly two degree-1 vertices. □

If a digraph D contains a path such that the degree-1 vertices of the path are v and w, then we say *there is a path between v and w in D*.

Proposition 1.1.6. *In a tree, there is a path between every pair of vertices.*

Proof. The proof is by induction on the number of vertices.

Clearly the proposition is true if the tree consists of a single vertex. If the tree contains exactly two vertices, then each vertex has degree 1 by proposition 1.1.3. So the tree contains a single arc between the two vertices, so it is a path.

Now suppose $T = (V, A)$ is a tree on $k > 2$ vertices and that the proposition is true for trees with less than k vertices. We need the following lemma to complete the induction step.

Lemma 1.1.7. *If $T = (V, A)$ is a tree and v is a degree-1 vertex in T incident with the arc a, then the subgraph $S = (V \setminus v, A \setminus a)$ is a tree.*

Proof. To prove this, we have to show that S is connected and that $|A \setminus a| = |V \setminus a| - 1$.

Since T is a tree, $|A| = |V| - 1$. Therefore $|A \setminus a| = |V \setminus a| - 1$.

Is S connected? Since T has a degree-1 vertex, V contains at least two vertices since it is not just a single vertex by proposition 1.1.2.

If $|V| = 2$, then T contains a single arc and S consists of a single vertex. So the result is true in this case.

If $|V| > 2$, let $W \subset V \setminus v$ be non-empty. This is possible since $V \setminus v$ contains at least two vertices. We just have to show that there is an arc between W and $(V \setminus v) \setminus W$. Since T is connected, there is an arc a between W and $V \setminus W$. Similarly there is an arc b between $W \cup v$ and $(V \setminus v) \setminus W$. If a and b are both incident with v, then a and b are distinct arcs since the other end of a is in W and the other end of b is in $(V \setminus v) \setminus W$. But then v would have degree 2. Since the degree of v is one, at most one of a or b is incident with v. So whichever arc is not incident with v is an arc between W and $(V \setminus v) \setminus W$. Therefore, S is connected, so S satisfies both tree conditions. □

Let us continue the proof of proposition 1.1.6. By proposition 1.1.3, T contains a degree-1 vertex v incident with an arc a. Let $S = (V \setminus v, A \setminus a)$. By lemma 1.1.7, S is a tree on less than k vertices. Therefore it satisfies the proposition by the induction hypothesis. So there is a path between every pair of vertices in S. This mean that in T, there is a path between every pair of vertices excluding v. We just need to show that in addition, there is a path between v and any other vertex $w \in T$. Let x be the other vertex incident with a. Since $x, w \in S$, there is a path between x and w in S. This path is also contained in T. Combining this path with arc between v and x in T yields a path between w and v in T. □

A *directed path*, or *dipath* is a path that consists of a single vertex, or a path that contains a single vertex of out-degree 1, a single vertex of in-degree 1 and all other vertices have in-degree and out-degree 1. In other words, a dipath is a path such that the direction of every arc in the path is consistent. If a digraph D contains a dipath such that the vertex with in-degree 1 is w and the vertex with out-degree 1 is v, then we say *there is a dipath in D from v to w*.

A digraph D is *strongly connected* if for any pair of vertices v and w in D, there is dipath from v to w and a dipath from w to v. A subgraph of D is called a *strong component* of D if it is a maximal strongly connected subgraph of D. A strong component C is *terminal* there are no arcs (v, w) in D such that $v \in C$ and $w \notin C$. A strong component C is *initial* there are no arcs (v, w) in D such that $v \notin C$ and $w \in C$.

Proposition 1.1.8. *The strong components of a digraph partition the vertex set of the digraph.*

Proof. Consider the relation $\sim$ on the vertex set where $v \sim w$ if there is dipath from v to w and a dipath from w to v. if $\sim$ is an equivalence relation, then the equivalence classes of $\sim$ are strong components of the digraph. So to prove the proposition, we will show that $\sim$ is an equivalence relation.

There is always a path from v to itself, so the relation is reflexive. The relation is clearly symmetric. If $v \sim w$ and $w \sim x$ then there is a dipath from v to w and from w to x. The union of two dipaths is not necessarily itself a dipath, but we can use the following lemma.

Lemma 1.1.9. *If P is a dipath from v to w and Q is a dipath from w to x, then $P \cup Q$ contains a dipath from v to x.*

Proof. Starting at v, we follow the dipath P until we encounter the first vertex y contained in Q. This is guaranteed to happen since P contains at least one vertex in Q, namely x. Let $P' \subseteq P$ be the part of P that goes from v to y. Let $Q' \subseteq Q$ be the part of Q that goes from y to x. Both P' and Q' are dipaths that only intersect at y since y is the earliest vertex in P that is also in Q. Therefore $P' \cup Q'$ is a dipath from v to x. $\square$

This lemma shows that $\sim$ is transitive. Therefore $\sim$ is an equivalence relation. $\square$

A *circle* in D is a connected subgraph $\mathcal{C} \subseteq D$ such that the degree of every vertex in $\mathcal{C}$ is 2. If every vertex of $\mathcal{C}$ has in-degree one and out-degree one, then we call $\mathcal{C}$ a *cycle*.

1.2 Signed Digraphs

A *signed digraph* is a pair $\Delta = (V, A)$ where V is the non-empty *vertex set of* Δ and $A \subseteq V \times V \times \{\pm\}$ is the *arc set of* Δ. The vertex set of Δ will also be denoted by $V(\Delta)$. If $(v, w, \sigma_{vw}) \in A$, we say there is an arc from v to w with sign σ_{vw}. An illustration of a signed digraph is given in figure 1.1.

Signed digraphs are themselves digraphs if you ignore the arc signs. So every digraph concept applies equally well to signed digraphs.

If there is at most one arc between any two vertices in the same direction, and at most one loop on any vertex, then the signed digraph can be represented with a matrix. If $\Delta = (V, A)$ is a signed digraph with this property, then the *adjacency matrix of* Δ, $\mathrm{Adj}(\Delta)$ is the $|V| \times |V|$ matrix whose ij entry is σ_{ij} if the arc (i, j, σ_{ij}) is contained in Δ and 0 otherwise.

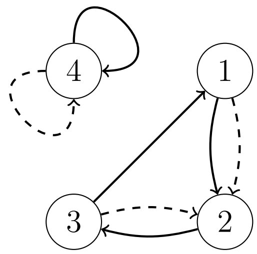

Figure 1.1: A signed digraph with arc set
$\{(1,2,+),(1,2,-),(2,3,+),(3,2,-),(3,1,+),(4,4,+),(4,4,-)\}$.

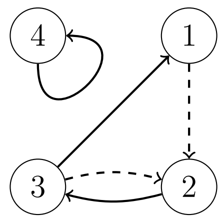

Figure 1.2: Signed digraph Δ that has an adjacency matrix.

For example, the signed digraph Δ in figure 1.2 has the following adjacency matrix.

$$\mathrm{Adj}(\Delta) = \begin{pmatrix} 0 & - & 0 & 0 \\ 0 & 0 & + & 0 \\ + & - & 0 & 0 \\ 0 & 0 & 0 & + \end{pmatrix}$$

However, the signed digraph in figure 1.1 does not have an adjacency matrix because it has a pair of parallel arcs from vertex 1 to vertex 2.

1.2.1 Balance

The *sign of a signed digraph* is the product of the signs of all of its arcs. The sign of Δ is written as $\operatorname{sgn}\Delta = \prod_{(i,j,\sigma_{ij})\in A}\sigma_{ij}$. So a signed digraph is positive (negative) if and only if it contains an even (odd) number of negative arcs.

A signed digraph Δ is *balanced* if every circle contained in Δ is positive. A signed digraph is *cycle-balanced* if every cycle it contains is positive. Examples of balanced and cycle-balanced digraphs are given in figure 1.3. The signed digraph in figure 1.1 is not balanced.

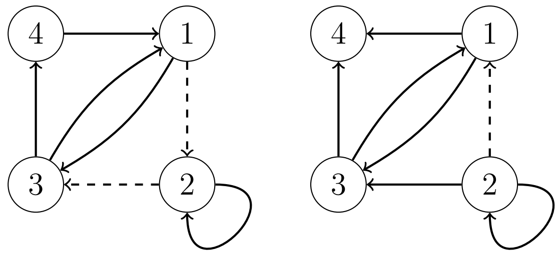

Figure 1.3: A balanced signed digraph on the left and a cycle-balanced digraph that is not balanced on the right.

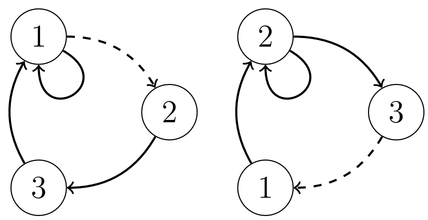

Figure 1.4: Switching isomorphic digraphs Δ and $\lambda\Delta$ where $\lambda = ((1\ 2\ 3), (-, +, -))$.

1.3 Switching Isomorphism

Given a signed digraph Δ with vertex set V. Let the bijection $\phi : V \to W$. We may write $W = \phi V$. Let $\zeta \in \{\pm\}^V$. We call the pair $\lambda = (\phi, \zeta)$ a *switching isomorphism of* Δ. If ϕ is a permutation, then we may also call λ a *signed permutation* of V. The switching isomorphism λ transforms Δ in the following way. The signed digraph $\lambda\Delta = (W, \{(\phi i, \phi j, \zeta_i \sigma_{ij} \zeta_j) \mid (i, j, \sigma_{ij}) \in \Delta\})$. If we write $\zeta\Delta$, this means that $\zeta\Delta = \lambda\Delta$ where the bijection ϕ in λ is the identity map. In this case we can call the map ζ a *switching of* Δ. Similarly $\phi\Delta = \lambda\Delta$ where $\zeta = \{+\}^V$. In this case we call ϕ an *isomorphism of* Δ. Let Δ_1 and Δ_2 be signed digraphs. If there is a switching isomorphism λ of Δ_1 such that $\lambda\Delta_1 = \Delta_2$, then we say that Δ_1 and Δ_2 are *switching isomorphic*. An example of a switching isomorphism is given in figure 1.3.

It is an elementary observation that switching isomorphisms preserve the signs of circles.

Proposition 1.3.1. *Let λ be a switching isomorphism of the signed digraph Δ. If $\mathcal{C} \subseteq \Delta$ is a circle, then* $\operatorname{sgn} \mathcal{C} = \operatorname{sgn} \lambda\mathcal{C}$.

Proof. Note that $\operatorname{sgn} \lambda C = \prod_{(\phi i, \phi j, \zeta_i \sigma_{ij} \zeta_j) \in \lambda C} \zeta_i \sigma_{ij} \zeta_j = \prod_{(i,j,\sigma_{ij}) \in C} \zeta_i \sigma_{ij} \zeta_j$. Since the degree of each vertex in C is two, ζ_i occurs twice in this product for any vertex i of C. Therefore $\prod_{(i,j,\sigma_{ij}) \in C} \zeta_i \sigma_{ij} \zeta_j = \prod_{(i,j,\sigma_{ij}) \in C} \sigma_{ij} = \operatorname{sgn} C$. $\square$

Here is arguably the most important result on balance and switching. This is corollary 3.3 in [8].

Theorem 1.3.2. *A signed digraph Δ is balanced if and only if there is $\zeta \in \{\pm\}^V$ such that every arc in $\zeta\Delta$ is positive.*

Proof. If there is $\zeta \in \{\pm\}^V$ such that every arc in $\zeta\Delta$ is positive, then every circle in Δ is positive by proposition 1.3.1.

Conversely, suppose Δ is balanced. We will construct ζ, a switching of Δ such that $\zeta\Delta$ is positive.

First we will prove this result specifically for trees.

Lemma 1.3.3. *If T is a tree, then there exists ζ, a switching of T, such that every arc of ζT is positive.*

Proof. The proof is by induction.

The proposition is trivially true if the tree consists of a single vertex since then there are no arcs.

Suppose the tree has $k > 1$ vertices and that the result is true for trees with less than k vertices. By proposition 1.1.3, T has a degree-1 vertex v incident with arc a between v and w with sign σ. By lemma 1.1.7, $S = (V \setminus v, A \setminus a)$ is a tree on fewer than k vertices. By assumption, there is ζ' a switching of S such that every arc in $\zeta' S$ is positive. Let ζ be the switching of T such that $\zeta_i = \zeta'_i$ if $i \neq v$ and $\zeta_v = \zeta'_w\sigma$. In ζT, the sign of a is $\zeta_w\zeta_v\sigma = \zeta'_w(\zeta'_w\sigma)\sigma = +$. Any other arc $(i, j, \zeta_i\zeta_j\sigma_{ij})$ in ζT is also positive because $\zeta_i\zeta_j\sigma_{ij} = \zeta'_i\zeta'_j\sigma_{ij} = +$ since $(i, j, \zeta'_i\zeta'_j\sigma_{ij}) \in S$, a positive tree by induction. Therefore every arc in ζT is positive. $\square$

Back to the proof of the theorem. Each connected component of Δ can be treated separately. So for the purposes of this proof we will assume that Δ is connected.

Since Δ is connected it contains a spanning-tree T by proposition 1.1.4. By lemma 1.3.3, there is ζ, a switching of T, such that ζT is positive. Since T and Δ have the same vertices, ζ is also a switching of Δ. Every arc in T will be positive in $\zeta\Delta$ so we just need to show that the arcs outside of T are also positive in $\zeta\Delta$.

Let (i, j, σ_{ij}) be an arc in $\zeta\Delta$ not contained in T. By proposition 1.1.6, there is a path in ζT between i and j. This path together with a forms a circle C in $\zeta\Delta$. Since Δ is balanced, C is positive in $\zeta\Delta$ by proposition 1.3.1. But every arc in the path is positive since ζT is positive. Therefore the sign of C is the sign of a. Therefore $\sigma_{ij} = +$. This shows that every arc in $\zeta\Delta$ is positive, whether in T or not. $\square$

Chapter 2

Interaction Graphs

2.1 Introduction

Interaction graphs are signed digraphs derived from a special kind of function, called an activation function. They are used to represent a group of interacting components, such as a gene regulatory network.

In section 2.2, activation functions and their interaction graphs are defined.

Section 2.3 deals with a useful way of transforming activation functions and their interaction graphs and establishes the relationship between the interaction graphs of an activation function and its conjugate activation functions.

Next are the main results of this chapter. Section 2.4 deals with the question of when a given signed directed graph Δ is a global interaction graph. A general method to construct an activation function whose interaction graph is Δ is then detailed.

Section 2.5 deals with the same question for local interaction graphs and provides a general method to construct an activation function whose local interaction graph is a given signed digraph.

Finally, section 2.6 covers the question of when a given signed directed graph is an interaction graph of a threshold activation function.

2.2 Interaction Graphs

2.2.1 Activation Functions

Let C_f be a finite set. For each $i \in C_f$, let S_i be a non-empty finite set of integers and let $\mathbb{S}_f = \prod_{i \in C_f} S_i$. An *activation function* is a function $f : \mathbb{S}_f \to \mathbb{S}_f$. The set C_f is called the *set of components of* f. The set $\mathbb{S}_f$ is the *state space of* f. We may also refer to $\mathbb{S}_f$ as the *states of* f and call elements of $\mathbb{S}_f$ states. Each component function $f_j : \mathbb{S}_f \to S_j$ is called a *component activation function.* For $\mathbf{x} \in \mathbb{S}_f$, S_i will normally denote the set of possible values of x_i.

If $|S_i| \leq 2$ for all $i \in C_f$, then f is called a *Boolean activation function*.

The states of an activation function f will be ordered using the product ordering. That is for two states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$, $\mathbf{x} \geq \mathbf{y}$ if $x_i \geq y_i$ for all $i \in C_f$.

If $f(\mathbf{x}) = \mathbf{x}$, then we call $\mathbf{x}$ a *steady state of* f.

We use Hamming metric to measure distance between states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. That is, the distance between $\mathbf{x}$ and $\mathbf{y}$ is $|\mathbf{x}, \mathbf{y}| := |\{i \in C_f \mid x_i \neq y_i\}|$.

2.2.2 Interaction Graphs of an Activation Function

For this section we are interested in how changing a single input of an activation function f affects the outputs of f.

First we will establish some notation to aid in the analysis. Let f be an activation function, $\mathbf{x} \in \mathbb{S}_f$, $i \in C_f$, and $a \in S_i$. Define $\mathbf{x}^{i\to a}$ to be the state in $\mathbb{S}_f$ such that

$$(\mathbf{x}^{i\to a})_j := x_j^{i\to a} = \begin{cases} x_j & j \neq i \\ a & j = i. \end{cases} \tag{2.1}$$

The idea here is we are only changing the value of the ith component of $\mathbf{x}$.

We shall be comparing the outputs of $f(\mathbf{x})$ and $f(\mathbf{x}^{i\to a})$ for different values of i and a. We will also want to see if the outputs of f increase or decrease when varying the ith input.

First some notation. Recall that for any real number x,

$$\operatorname{sgn} x = \begin{cases} + & x > 0 \\ 0 & x = 0 \\ - & x < 0. \end{cases}$$

For any $\mathbf{x} \in \mathbb{S}_f$ and $a \in S_i$, let

$$\partial^{i\to a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x})][a - x_i]. \tag{2.2}$$

Let us examine what this definition means. The value of $\partial^{i\to a} f_j(\mathbf{x})$ signifies what effect changing the ith input of f has on the jth output of f. Specifically, suppose $\partial^{i\to a} f_j(\mathbf{x}) = +$. Then $\operatorname{sgn}(f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x})) = \operatorname{sgn}(a - x_i) \neq 0$. So either $f_j(\mathbf{x}^{i\to a}) < f_j(\mathbf{x})$ and $a < x_i$, or $f_j(\mathbf{x}^{i\to a}) > f_j(\mathbf{x})$ and $a > x_i$. Therefore increasing the ith input of f increased the jth output, or decreasing the ith input decreased the jth output. So the jth output of f changes in the same way as the ith input was changed. Similarly, if $\partial^{i\to a} f_j(\mathbf{x}) = -$, then $\operatorname{sgn}(f_j(\mathbf{x}^{i\to y_i}) - f_j(\mathbf{x})) = -\operatorname{sgn}(y_i - x_i) \neq 0$.

So in this case increasing the ith input of f decreased the jth output, or decreasing the ith input increased the jth output. So the value of f_j changed contrary to the change of x_i. If $\partial^{i\to a} f_j(\mathbf{x}) = 0$, then $f_j(\mathbf{x}^{i\to a}) = f_j(\mathbf{x})$ regardless or what the value of a is, so changing x_i has no effect on f_j.

Here is one fact about $\partial^{i\to a} f_j(\mathbf{x})$ that will prove useful. In $\partial^{i\to a} f_j(\mathbf{x})$ we start with $\mathbf{x}$ and change the ith input to a. But what if we start with $\mathbf{x}^{i\to a}$ and change the ith input to x_i? It turns out we get the same sign.

Proposition 2.2.1. *Given an activation function f, $\mathbf{x} \in \mathbb{S}_f$ and $a \in S_i$, then*

$$\partial^{i\to x_i} f_j(\mathbf{x}^{i\to a}) = \partial^{i\to a} f_j(\mathbf{x}).$$

Proof. This is a straightforward calculation based on equation (2.2).

$$\begin{aligned}
\partial^{i\to x_i} f_j(\mathbf{x}^{i\to a}) &= \operatorname{sgn}[f_j([\mathbf{x}^{i\to a}]^{i\to x_i}) - f_j(\mathbf{x}^{i\to a})][x_i - a] \\
&= \operatorname{sgn}[f_j(\mathbf{x}) - f_j(\mathbf{x}^{i\to a})][x_i - a] \\
&= \operatorname{sgn}[-(f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x}))][-(a - x_i)] \\
&= \operatorname{sgn}[f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x})][a - x_i] \\
&= \partial^{i\to a} f_j(\mathbf{x})
\end{aligned}$$

□

Now we will represent graphically how changing the inputs of an activation function affects the outputs. Let f be an activation function and $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. The *local interaction graph of f at $\mathbf{x}$ in the direction of* $\mathbf{y}$ is the signed digraph $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ with vertex set C_f and with arc set $\{(i, j, \sigma_{ij}) \mid \sigma_{ij} = \partial^{i\to y_i} f_j(\mathbf{x}) \neq 0\}$.

For example, let the activation function $f(x_1, x_2) = (1 - 2x_2^2, x_2(x_1 - 1)/2)$ where $S_1 = \{\pm 1\}$ and $S_2 = \{0, \pm 1\}$. Let us find the local interaction graph $\mathcal{I}_f(\langle -1, -1\rangle, \langle 1, 0\rangle)$. We just have to calculate $\partial^{i\to y_i} f_j(\langle -1, -1\rangle)$ for each possible i and j where $\mathbf{y} = \langle 1, 0\rangle$.

$$\begin{aligned}
\partial^{1\to 1} f_1(\langle -1, -1\rangle) &= \operatorname{sgn}[f_1(\langle 1, -1\rangle) - f_1(\langle -1, -1\rangle)][1 - (-1)] = \operatorname{sgn}[-1 - (-1)] = 0 \\
\partial^{2\to 0} f_1(\langle -1, -1\rangle) &= \operatorname{sgn}[f_1(\langle -1, 0\rangle) - f_1(\langle -1, -1\rangle)][0 - (-1)] = \operatorname{sgn}[1 - (-1)] = + \\
\partial^{1\to 1} f_2(\langle -1, -1\rangle) &= \operatorname{sgn}[f_2(\langle 1, -1\rangle) - f_2(\langle -1, -1\rangle)][1 - (-1)] = \operatorname{sgn}[0 - 1] = - \\
\partial^{2\to 0} f_2(\langle -1, -1\rangle) &= \operatorname{sgn}[f_2(\langle -1, 0\rangle) - f_2(\langle -1, -1\rangle)][0 - (-1)] = \operatorname{sgn}[0 - 1] = -
\end{aligned}$$

Therefore the arc set of $\mathcal{I}_f(\langle -1, -1\rangle, \langle 1, 0\rangle)$ is $\{(2, 1, +), (1, 2, -), (2, 2, -)\}$. An illustration of $\mathcal{I}_f(\langle -1, -1\rangle, \langle 1, 0\rangle)$ is given in figure 2.1. In the same figure we also have $\mathcal{I}_f(\langle 1, 1\rangle, \langle -1, 0\rangle)$. You can easily check that $\partial^{1\to -1} f_1(\langle 1, 1\rangle) = \partial^{2\to 0} f_2(\langle 1, 1\rangle) = 0$, $\partial^{1\to -1} f_2(\langle 1, 1\rangle) = +$ and $\partial^{2\to 0} f_1(\langle 1, 1\rangle) = -$.

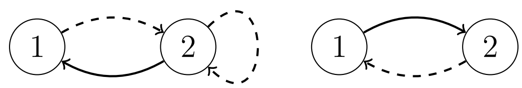

Figure 2.1: Local interaction graphs $\mathcal{I}_f(\langle -1,-1\rangle,\langle 1,0\rangle)$ and $\mathcal{I}_f(\langle 1,1\rangle,\langle -1,0\rangle)$.

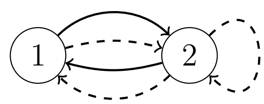

Figure 2.2: Global interaction graph $\mathcal{I}_f$

Since local interaction graphs have at most one arc from any vertex to another, they can be represented by an adjacency matrix. The ijth entry of $\mathrm{Adj}(\mathcal{I}_f(\mathbf{x},\mathbf{y}))$ is $\partial^{i\to y_i} f_j(\mathbf{x})$. We will abbreviate the ith row of this adjacency matrix by $\partial^{i\to y_i} f(\mathbf{x}) := \mathrm{sgn}[f(\mathbf{x}^{i\to y_i}) - f(\mathbf{x})][y_i - x_i]$. So for the activation function $f(x_1, x_2) = (1 - 2x_2^2, x_2(x_1 - 1)/2)$,

$$\mathrm{Adj}(\mathcal{I}_f(\langle -1,-1\rangle,\langle 1,0\rangle)) = \begin{pmatrix} \partial^{1\to 1} f(\langle -1,-1\rangle) \\ \partial^{2\to 0} f(\langle -1,-1\rangle) \end{pmatrix} = \begin{pmatrix} 0 & - \\ + & - \end{pmatrix},$$

$$\mathrm{Adj}(\mathcal{I}_f(\langle 1,1\rangle,\langle -1,0\rangle)) = \begin{pmatrix} \partial^{1\to -1} f(\langle 1,1\rangle) \\ \partial^{2\to 0} f(\langle 1,1\rangle) \end{pmatrix} = \begin{pmatrix} 0 & + \\ - & 0 \end{pmatrix}.$$

The *global interaction graph* of an activation function f, or just the interaction graph of f, is the union of all the local interaction graphs of f. Put another way, $\mathcal{I}_f$ is the signed digraph with vertex set C_f and with arc set $\{(i, j, \sigma_{ij}) \mid \exists\, \mathbf{x} \in \mathbb{S}_f, a \in S_i \text{ such that } \sigma_{ij} = \partial^{i\to a} f_j(\mathbf{x}) \neq 0\}$.

An illustration of the global interaction graph of the activation function $f(x_1, x_2) = (1 - 2x_2^2, x_2(x_1 - 1)/2)$ is given in figure 2.2. Notice that it includes all of the arcs from the two local interaction graphs above. Since f_1 does not depend on x_1 there will be no loops on vertex 1 in the graph. You can also check that $\partial^{2\to y_i} f_2(\mathbf{x}) \leq 0$, so there is no positive loop on vertex 2.

2.2.3 Boolean Interaction Graphs

Given an activation function f, if $|S_i| \leq 2$ for all $i \in C_f$, then f is called a *Boolean activation function* since each x_i has exactly two possible values for each $i \in C_f$.

For Boolean activation functions, $\mathcal{I}_f(\mathbf{x})$ will denote the local interaction graph $\mathcal{I}_f(\mathbf{x},\mathbf{y})$ where $\mathbf{y}$ is the state in $\mathbb{S}_f$ such that $y_i \neq x_i$ for each $i \in C_f$. This is well defined since $|S_i| = 2$ for each

$i \in C_f$. Also $\mathbf{x}^{i\to a}$ where $a \in S_i$ and $a \neq x_i$ will be denoted by $\mathbf{x}^{i\to}$ since the ith coordinate of $\mathbf{x}$ can only be changed to a single value that differs from x_i.

Here is proposition 2.2.1 for Boolean activation functions with our new shorthand notation. The proof is practically identical.

Proposition 2.2.2. *Given a Boolean activation function f, if $\mathbf{x} \in \mathbb{S}_f$, then*

$$\partial^{i\to} f_j(\mathbf{x}^{i\to}) = \partial^{i\to} f_j(\mathbf{x}).$$

2.2.4 A Restriction on Interaction Graphs

My definition of a local interaction graph is a bit looser than standard. Others, such as the definition in [4] and [5], compare the outputs of the activation function when only changing the value of a component by the smallest amount possible. That is, they only consider $\partial^{i\to a} f_j(\mathbf{x})$ when $a \lessdot x_i$ or $a \gtrdot x_i$. The relation $a \lessdot b$ is the cover relation: b covers a, i.e., $a \lessdot x_i$ means that $a < x_i$, and there is no $b \in S_i$ such that $a < b < x_i$. Similarly, $a \gtrdot x_i$ means that $a > x_i$, and there is no $b \in S_i$ such that $a > b > x_i$.

With this in mind, we will define a restricted version of the interaction graphs. If $a > x_i$, let $\partial^{i\to^* a} f_j(\mathbf{x}) := \partial^{i\to b} f_j(\mathbf{x})$ where $b \gtrdot x_i$. Similarly, if $a < x_i$, let $\partial^{i\to^* a} f_j(\mathbf{x}) := \partial^{i\to b} f_j(\mathbf{x})$ where $b \lessdot x_i$. The *restricted local interaction graph of f at* $\mathbf{x}$ *in the direction of* $\mathbf{y}$ is the signed digraph $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$ with vertex set C_f and with arc set $\{(i, j, \sigma_{ij}) \mid \sigma_{ij} = \partial^{i\to^* y_i} f_j(\mathbf{x}) \neq 0\}$.

For any states $\mathbf{x}$ and $\mathbf{y}$, let $\mathbf{x}^{\to^*\mathbf{y}}$ be the state whose ith coordinate satisfies $(\mathbf{x}^{\to^*\mathbf{y}})_i \lessdot x_i$ if $y_i < x_i$, $(\mathbf{x}^{\to^*\mathbf{y}})_i \gtrdot x_i$ if $y_i > x_i$, and $(\mathbf{x}^{\to^*\mathbf{y}})_i = x_i$ otherwise. This way, if $\mathbf{z} = \mathbf{x}^{\to^*\mathbf{y}}$, then

$$\mathbf{x}^{i\to^* y_i} = \mathbf{x}^{i\to z_i}. \tag{2.3}$$

So $\partial^{i\to^* y_i} f_j(\mathbf{x}) = \partial^{i\to z_i} f_j(\mathbf{x})$. From this we get the following result immediately.

Proposition 2.2.3. $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y}) = \mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})$.

Many of our future results on interaction graphs also apply to restricted interaction graphs because of this fact.

Since Boolean interaction graphs can take at most two values on each of their components, there is no distinction between local interaction graphs and restricted ones in this case.

Proposition 2.2.4. *If f is a Boolean activation function, then $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y}) = \mathcal{I}_f(\mathbf{x}, \mathbf{y})$.*

So we do not need to concern ourselves with restricted interaction graphs when working with Boolean activation functions.

The *restricted global interaction graph*, $\mathcal{I}_f^*$ is the union of all the restricted local interaction graphs of f. Some results about local interaction graphs, such as those in section 3.3.1, can be proved for the first kind of local interaction graphs, but are stronger if you prove them for restricted interaction graphs. You may suspect that the same is true for global interaction graphs. Fortunately, we have the following result.

Proposition 2.2.5. *For an activation function f, $\mathcal{I}_f^* = \mathcal{I}_f$.*

Proof. By proposition 2.2.3, $\mathcal{I}_f^* \subseteq \mathcal{I}_f$. So we just need to show the reverse containment. Since both interaction graphs have the same vertex set, we only need to show that any arc contained in $\mathcal{I}_f$ is also contained in $\mathcal{I}_f^*$. To do this we will use the following lemma.

Lemma 2.2.6. *Given an activation function f.*

1. *If the arc $(i, j, +) \notin \mathcal{I}_f^*$, then f_j is weakly decreasing in the ith component. I.e., if $a > x_i$, then $f_j(\mathbf{x}) \geq f_j(\mathbf{x}^{i\to a})$.*

2. *If the arc $(i, j, -) \notin \mathcal{I}_f^*$, then f_j is weakly increasing in the ith component. I.e., if $a > x_i$, then $f_j(\mathbf{x}) \leq f_j(\mathbf{x}^{i\to a})$.*

Proof. For the first item, suppose the arc $(i, j, +) \notin \mathcal{I}_f^*$. Let $\mathbf{x} = \mathbf{x}_1$. So $\partial^{i\to^* a} f_j(\mathbf{x}_1) \neq +$. Using equation (2.2), since $a > x_i$,

$$\partial^{i\to^* a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i\to^* a}) - f_j(\mathbf{x})][a - x_i] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to^* a}) - f_j(\mathbf{x})].$$

So $f_j(\mathbf{x}_1^{i\to^* a}) \leq f_j(\mathbf{x}_1)$.

Now let $\mathbf{x}_2 = \mathbf{x}^{i\to^* a}$. Again, since $(i, j, +) \notin \mathcal{I}_f^*$, $\partial^{i\to^* a} f_j(\mathbf{x}_2) \neq +$. Since $a > x_i$, $\partial^{i\to^* a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}_2^{i\to^* a}) - f_j(\mathbf{x}_2)] \neq +$. So $f_j(\mathbf{x}_2^{i\to^* a}) \leq f_j(\mathbf{x}_2)$ again.

Now let $\mathbf{x}_3 = \mathbf{x}_2^{i\to^* a}$ and repeat this process. Eventually $\mathbf{x}_n = \mathbf{x}^{i\to a}$ after some number of steps and $f_j(\mathbf{x}_{k+1}) \leq f_j(\mathbf{x}_k)$ for all $k \in [n-1]$. Therefore $f_j(\mathbf{x}_n) \leq f_j(\mathbf{x}_1)$, i.e $f_j(\mathbf{x}^{i\to a}) \leq f_j(\mathbf{x})$.

The second item follows by a similar argument. If $(i, j, -) \notin \mathcal{I}_f^*$, then $\partial^{i\to^* a} f_j(\mathbf{x}_1) \neq -$. Since $a > x_i$, $\partial^{i\to^* a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}_1^{i\to^* a}) - f_j(\mathbf{x}_1)]$. So $f_j(\mathbf{x}_1^{i\to^* a}) \geq f_j(\mathbf{x}_1)$.

Now let $\mathbf{x}_2 = \mathbf{x}^{i\to^* a}$. Again, since $(i, j, -) \notin \mathcal{I}_f^*$, $\partial^{i\to^* a} f_j(\mathbf{x}_2) \neq -$. Since $a > x_i$, $\partial^{i\to^* a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}_2^{i\to^* a}) - f_j(\mathbf{x}_2)]$. So $f_j(\mathbf{x}_2^{i\to^* a}) \geq f_j(\mathbf{x}_2)$ again.

Now let $\mathbf{x}_3 = \mathbf{x}_2^{i\to^* a}$ and repeat this process. Eventually $\mathbf{x}_n = \mathbf{x}^{i\to a}$ after some number of steps and $f_j(\mathbf{x}_{k+1}) \geq f_j(\mathbf{x}_k)$ for all $k \in [n-1]$. Therefore $f_j(\mathbf{x}_n) \geq f_j(\mathbf{x}_1)$, i.e $f_j(\mathbf{x}^{i\to a}) \leq f_j(\mathbf{x})$. $\square$

Back to the proof of proposition 2.2.5. Suppose the arc $(i, j, +) \in \mathcal{I}_f$. Then $\partial^{i \to y_i} f_j(\mathbf{x}) = +$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. If $y_i > x_i$, using equation (2.2),

$$\partial^{i \to y_i} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})][y_i - x_i] = \operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})].$$

Since $\partial^{i \to y_i} f_j(\mathbf{x}) = +$, $f_j(\mathbf{x}^{i \to y_i}) > f_j(\mathbf{x})$. By lemma 2.2.6, if $(i, j, +) \notin \mathcal{I}_f^*$ then $f_j(\mathbf{x}^{i \to y_i}) \leq f_j(\mathbf{x})$, a contradiction. Therefore $(i, j, +) \in \mathcal{I}_f^*$. If instead $y_i < x_i$, then

$$\partial^{i \to y_i} f_j(\mathbf{x}) = -\operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})] = +.$$

So $f_j(\mathbf{x}^{i \to y_i}) < f_j(\mathbf{x})$. But if $(i, j, +) \notin \mathcal{I}_f^*$ then $f_j(\mathbf{x}^{i \to y_i}) \geq f_j(\mathbf{x})$ by lemma 2.2.6. Again, this is a contradiction. Therefore $(i, j, +) \in \mathcal{I}_f^*$.

If the arc $(i, j, -) \in \mathcal{I}_f$, we can show that $(i, j, -) \in \mathcal{I}_f^*$ using essentially the same argument. Since $(i, j, -) \in \mathcal{I}_f$, $\partial^{i \to y_i} f_j(\mathbf{x}) = -$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. If $y_i > x_i$, using equation (2.2), $\partial^{i \to y_i} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})]$. Since $\partial^{i \to y_i} f_j(\mathbf{x}) = -$, $f_j(\mathbf{x}^{i \to y_i}) < f_j(\mathbf{x})$. By lemma 2.2.6, if $(i, j, -) \notin \mathcal{I}_f^*$ then $f_j(\mathbf{x}^{i \to y_i}) \geq f_j(\mathbf{x})$, a contradiction. Therefore $(i, j, +) \in \mathcal{I}_f^*$. If instead $y_i < x_i$, then $\partial^{i \to y_i} f_j(\mathbf{x}) = -\operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})] = +$. So $f_j(\mathbf{x}^{i \to y_i}) > f_j(\mathbf{x})$. But if $(i, j, +) \notin \mathcal{I}_f^*$ then $f_j(\mathbf{x}^{i \to y_i}) \leq f_j(\mathbf{x})$ by lemma 2.2.6. Again, this is a contradiction. Therefore $(i, j, -) \in \mathcal{I}_f^*$.

So every arc in $\mathcal{I}_f$ is also contained in $\mathcal{I}_f^*$. Therefore $\mathcal{I}_f = \mathcal{I}_f^*$. $\square$

So when working with global interaction graphs, we can use their more convenient standard definition and all results we prove will also be true for the restricted global interaction graph.

2.3 Conjugate Activation Functions

Given an activation function f, let $\phi : C_f \to W$ be a bijection and $\zeta \in \{\pm\}^{C_f}$. In particular, $\lambda = (\phi, \zeta)$ is a switching isomorphism of $\mathcal{I}_f$ and of $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ for any $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. We will also call λ a switching isomorphism of f for reasons which will become clear in the next section. Also note that $W = \phi C_f$, since it is the codomain of ϕ.

There is a natural way for λ to transform states in $\mathbb{S}_f$. First we will establish how the bijection ϕ transforms states in $\mathbb{S}_f$. If $\mathbf{x} \in \mathbb{S}_f$, $\phi\mathbf{x}$ is the state whose components are in ϕC_f and are such that $\phi_j \mathbf{x} := (\phi\mathbf{x})_j = x_{\phi^{-1} j}$, or $\phi_{\phi i} \mathbf{x} = x_i$.

Let $\circ$ be the Hadamard component-wise product. Define $\lambda\mathbf{x} = (\phi, \zeta)\mathbf{x} := \phi(\zeta \circ \mathbf{x})$. Since $\circ$ is the only kind of product between states and sign-vectors, we normally suppress the $\circ$ notation.

Note that even though $\zeta\mathbf{x}$ may not be in $\mathbb{S}_f$, its components are C_f, so it can still be transformed by ϕ as defined above. So if $j \in \phi C_f$, then

$$\lambda_j\mathbf{x} := (\lambda\mathbf{x})_j = \phi_j(\zeta\mathbf{x}) = (\zeta\mathbf{x})_{\phi^{-1}j} = \zeta_{\phi^{-1}j}x_{\phi^{-1}j}.$$

Or more cleanly,

$$\lambda_{\phi i}\mathbf{x} = \zeta_i x_i. \tag{2.4}$$

So when the switching isomorphism $\lambda = (\phi, \zeta)$ transforms states in $\mathbb{S}_f$ this way, what is the codomain? Let $\zeta_i S_i := \{\zeta_i x \mid x \in X_i\}$. For $j \in \phi C_f$, define $S_j := \zeta_{\phi^{-1}j}S_{\phi^{-1}j}$. More cleanly, $S_{\phi i} := \zeta_i S_i$. Let $\lambda\mathbb{S}_f := \prod_{j\in W} S_j$. This way, $\lambda\mathbb{S}_f$ is the codomain of λ, i.e $\lambda : \mathbb{S}_f \to \lambda\mathbb{S}_f$.

Elements of $\{\pm\}^{C_f}$ can be transformed by ϕ the same way states are, since its components are also indexed by C_f. That is $\phi\zeta \in \{\pm\}^{\phi C_f}$ where $\phi_j\zeta := (\phi\zeta)_j = \zeta_{\phi^{-1}j}$, or $\phi_{\phi i}\zeta = \zeta_i$. It follows that

$$\lambda_j\mathbf{x} = \zeta_{\phi^{-1}j}x_{\phi^{-1}j} = (\phi_j\zeta)(\phi_j\mathbf{x}).$$

Combining this with the previous definition of $\lambda\mathbf{x}$ we get

$$\lambda\mathbf{x} = \phi(\zeta\mathbf{x}) = (\phi\zeta)(\phi\mathbf{x}). \tag{2.5}$$

By this definition, the transformation $\lambda : \mathbb{S}_f \to \lambda\mathbb{S}_f$ is a bijection with inverse transformation $\lambda^{-1} = (\phi^{-1}, \phi\zeta)$ since

$$(\phi^{-1}, \phi\zeta)[(\phi, \zeta)\mathbf{x}] = (\phi^{-1}, \phi\zeta)[(\phi\zeta)(\phi\mathbf{x})] = \phi^{-1}[(\phi\zeta)(\phi\zeta)(\phi\mathbf{x})] = \phi^{-1}[\phi\mathbf{x}] = \mathbf{x}.$$

If λ is a switching isomorphism of an activation function f, the *λ-conjugate activation function of f* is the activation function $f^\lambda : \lambda\mathbb{S}_f \to \lambda\mathbb{S}_f$ with the property that

$$f^\lambda(\lambda\mathbf{x}) = \lambda f(\mathbf{x}). \tag{2.6}$$

The conjugate activation function is defined this way so that figure 2.3 is a commutative diagram. Note that $C_{f^\lambda} = \phi C_f$ and $\mathbb{S}_{f^\lambda} = \lambda\mathbb{S}_f$. Also

$$f^\lambda_{\phi i}(\lambda\mathbf{x}) = [f^\lambda(\lambda\mathbf{x})]_{\phi i} = [\lambda f(\mathbf{x})]_{\phi i} = \zeta_i f_i(\mathbf{x}). \tag{2.7}$$

If $\zeta \in \{\pm\}^{C_f}$, then $f^\zeta = f^\lambda$ where $\lambda = (\phi, \zeta)$ and ϕ is the identity function on C_f. Similarly, if $\phi : C_f \to \phi C_f$, then $f^\phi = f^\lambda$ where $\lambda = (\{+\}^{C_f}, \phi)$. Two activation functions f and g are called *conjugate* if $g = f^\lambda$ for some switching isomorphism λ.

$$\begin{tikzcd} \mathbb{S}_f \arrow[r, "f"] \arrow[d, "\lambda"'] & \mathbb{S}_f \arrow[d, "\lambda"] \\ \lambda\mathbb{S}_f \arrow[r, "f^\lambda"'] & \lambda\mathbb{S}_f \end{tikzcd}$$

Figure 2.3: Commutative diagram that defines the conjugate activation function.

2.3.1 Interaction Graphs of Conjugate Activation Functions

It turns out that the interaction graphs of an activation function and its conjugate activation functions are related in a nice way. This will justify why we are also calling a switching isomorphism of $\mathcal{I}_f$ a switching isomorphism of the activation function f itself.

Theorem 2.3.1. *Let f be an activation function, let $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ and let λ be a switching isomorphism of f. Then;*

1. $\mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y}) = \lambda[\mathcal{I}_f(\mathbf{x}, \mathbf{y})]$, *i.e., $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ and $\mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y})$ are switching isomorphic.*

2. $\mathcal{I}^*_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y}) = \lambda[\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})]$, *i.e., $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$ and $\mathcal{I}^*_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y})$ are switching isomorphic.*

3. $\mathcal{I}_{f^\lambda} = \lambda\mathcal{I}_f$, *i.e., $\mathcal{I}_f$ and $\mathcal{I}_{f^\lambda}$ are switching isomorphic.*

4. $\mathcal{I}^*_{f^\lambda} = \lambda\mathcal{I}^*_f$, *i.e., $\mathcal{I}^*_f$ and $\mathcal{I}^*_{f^\lambda}$ are switching isomorphic.*

Proof. To prove the first item, we will show that $\partial^{\phi i \to \lambda_{\phi i}\mathbf{y}} f^\lambda_{\phi j}(\lambda\mathbf{x}) = \zeta_i \partial^{i \to y_i} f_j(\mathbf{x})\zeta_j$, so that $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$ if and only if $(\phi i, \phi j, \zeta_i\sigma_{ij}\zeta_j) \in \mathcal{I}_{f^\lambda}[\lambda\mathbf{x}, \lambda\mathbf{y}]$. Note that

$$\partial^{\phi i \to \lambda_{\phi i}\mathbf{y}} f^\lambda_{\phi j}(\lambda\mathbf{x}) = \operatorname{sgn}[f^\lambda_{\phi j}(\lambda\mathbf{x}^{\phi i \to \lambda_{\phi i}\mathbf{y}}) - f^\lambda_{\phi j}(\lambda\mathbf{x})][\lambda_{\phi i}\mathbf{y} - \lambda_{\phi i}\mathbf{x}].$$

To complete a calculation, we need the following lemma.

Lemma 2.3.2. *If $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$, then $\lambda(\mathbf{x}^{i \to y_i}) = (\lambda\mathbf{x})^{\phi i \to \lambda_{\phi i}\mathbf{y}}$.*

Proof. We just apply equation (2.4).

$$\begin{aligned}
\lambda_{\phi j}(\mathbf{x}^{i\to y_i}) &= \zeta_j \mathbf{x}_j^{i\to y_i} \\
&= \begin{cases} \zeta_j x_j & j \neq i \\ \zeta_i y_i & j = i \end{cases} \\
&= \begin{cases} \lambda_{\phi j}\mathbf{x} & \phi j \neq \phi i \\ \lambda_{\phi i}\mathbf{y} & \phi j = \phi i \end{cases} \\
&= (\lambda \mathbf{x})_{\phi j}^{\phi i \to \lambda_{\phi i}\mathbf{y}}
\end{aligned}$$

□

We are now ready to complete the proof. Using this lemma, equation (2.7) and equation (2.4),

$$\begin{aligned}
\partial^{\phi i\to\lambda_{\phi i}\mathbf{y}} f_{\phi j}^{\lambda}(\lambda\mathbf{x}) &= \operatorname{sgn}[f_{\phi j}^{\lambda}(\lambda\mathbf{x}^{\phi i\to\lambda_{\phi i}\mathbf{y}}) - f_{\phi j}^{\lambda}(\lambda\mathbf{x})][\lambda_{\phi i}\mathbf{y} - \lambda_{\phi i}\mathbf{x}] \\
&= \operatorname{sgn}[f_{\phi j}^{\lambda}(\lambda[\mathbf{x}^{i\to y_i}]) - f_{\phi j}^{\lambda}(\lambda\mathbf{x})][\lambda_{\phi i}\mathbf{y} - \lambda_{\phi i}\mathbf{x}] \\
&= \operatorname{sgn}[\zeta_j f_j(\mathbf{x}^{i\to y_i}) - \zeta_j f_j(\mathbf{x})][\lambda_{\phi i}\mathbf{y} - \lambda_{\phi i}\mathbf{x}] \\
&= \operatorname{sgn}[\zeta_j f_j(\mathbf{x}^{i\to y_i}) - \zeta_j f_j(\mathbf{x})][\zeta_i y_i - \zeta_i x_i] \\
&= \operatorname{sgn}\zeta_j[f_j(\mathbf{x}^{i\to y_i}) - f_j(\mathbf{x})]\zeta_i[y_i - x_i] \\
&= \zeta_i(\operatorname{sgn}[f_j(\mathbf{x}^{i\to y_i}) - f_j(\mathbf{x})][y_i - x_i])\zeta_j \\
&= \zeta_i \partial^{i\to y_i} f_j(\mathbf{x})\zeta_j.
\end{aligned}$$

Now if $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$, then $\partial^{i\to y_i} f_j(\mathbf{x}) = \sigma_{ij}$. Therefore $(\phi i, \phi j, \zeta_i\sigma_{ij}\zeta_j) \in \mathcal{I}_{f^\lambda}$ if and only if $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

For the second item, we start with the left hand side. By proposition 2.2.3, $\mathcal{I}^*_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y}) = \mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, (\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}})$. Using the same proposition, the right hand side $\lambda[\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})] = \lambda[\mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})]$. By the first item in this theorem, $\lambda[\mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})] = \mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda(\mathbf{x}^{\to^*\mathbf{y}}))$. So to prove the second item, we will use the following lemma.

Lemma 2.3.3. *Given states* $\mathbf{x}, \mathbf{y}$ *and a switching isomorphism* λ *of an activation function* f. *Then* $(\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}} = \lambda(\mathbf{x}^{\to^*\mathbf{y}})$.

Proof. Let $\mathbf{z} = \mathbf{x}^{\to^*\mathbf{y}}$ and $\mathbf{z}' = (\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}}$. We want to show that $\lambda\mathbf{z} = \mathbf{z}'$. We will show this component-wise, i.e., that $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$ for all $i \in C_f$.

First suppose that $y_i > x_i$. Then $z_i \gtrdot x_i$ by definition. If $\zeta_i = +$, then $\zeta_i y_i > \zeta_i x_i$. By equation (2.4), $\lambda_{\phi i}\mathbf{y} > \lambda_{\phi i}\mathbf{x}$. So $z'_{\phi i} \gtrdot \lambda_{\phi i}\mathbf{x}$ by definition. But since $z_i \gtrdot x_i$, $\zeta_i z_i \gtrdot \zeta_i x_i$. So by equation (2.4), $\lambda_{\phi i}\mathbf{z} \gtrdot \lambda_{\phi i}\mathbf{x}$. Therefore $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$. If $\zeta_i = -$, then $\zeta_i y_i < \zeta_i x_i$. By equation (2.4), $\lambda_{\phi i}\mathbf{y} < \lambda_{\phi i}\mathbf{x}$. So $z'_{\phi i} \lessdot \lambda_{\phi i}\mathbf{x}$ by definition. But since $z_i \gtrdot x_i$, $\zeta_i z_i \lessdot \zeta_i x_i$. So by equation (2.4), $\lambda_{\phi i}\mathbf{z} \lessdot \lambda_{\phi i}\mathbf{x}$. Therefore $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$.

Similarly, if $y_i < x_i$, then $z_i \lessdot x_i$ by definition. If $\zeta_i = +$, then $\zeta_i y_i < \zeta_i x_i$. By equation (2.4), $\lambda_{\phi i}\mathbf{y} < \lambda_{\phi i}\mathbf{x}$. So $z'_{\phi i} \lessdot \lambda_{\phi i}\mathbf{x}$ by definition. But since $z_i \lessdot x_i$, $\zeta_i z_i \lessdot \zeta_i x_i$. So by equation (2.4), $\lambda_{\phi i}\mathbf{z} \lessdot \lambda_{\phi i}\mathbf{x}$. Therefore $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$. If $\zeta_i = -$, then $\zeta_i y_i > \zeta_i x_i$. By equation (2.4), $\lambda_{\phi i}\mathbf{y} > \lambda_{\phi i}\mathbf{x}$. So $z'_{\phi i} \gtrdot \lambda_{\phi i}\mathbf{x}$ by definition. But since $z_i \lessdot x_i$, $\zeta_i z_i \gtrdot \zeta_i x_i$. So by equation (2.4), $\lambda_{\phi i}\mathbf{z} \gtrdot \lambda_{\phi i}\mathbf{x}$. Therefore $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$.

We have shown that $\lambda_{\phi i}\mathbf{z} = z'_{\phi i}$ for all $i \in C_f$. Therefore $\lambda\mathbf{z} = \mathbf{z}'$, or $(\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}} = \lambda(\mathbf{x}^{\to^*\mathbf{y}})$. $\square$

By lemma 2.3.3, $\mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, (\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}}) = \mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda(\mathbf{x}^{\to^*\mathbf{y}}))$. By the first item of this theorem $\mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda(\mathbf{x}^{\to^*\mathbf{y}})) = \lambda[\mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})]$. Therefore, $\mathcal{I}^*_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y}) = \lambda[\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})]$ by proposition 2.2.3 since $\mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, (\lambda\mathbf{x})^{\to^*\lambda\mathbf{y}}) = \lambda[\mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})]$. This proves the second part of the theorem.

Finally the same results hold for the global interaction graphs since they are the unions of local interaction graphs. This proves the third and fourth items. $\square$

2.4 Forbidden Interaction Graphs

Can every signed digraph be an interaction graph? More specifically, given any signed digraph Δ, is there an activation function f such that $\mathcal{I}_f = \Delta$? Is there an activation function g such that $\Delta = \mathcal{I}_g(\mathbf{x}, \mathbf{y})$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{S}_g$? If there is no such activation function, what restrictions are there on the structure of Δ to insure that it is an interaction graph? These are the questions that motivate the next few sections.

It turns out that every signed digraph is indeed the interaction graph of some activation function. However, there are restrictions on what kind of signed directed graphs can be the interaction graph of a Boolean activation function, as the following result shows.

Theorem 2.4.1. *Let f be a Boolean activation function. If both $(i, j, +), (i, j, -) \in \mathcal{I}_f$, then the in-degree of j is at least four.*

Proof. Suppose that $(i, j, +), (i, j, -) \in \mathcal{I}_f$ and the in-degree of j is less than four. Then all arcs into j come only from at most two vertices, i and k. That means $\partial^{i'\to} f_j(\mathbf{x}) = 0$ for $i' \neq i, k$, i.e.,

f_j depends only on the ith and kth inputs. Because of this, we will write $f_j(\mathbf{x}) = f_j(x_i, x_k)$ for the purpose of this proof. We will show that under these conditions, both $(k, j, +)$ and $(k, j, -)$ are also in $\mathcal{I}_f$ which will contradict our assumption about the in-degree of j.

There is a state $\mathbf{x}$ such that $\partial^{i\rightarrow} f_j(\mathbf{x}) = +$ because $(i, j, +) \in \mathcal{I}_f$. Let $S_i = \{x_i, x_i'\}$. Since $\partial^{i\rightarrow} f_j(x_i, x_k) = \partial^{i\rightarrow} f_j(x_i', x_k)$ by proposition 2.2.2, we may assume for the sake of argument that $x_i < x_i'$. Putting this assumption into equation (2.2),

$$\partial^{i\rightarrow} f_j(x_i, x_k) = \operatorname{sgn}[f_j(x_i', x_k) - f_j(x_i, x_k)][x_i' - x_i] = \operatorname{sgn}[f_j(x_i', x_k) - f_j(x_i, x_k)].$$

Since $\partial^{i\rightarrow} f_j(x_i, x_k) = +$, $f_j(x_i', x_k) > f_j(x_i, x_k)$. Also $S_j = \{z, z'\}$ where $z < z'$ since f is Boolean. So $f_j(x_i', x_k) = z'$ and $f_j(x_i, x_k) = z$.

Similarly, since $(i, j, -) \in \mathcal{I}_f$, there is a state $\mathbf{y} = (y_i, y_k)$ such that $\partial^{i\rightarrow} f_j(\mathbf{y}) = -$. Now $y_k \neq x_k$ since otherwise $\partial^{i\rightarrow} f_j(\mathbf{y})$ would be the same as $\partial^{i\rightarrow} f_j(x_i, x_k)$ or $\partial^{i\rightarrow} f_j(x_i', x_k)$ which are both positive. So let $y_k = x_k'$. Now by a similar calculation to that above, it follows that $f_j(x_i, x_k') > f_j(x_i', x_k')$, so $f_j(x_i, x_k') = z'$ and $f_j(x_i', x_k') = z$. Based on this,

$$\partial^{k\rightarrow} f_j(x_i, x_k) = \operatorname{sgn}[f_j(x_i, x_k') - f_j(x_i, x_k)][x_k' - x_k] = \operatorname{sgn}[x_k' - x_k].$$

Similarly $\partial^{k\rightarrow} f_j(x_i', x_k) = -\operatorname{sgn}[x_k' - x_k]$. Therefore $\partial^{k\rightarrow} f_j(x_i, x_k) = -\partial^{k\rightarrow} f_j(x_i', x_k)$, and so both $(k, j, +)$ and $(k, j, -)$ are in $\mathcal{I}_f$, so the in-degree of j is four. $\square$

This theorem tells us that for a signed digraph Δ to be an interaction graph of a Boolean activation function, it is necessary that if both $(i, j, +), (i, j, -) \in \Delta$, then the in-degree of j is at least four. It turns out that this condition is also sufficient, which will be proved shortly.

What if the Boolean restriction is removed? It turns out that every signed digraph is an interaction graph of a ternary function. That is an activation function $f : \{0, \pm 1\}^{C_f} \rightarrow \{0, \pm 1\}^{C_f}$. In fact we can be even more specific.

Theorem 2.4.2. *Given a signed digraph Δ. There is an activation function f such that $\mathcal{I}_f = \Delta$, and:*

1. *each component activation function f_i is a polynomial,*

2. *$S_i = \{0, \pm 1\}$ if there is $j \in C_f$ such that $(i, j, +), (i, j, -) \in \Delta$ and the in-degree of j is two or three,*

3. *$S_i = \{\pm 1\}$ otherwise.*

As an immediate corollary to this theorem, the "in-degree four" property is sufficient for a signed digraph Δ to be an interaction graph of a Boolean activation function.

Corollary 2.4.3. *Let Δ be a signed digraph with the property that if both $(i, j, +), (i, j, -) \in \Delta$, then the in-degree of j is at least four. Then there is a Boolean activation function f such that $\Delta = \mathcal{I}_f$, and each f_j is a polynomial.*

Proof. If there are no vertices $i \in C_f$ with the property that $(i, j, +), (i, j, -) \in \Delta$ and the in-degree of j is two or three, then theorem 2.4.2 says there is an activation function f such that $\mathcal{I}_f = \Delta$, each f_i is a polynomial, and $S_i = \{\pm 1\}$ for all $i \in C_f$, i.e., f is a Boolean activation function. $\square$

Proof of theorem 2.4.2. For each $j \in V(\Delta)$, we will provide a method to construct a component activation polynomial f_j of an activation function f such that $\mathcal{I}_f = \Delta$. We will do this by showing that j has the same in-star in $\mathcal{I}_f$ as in Δ, where the *in-star* of j is the set of all arcs into j. For brevity, we will call the in-star of j in $\mathcal{I}_f$ the *in-star of f_j*.

Every component activation polynomial f_j we will use in the proof maps from $\{0, \pm 1\}^{C_f}$ to $\{\pm 1\}$. This way f_j can be used when $S_j = \{\pm 1\}$. We also need to insure that if $(i, j, \sigma_{ij}) \in \Delta$, then there is a state $\mathbf{x} \in \mathbb{S}_f$ such that $x_i = -1$ and $\partial^{i \to 1} f_j(\mathbf{x}) = \sigma_{ij}$ so that if $0 \notin S_i$, $\partial^{i \to 1} f_j(\mathbf{x})$ is well defined for f_j.

Consider the in-star of j in Δ. The in-star contains either no parallel arcs, a single pair of parallel arcs or, multiple pairs of parallel arcs. We will break down the proof by dealing with these three cases separately.

First let us deal with the case when there are no parallel arcs in the in-star of j in Δ. There may be no arcs into j in Δ. Then let f_j be the constant polynomial $f_j(\mathbf{x}) = 1$. This does not depend on any of its inputs so the in-star of f_j will be empty.

Now we provide a way to add an additional non-parallel arc into j in the in-star of a seed component activation function g_j in the interaction graph. We will construct a component activation function f_j whose in-star includes the in-star of g_j and the additional arc. Also, if g_j is a polynomial, then the new function f_j will also be a polynomial. Moreover, the values of the initial component of the new arc need not include 0. This lemma can then be iterated to add as many additional non-parallel arcs as we wish.

Lemma 2.4.4. *Let $g_j : \{0, \pm 1\}^{C_g} \to \{\pm 1\}$ be a component activation function that is not identically -1 and does not depend on input $i \in C_g$. For $\sigma \in \{\pm\}$, let $f_j : \{0, \pm 1\}^{C_g} \to \{\pm 1\}$ be the* *component activation function*

$$f_j(\mathbf{x}) := \frac{1}{2}(g_j(\mathbf{x})+1)(\sigma x_i + x_i^2) - 1 = \begin{cases} 1 & \textit{if } g_j(\mathbf{x}) = \sigma x_i = 1 \\ -1 & \textit{otherwise.} \end{cases} \tag{2.8}$$

Then:

1. *the in-star of f_j is $T \cup (i, j, \sigma)$ where T is the in-star of g_j,*

2. *there is $\mathbf{x} \in \mathbb{S}_f$ such that $x_i = -1$ and $\sigma = \partial^{i\to 1} f_j(\mathbf{x})$.*

Proof. We have to show that the in-star of f_j contains all arcs in the set $T \cup (i, j, \sigma)$ and nothing more.

Let us start by showing that the in-star of f_j contains (i, j, σ) and there is $\mathbf{x} \in \mathbb{S}_g$ such that $x_i = -1$ and $\sigma = \partial^{i\to 1} f_j(\mathbf{x})$. Since g_j is not identically -1, we know that there is a state $\mathbf{x}$ such that $g_j(\mathbf{x}) = 1$. Since g_j does not depend on input i we may assume that $x_i = -1$. This simplifies equation (2.2) slightly since

$$\begin{aligned} \partial^{i\to 1} f_j(\mathbf{x}) &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})][1 - x_i] \\ &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})][1 - (-1)] \\ &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})]. \end{aligned}$$

If $\sigma = +$, then $\sigma x_i = -1$. So by the definition of f_j, $f_j(\mathbf{x}) = -1$. Since g_j does not depend on input i , $g_j(\mathbf{x}) = g_j(\mathbf{x}^{i\to 1}) = 1 = \sigma \mathbf{x}_i^{i\to 1}$. So by the definition of f_j again, $f_j(\mathbf{x}^{i\to 1}) = 1$. Therefore

$$\begin{aligned} \partial^{i\to 1} f_j(\mathbf{x}) &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})] \\ &= \operatorname{sgn}[1 + (-1)] \\ &= +. \end{aligned}$$

Similarly, if $\sigma = -$, then $\partial^{i\to 1} f_j(\mathbf{x}) = \operatorname{sgn}[-1 - 1] = -$. Either way $\partial^{i\to 1} f_j(\mathbf{x}) = \sigma$, so (i, j, σ) is indeed contained in the in-star of f_j and the state $\mathbf{x}$ satisfies our $\{\pm 1\}$ input requirements.

Next, let us show that every arc in T, the in-star of g_j, is also contained in the in-star of f_j. If $(k, j, \sigma_{kj}) \in T$, then there is a state $\mathbf{x}$ such that $\partial^{k\to a} g_j(\mathbf{x}) = \sigma_{kj}$. Since g_j does not depend on input i, we may assume that $\sigma x_i = 1$. If $g_j(\mathbf{x}) = 1$, then $f_j(\mathbf{x}) = 1$ by definition. Similarly, if $g_j(\mathbf{x}) = -1$, then $f_j(\mathbf{x}) = -1$. Therefore $f_j(\mathbf{x}) = g_j(\mathbf{x})$ if $\sigma x_i = 1$. So $\partial^{k\to a} g_j(\mathbf{x}) = \partial^{k\to a} f_j(\mathbf{x}) = \sigma_{kj}$. Therefore (k, j, σ_{kj}) is also contained in the in-star of f_j. It follows that in-star of f_j contains every arc in T.

Next we have to show that every arc in the in-star of f_j is either (i, j, σ), or is contained in the in-star of g_j. Suppose the in-star of f_j contains (k, j, σ_{kj}), i.e., $\partial^{k\to a} f_j(\mathbf{x}) = \sigma_{kj}$. This means that $f_j(\mathbf{x}) \neq f_j(\mathbf{x}^{k\to a})$, so either $f_j(\mathbf{x}) = 1$ or $f_j(\mathbf{x}^{k\to a}) = 1$.

First suppose that $f_j(\mathbf{x}) = 1$. Then $g_j(\mathbf{x}) = \sigma x_i = 1$ by the definition of f_j. Recall that if $\sigma x_i = 1$, then $f_j(\mathbf{x}) = g_j(\mathbf{x})$. So if $k \neq i$,

$$\begin{aligned}\partial^{k\to a} g_j(\mathbf{x}) &= \operatorname{sgn}[g_j(\mathbf{x}^{k\to a}) - g_j(\mathbf{x})][a - x_k] \\ &= \operatorname{sgn}[f_j(\mathbf{x}^{k\to a}) - f_j(\mathbf{x})][a - x_k] \\ &= \partial^{k\to a} f_j(\mathbf{x}) \\ &= \sigma_{kj}.\end{aligned}$$

Therefore $(k, j, \sigma_{kj}) \in T$. Suppose $k = i$, i.e., (i, j, σ_{ij}) is in the in-star of f_j. Since $\sigma x_i = 1$, $\sigma a \neq 1$ because $a \neq x_i$. So by the definition of f_j, $f_j(\mathbf{x}^{i\to a}) = -1$. Therefore

$$\begin{aligned}\sigma_{ij} &= \partial^{i\to a} f_j(\mathbf{x}) \\ &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x})][a - x_i] \\ &= \operatorname{sgn}[-1 - 1][a - x_i] \\ &= -\operatorname{sgn}[a - x_i].\end{aligned}$$

If $\sigma = +$, then $x_i = \sigma x_i = 1$ and $a = -1$ or $a = 0$, so $a - x_i < 0$. It follows that $\sigma_{ij} = +$. Similarly, if $\sigma = -$, then $x_i = -1$. Since a is 0 or 1, $a - x_i > 0$. Again $\sigma_{ij} = -$. In all cases, $\sigma_{ij} = \sigma$. So if the in-star of f_j contains (i, j, σ_{ij}), then $(i, j, \sigma_{ij}) = (i, j, \sigma)$ and that is the only possibility for arcs from i to j.

If we had instead assumed that $f_j(\mathbf{x}^{k\to a}) = 1$, then $g_j(\mathbf{x}^{k\to a}) = \sigma \mathbf{x}_i^{k\to a} = 1$. If $k \neq i$, then $\sigma \mathbf{x}_i^{k\to a} = \sigma x_i = 1$, and $g_j(\mathbf{x}) = f_j(\mathbf{x})$ again. So $\partial^{k\to a} g_j(\mathbf{x}) = \sigma_{kj}$ as before. If $k = i$, then $\sigma \mathbf{x}_i^{k\to a} = \sigma \mathbf{x}_i^{i\to a} = \sigma a = 1$ and (i, j, σ_{ij}) is contained in the in-star of f_j. So

$$\begin{aligned}\sigma_{ij} &= \partial^{i\to a} f_j(\mathbf{x}) \\ &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to a}) - f_j(\mathbf{x})][a - x_i] \\ &= \operatorname{sgn}[1 - (-1)][a - x_i] \\ &= \operatorname{sgn}[a - x_i].\end{aligned}$$

If $\sigma = +$, then $a = \sigma a = 1$ and $x_i = -1$ or $x_i = 0$, so $a - x_i > 0$. Therefore $\sigma_{ij} = +$. Similarly, if $\sigma = -$, then $a = -1$, x_i is 0 or 1, so $a - x_i < 0$ and $\sigma_{ij} = -$. Again $\sigma_{ij} = \sigma$ in all cases. So if the in-star of f_j contains (i, j, σ_{ij}), then $(i, j, \sigma_{ij}) = (i, j, \sigma)$ as before and that is the only possibility for arcs from i to j. Therefore every arc in the in-star of f_j is either (i, j, σ) or contained in T, the in-star of g_j. And so the in-star of f_j is $T \cup (i, j, \sigma)$. □

If we want a component activation polynomial f_j whose in-star contains a single arc, we can start with the constant polynomial $g_j(\mathbf{x}) = 1$ and apply lemma 2.4.4 to get a suitable component activation polynomial f_j. If we want to construct a component activation polynomial f_j whose in-star contains as many non-parallel arcs of any sign that we wish, then we can apply lemma 2.4.4 iteratively to g_j to get the desired activation function. So if the in-star of j in Δ contains no parallel arcs, we can construct a suitable component activation polynomial f_j with the same in-star.

It is worth pointing out that when iterating lemma 2.4.4 as described previously, we get a nice function. If the in-star of j in Δ is $\{(i, j, \sigma_{ij}) \mid i \in A\}$, then the resulting $f_j(\mathbf{x}) = 1$ if and only if $\sigma_{ij} x_i = 1$ for all $i \in A$.

Now let us handle the case when the in-star of j contains a single pair of parallel arcs. Recall that in the Boolean case, there are only two restrictions on a signed graph Δ that prevent it from being the interaction graph of a Boolean activation function. There cannot be a vertex $j \in \Delta$ whose in-star is $\{(i, j, +), (i, j, -)\}$ or $\{(i, j, +), (i, j, -), (k, j, \sigma_{kj})\}$. So let us begin with ternary component activation polynomials with these as in-stars.

Lemma 2.4.5. *Let $f_j : \{0, \pm 1\}^{C_f} \to \{\pm 1\}$ be the component activation polynomial $f_j(\mathbf{x}) = 2x_i^2 - 1$. Then the in-star of f_j is $\{(i, j, +), (i, j, -)\}$.*

Proof. Observe that

$$\partial^{i \to 0} f_j(-1) = \operatorname{sgn}[f_j(0) - f_j(-1)][0 - (-1)] = \operatorname{sgn}[-1 - 1] = -.$$

Also

$$\partial^{i \to 1} f_j(0) = \operatorname{sgn}[f_j(1) - f_j(0)][1 - 0] = \operatorname{sgn}[1 - (-1)] = +.$$

Therefore the arcs $(i, j, +)$ and $(i, j, -)$ are contained in the in-star of j in $\mathcal{I}_f$. Since f_j depends only on x_i, there are no other arcs in the in-star of j. □

If we apply lemma 2.4.4 to the polynomial in lemma 2.4.5, we can construct a component activation polynomial f_j with $\{(i, j, +), (i, j, -), (k, j, \sigma_{kj})\}$ as its in-star where $S_i = \{0, \pm 1\}$. These are the only two cases where we specifically need ternary polynomials.

Next we will show how to produce a component activation polynomial whose in-star contains a single pair of parallel arcs and has in-degree four or more. We start with a polynomial f_j whose in-star contains exactly four arcs. Its inputs are all ± 1 to satisfy the hypotheses of theorem 2.4.2. This polynomial depends only on three inputs x_{i_1}, x_{i_2} and x_{i_3} so we will write $f_j(\mathbf{x}) = f_j(x_{i_1}, x_{i_2}, x_{i_3})$ for the following lemma.

Lemma 2.4.6. *For $\sigma_1, \sigma_2 \in \{\pm\}$, let $f_j : \{0, \pm 1\}^{C_f} \to \{\pm 1\}$ be the component activation polynomial*

$$\begin{aligned} f_j(\mathbf{x}) &= \frac{1}{2}(x_{i_3}[\sigma_1 x_{i_1}(\sigma_1 x_{i_1} + 1)(x_{i_3} + 1) + \sigma_2 x_{i_2}(\sigma_2 x_{i_2} + 1)(x_{i_3} - 1)]) - 1 \\ &= \frac{1}{2}[(x_{i_1}^2 + \sigma_1 x_{i_1} - 1)(x_{i_3}^2 + x_{i_3}) + (x_{i_2}^2 + \sigma_2 x_{i_2} - 1)(x_{i_3}^2 - x_{i_3})] + (x_{i_3}^2 - 1) \\ &= \begin{cases} 1 & \sigma_1 x_{i_1} = x_{i_3} = 1 \textit{ or } \sigma_2 x_{i_2} = -x_{i_3} = 1 \\ -1 & \textit{otherwise.} \end{cases} \end{aligned}$$

Then:

1. *the in-star of f_j is $\{(i_1, j, \sigma_1), (i_2, j, \sigma_2), (i_3, j, +), (i_3, j, -)\}$,*
2. *$\partial^{i_3 \to 1} f_j(x_{i_1}, x_{i_2}, -1) = +$ if $\sigma_1 x_{i_1} = 1$ and $\sigma_2 x_{i_2} = -1$,*
3. *$\partial^{i_3 \to 1} f_j(x_{i_1}, x_{i_2}, -1) = -$,*
4. *$\partial^{i_1 \to 1} f_j(-1, x_{i_2}, 1) = \sigma_1$,*
5. *$\partial^{i_2 \to 1} f_j(x_{i_1}, -1, -1) = \sigma_2$.*

Proof. There are three parts to the proof. First we will show that the in-star of f_j contains the arcs $(i_3, j, +)$ and $(i_3, j, -)$. Then we will show that (i_1, j, σ_1) is the only arc from i_1 to j in the in-star of f_j and that (i_2, j, σ_2) is the only arc from i_2 to j in the in-star of f_j.

First some useful observations about f_j. Notice that

$$\begin{aligned} f_j(x_{i_1}, x_{i_2}, 1) &= \frac{1}{2}[(x_{i_1}^2 + \sigma_1 x_{i_1} - 1)(1^2 + 1) + (x_{i_2}^2 + \sigma_2 x_{i_2} - 1)(1^2 - 1)] + (1^2 - 1) \\ &= x_{i_1}^2 + \sigma_1 x_{i_1} - 1. \end{aligned} \tag{2.9}$$

If in addition $x_{i_1} = \pm 1$, then

$$f_j(x_{i_1}, x_{i_2}, 1) = \sigma_1 x_{i_1}. \tag{2.10}$$

Similarly

$$f_j(x_{i_1}, x_{i_2}, -1) = x_{i_2}^2 + \sigma_2 x_{i_2} - 1. \tag{2.11}$$

If in addition $x_{i_2} = \pm 1$, then

$$f_j(x_{i_1}, x_{i_2}, -1) = \sigma_2 x_{i_2}. \tag{2.12}$$

So if $x_{i_1} = \pm 1$ and $x_{i_2} = \pm 1$, by equations (2.10) and (2.12),

$$\begin{aligned}
\partial^{i_3 \to 1} f_j(x_{i_1}, x_{i_2}, -1) &= \operatorname{sgn}[f_j(x_{i_1}, x_{i_2}, 1) - f_j(x_{i_1}, x_{i_2}, -1)][1 - (-1)] \\
&= \operatorname{sgn}[f_j(x_{i_1}, x_{i_2}, 1) - f_j(x_{i_1}, x_{i_2}, -1)] \\
&= \operatorname{sgn}[\sigma_1 x_{i_1} - \sigma_2 x_{i_2}].
\end{aligned} \tag{2.13}$$

Now suppose that $\sigma_1 x_{i_1} = 1$ and $\sigma_2 x_{i_2} = -1$. In particular, this means that $x_{i_1} = \pm 1$ and $x_{i_2} = \pm 1$. So by equation (2.13),

$$\partial^{i_3 \to 1} f_j(x_{i_1}, x_{i_2}, -1) = \operatorname{sgn}[\sigma_1 x_{i_1} - \sigma_2 x_{i_2}] = \operatorname{sgn}[1 - (-1)] = +.$$

Similarly if $\sigma_1 x_{i_1} = -1$ and $\sigma_2 x_{i_2} = 1$, then

$$\partial^{i_3 \to 1} f_j(x_{i_1}, x_{i_2}, -1) = \operatorname{sgn}[-1 - 1] = -.$$

Therefore $(i_3, j, +)$ and $(i_3, j, -)$ are both contained in the in-star of f_j.

Next we will show that (i_1, j, σ_1) is the only arc from i_1 to j in the in-star of f_j. Consider $\partial^{i_1 \to d} f_j(a, b, c)$. Using lemma 2.2.1 we can assume that $d > a$, so $\partial^{i_1 \to d} f_j(a, b, c) = \operatorname{sgn}[f_j(d, b, c) - f_j(a, b, c)][d - a] = \operatorname{sgn}[f_j(d, b, c) - f_j(a, b, c)]$. Notice that $f_j(d, b, -1) = f_j(a, b, -1)$ by equation (2.11). Also $f_j(d, b, 0) = f_j(a, b, 0) = -1$ from the definition of f_j, so $\partial^{i_1 \to d} f_j(a, b, c) = 0$ if $c \neq 1$. So we need only consider the case when $c = 1$. Using equation (2.9),

$$\begin{aligned}
\partial^{i_1 \to d} f_j(a, b, 1) &= \operatorname{sgn}[f_j(d, b, 1) - f_j(a, b, 1)] \\
&= \operatorname{sgn}[d^2 + \sigma_1 d - 1 - (a^2 + \sigma_1 a - 1)] \\
&= \operatorname{sgn}[(d^2 - a^2) + \sigma_1(d - a)] \\
&= \operatorname{sgn}[(d + a)(d - a) + \sigma_1(d - a)] \\
&= \operatorname{sgn}[(d - a)][d + a + \sigma_1(1)].
\end{aligned}$$

Since $d > a$,

$$\partial^{i_1 \to d} f_j(a, b, 1) = \operatorname{sgn}[(d + a) + \sigma_1(1)].$$

Now $a, d \in \{0 \pm 1\}$, so $-1 \leq (d + a) \leq 1$. Then $(d + a) + 1 \geq 0$ and $(d + a) - 1 \leq 0$.

If $\sigma_1 = +$, then $\partial^{i_1 \to d} f_j(a, b, 1) = \text{sgn}[(d + a) + 1] \geq 0$. If $\sigma_1 = -$, then $\partial^{i_1 \to d} f_j(a, b, 1) = \text{sgn}[(d + a) - 1] \leq 0$. In either case $\partial^{i_1 \to d} f_j(a, b, 1)$ is σ_1 or 0. Specifically,

$$\partial^{i_1 \to 1} f_j(-1, b, 1) = \text{sgn}[(1 + (-1)) + \sigma_1(1)] = \sigma_1.$$

Therefore (i_1, j, σ_1) is the only arc from i_1 to j in the in-star of f_j.

Using a similar argument, we will show that (i_2, j, σ_2) is the only arc from i_2 to j in the in-star of f_j.

Consider $\partial^{i_2 \to d} f_j(a, b, c)$. Using lemma 2.2.1 again we can assume that $d > b$. It follows that $\partial^{i_2 \to d} f_j(a, b, c) = \text{sgn}[f_j(a, d, c) - f_j(a, b, c)]$. Similarly to before, if $c \neq -1$, then $f_j(a, d, c) = f_j(a, b, c)$, so $\partial^{i_2 \to d} f_j(a, b, c) = 0$. By equation (2.11),

$$\begin{aligned}
\partial^{i_2 \to d} f_j(a, b, -1) &= \text{sgn}[d^2 + \sigma_2 d - 1 - (b^2 + \sigma_2 b - 1)] \\
&= \text{sgn}[(d - b)][d + b + \sigma_2(1)] \\
&= \text{sgn}[(d + b) + \sigma_2(1)].
\end{aligned}$$

Again $-1 \leq (d + b) \leq 1$ since $b, d \in \{0, \pm 1\}$. So $(d + b) + 1 \geq 0$ and $(d + b) - 1 \leq 0$.

If $\sigma_2 = +$, then $\partial^{i_2 \to d} f_j(a, b, -1) = \text{sgn}[(d+b)+1] \geq 0$. If $\sigma_2 = -$, then $\text{sgn}[(d+b)-1] \leq 0$. Either way $\partial^{i_2 \to d} f_j(a, b, -1)$ is σ_2 or 0. Specifically,

$$\partial^{i_2 \to 1} f_j(x_{i_1}, -1, -1) = \text{sgn}[(1 + -1) + \sigma_2(1)] = \sigma_2.$$

Therefore (i_2, j, σ_2) is the only arc from i_2 to j in the in-star of f_j. $\square$

Starting with the polynomial in lemma 2.4.6, we can then iteratively apply lemma 2.4.4 to construct a component activation polynomial f_j whose in-star contains $\{(i_1, j, \sigma_1), (i_2, j, \sigma_2), (i_3, j, +), (i_3, j, -)\}$ and as many additional non-parallel arcs as we wish and the inputs can be taken exclusively from $\{\pm 1\}$. This covers the case when the in-star of j in Δ contains a single pair of parallel arcs.

Now suppose there are multiple pairs of parallel arcs into j in Δ.

First we start with a polynomial f_j whose in-star contains only two pairs of parallel arcs and whose inputs can be exclusively from $\{\pm 1\}$. The polynomial depends only on inputs i and k so we will write $f_j(\mathbf{x}) = f_j(x_i, x_k)$ for the purposes of the lemma. Note that $\veebar$ means exclusive-or.

Lemma 2.4.7. *Let $f_j : \{0, \pm 1\}^{C_f} \to \{\pm 1\}$ be a component activation function such that*

$$f_j(\mathbf{x}) = (1 + x_i - x_i^2)(1 + x_k - x_k^2) = \begin{cases} -1 & \textit{if } (x_i = -1) \veebar (x_k = -1) \\ 1 & \textit{otherwise.} \end{cases}$$

Then:

1. *the in-star of f_j is $\{(i,j,+),(i,j,-),(k,j,+),(k,j,-)\}$,*

2. $\partial^{i\to 1} f_j(-1,1) = \partial^{k\to 1} f_j(1,-1) = +$,

3. $\partial^{i\to 1} f_j(-1,-1) = \partial^{k\to 1} f_j(-1,-1) = -$.

Proof. Since f_j only depends on x_i and x_k, all we need to do is calculate $\partial^{i\to 1} f_j(-1,1)$, $\partial^{k\to 1} f_j(1,-1)$, $\partial^{i\to 1} f_j(-1,-1)$, and $\partial^{k\to 1} f_j(-1,-1)$.

$$\begin{aligned}
\partial^{i\to 1} f_j(-1,1) &= \mathrm{sgn}[f_j(1,1) - f_j(-1,1)][1-(-1)] = \mathrm{sgn}[1-(-1)] = +.\\
\partial^{k\to 1} f_j(1,-1) &= \mathrm{sgn}[f_j(1,1) - f_j(1,-1)][1-(-1)] = \mathrm{sgn}[1-(-1)] = +.\\
\partial^{i\to 1} f_j(-1,-1) &= \mathrm{sgn}[f_j(1,-1) - f_j(-1,-1)][1-(-1)] = \mathrm{sgn}[-1-1] = -.\\
\partial^{k\to 1} f_j(-1,-1) &= \mathrm{sgn}[f_j(-1,1) - f_j(-1,-1)][1-(-1)] = \mathrm{sgn}[-1-1] = -.
\end{aligned}$$

This shows that the in-star of f_j is $\{(i,j,+),(i,j,-),(k,j,+),(k,j,-)\}$. $\square$

Combining lemma 2.4.7 with the next, we will be able to construct a component activation polynomial f_j whose in-star contains more than two pairs of parallel arcs.

Lemma 2.4.8. *Let $g_j : \{0,\pm 1\}^{C_g} \to \{\pm 1\}$ be a non-constant component activation function that does not depend on input $i \in C_g$. Let f_j be the component activation function such that $C_f = C_g$ and*

$$f_j(\mathbf{x}) = (1 + x_i - x_i^2)g_j(\mathbf{x}) = \begin{cases} -1 & \text{if } (g_j(\mathbf{x}) = -1) \veebar (x_i = -1) \\ 1 & \text{otherwise.} \end{cases}$$

Then the in-star of f_j contains $T \cup \{(i,j,+),(i,j,-)\}$ where T is the in-star of g_j and there are states $\mathbf{x},\mathbf{y} \in \mathbb{S}_f$ such that $x_i = y_i = -1$, $\partial^{i\to 1} f_j(\mathbf{x}) = +$ and $\partial^{i\to 1} f_j(\mathbf{y}) = -$. Also, if $k \in C_f$ such that $k \neq i$ and there is no arc from k to j in T, then there are no arcs from k to j in the in-star of f_j.

Proof. To prove this, we have to show that the in-star of f_j contains $(i,j,+)$, $(i,j,-)$, and all the arcs in T.

First let us show that every arc contained in T, the in-star of g_j, is also contained in the in-star of f_j. If $(k,j,\sigma_{kj}) \in T$, then there is a state $\mathbf{x}$ such that $\partial^{k\to a} g_j(\mathbf{x}) = \sigma_{kj}$. Since g_j does not depend on input i we may assume that $x_i = 1$. Notice that if $x_i = 1$, then $f_j(\mathbf{x}) = (1+1-1)g_j(\mathbf{x}) = g_j(\mathbf{x})$. Therefore $\partial^{k\to a} f_j(\mathbf{x}) = \partial^{k\to a} g_j(\mathbf{x})$. So the in-star of f_j also contains (k,j,σ_{kj}).

Now let us show that the in-star of f_j contains $(i, j, +)$ and $(i, j, -)$. There is a state $\mathbf{x}$ such that $g_j(\mathbf{x}) = 1$ since g_j is not constant. Since g_j does not depend on input i, $g_j(\mathbf{x}^{i\to 1}) = g_j(\mathbf{x})$ and we may assume that $x_i = -1$. Putting this into the formula for f_j,

$$\begin{aligned}
\partial^{i\to 1} f_j(\mathbf{x}) &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})][1 - (-1)] \\
&= \operatorname{sgn}[f_j(\mathbf{x}^{i\to 1}) - f_j(\mathbf{x})] \\
&= \operatorname{sgn}[(1 + 1 - 1)g_j(\mathbf{x}^{i\to 1}) - (1 - 1 - 1)g_j(\mathbf{x})] \\
&= \operatorname{sgn}[g_j(\mathbf{x}) + g_j(\mathbf{x})] \\
&= \operatorname{sgn}[1 + 1] \\
&= +.
\end{aligned}$$

By a similar argument there is a state $\mathbf{y}$ such that $g_j(\mathbf{y}) = -1$ since g_j is not constant and we may assume that $y_i = -1$. In this case

$$\begin{aligned}
\partial^{i\to 1} f_j(\mathbf{y}) &= \operatorname{sgn}[f_j(\mathbf{y}^{i\to 1}) - f_j(\mathbf{y})] \\
&= \operatorname{sgn}[(1 + 1 - 1)g_j(\mathbf{y}^{i\to 1}) - (1 - 1 - 1)g_j(\mathbf{y})] \\
&= \operatorname{sgn}[g_j(\mathbf{y}) + g_j(\mathbf{y})] \\
&= \operatorname{sgn}[-1 - 1] \\
&= -.
\end{aligned}$$

Therefore the in-star of f_j contains $(i, j, +)$ and $(i, j, -)$ also.

Finally, if $k \in C_f$, and there is no arc from k to j in T, then $g(\mathbf{x}^{k\to a}) = g(\mathbf{x})$ always since g does not depend on k. Therefore $f(\mathbf{x}^{k\to a}) = f(\mathbf{x})$, if $k \neq i$. □

Notice that if g_j in lemma 2.4.8 is a polynomial, then f_j is also a polynomial.

Now we can apply lemma 2.4.8 iteratively to the polynomial in lemma 2.4.7 to give us a component activation polynomial f_j whose in-star contains two or more pairs of parallel arcs.

Now while applying lemma 2.4.8, there is the possibility to add an additional arc from k to j in the in-star of f_j if there is already an arc from k to j in the in-star of g_j, T. But since the polynomial in 2.4.7 starts with pairs of parallel arcs, this is not an issue for our purposes since all possible arcs that could be contained in the in-star of j are present at each step when we iteratively apply lemma 2.4.8 to the polynomial in lemma 2.4.7.

As we have done before, we can then iteratively apply lemma 2.4.4 to construct a component activation polynomial whose in-star contains as many pairs of parallel arcs as we like and as many non-parallel arcs as we like. This covers the case when the in-star of j in Δ contains multiple parallel arc pairs. And at last we have covered all possible cases for the proof of the theorem. So we can indeed construct a component activation polynomial with any desired in-star. $\square$

2.4.1 Restricting to Boolean Activation Functions

It is worth noting that if we restrict to Boolean activation functions, then the polynomials used in the proof of theorem 2.4.2 simplify because if $x \in \{\pm 1\}$, then $x^2 = 1$. The function in lemma 2.4.4 becomes

$$f_j(\mathbf{x}) = \frac{1}{2}(g_j(\mathbf{x}) + 1)(\sigma_{ij}x_i + 1) - 1 = \min(g_j(\mathbf{x}), \sigma_{ij}x_i).$$

So if we want a component activation function f_j whose in-star is $\{(i, j, \sigma_{ij}) \mid i \in A\}$, then $f_j(\mathbf{x}) = \min_{i \in A}(\sigma_{ij}x_i)$ works in the Boolean case. The function in lemma 2.4.6 becomes

$$f_j(\mathbf{x}) = \frac{1}{2}[\sigma_1 x_{i_1}(1 + x_{i_3}) + \sigma_2 x_{i_2}(1 - x_{i_3})] = \begin{cases} \sigma_1 x_{i_1} & x_{i_3} = 1 \\ \sigma_2 x_{i_2} & x_{i_3} = -1. \end{cases}$$

The function in lemma 2.4.7 becomes $f_j(\mathbf{x}) = x_{i_1}x_{i_2}$ and the function in lemma 2.4.8 becomes $f_j(\mathbf{x}) = x_i g_j(\mathbf{x})$. So when lemma 2.4.8 is iteratively applied to the function in lemma 2.4.7, we just get a product. Specifically, if we want a component activation polynomial f_j whose in-star is $\{(i, j, +), (i, j, -) \mid i \in A\}$ where A contains at least two vertices, then $f_j(\mathbf{x}) = \prod_{i \in A} x_i$ will work in the Boolean case.

2.5 Forbidden Local Interaction Graphs

In this section we will address the same question as in the previous section for local interaction graphs. That is, for a given signed digraph Δ, when can we find an activation function f and states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$? We will only discuss the first kind of local interaction graphs, but by proposition 2.2.3, all the results in this section apply to restricted local interaction graphs too.

The first observation is that for this to be possible, Δ cannot contain any parallel arcs since there is a single arc from i to j in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ if and only if $\partial^{i \to y_i} f_j(\mathbf{x})$ is non-zero. It turns out that if this condition is met, then Δ is a local interaction graph. Moreover f is at worst ternary.

For a set of vertices $W \subseteq V(\Delta)$, we can also give a condition for when it is possible to have $|S_i| \leq 2$ for all $i \in W$ for this activation function. For a vertex j in a signed digraph, if every arc into j is positive, then call j *positive inward-homogeneous*. Similarly, if every arc into j is negative, call j *negative inward-homogeneous*. Call j *inward-homogeneous* if it is either positive, or negative inward-homogeneous. For $W \subseteq V$, call W *inward-homogeneous* if every vertex in W is inward-homogeneous. We call a signed digraph *inward-homogeneous* if all of its vertices are inward-homogeneous.

Theorem 2.5.1. *Given a signed digraph Δ containing no parallel arcs.*

1. *There are an activation function f and states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$.*

2. *For each $j \in C_f$, S_j can be chosen to be $\{\pm 1\}$ or $\{0, \pm 1\}$.*

3. *For $W \subseteq C_f$, there is an activation function f such that $|S_j| \leq 2$ for all $j \in W$ if and only if there is $\zeta \in \{\pm\}^{V(\Delta)}$ such that W is inward-homogeneous in $\zeta\Delta$.*

This theorem has an immediate corollary for Boolean activation functions.

Corollary 2.5.2. *A signed digraph Δ is a local interaction graph of a Boolean activation function if and only if Δ contains no parallel arcs and Δ is switching equivalent to an inward-homogeneous digraph.*

Proof. This is simply part three of theorem 2.5.1 when $W = C_f$ □

A natural question arises from this corollary. Given a signed digraph, is there a way to quickly identify whether it is switching equivalent to an inward-homogeneous digraph or not? As of this writing, I still have not found a satisfying answer to this question.

Proof of theorem 2.5.1. Let us begin with the first part of the theorem, that for any signed digraph we can in fact find an appropriate activation function.

We will start with a method of constructing an activation function f such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$ in a simpler case that also addresses the size of S_i for each $i \in C_f$.

Lemma 2.5.3. *Let Δ be a signed digraph that does not contain any parallel arcs. Let $W \subseteq V(\Delta)$ be inward-homogeneous. Then there are an activation function f and $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$ and $S_i = \{\pm 1\}$ if $i \in W$ and $S_i = \{0, \pm 1\}$ otherwise.*

Proof. We have to find an activation function f such that $\partial^{i \to y_i} f_j(\mathbf{x}) = \sigma_{ij}$ if $(i, j, \sigma_{ij}) \in \Delta$ and $\partial^{i \to y_i} f_j(\mathbf{x}) = 0$ otherwise. To completely determine $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$, it will suffice to know the values of $f_j(\mathbf{x})$ and $f_j(\mathbf{x}^{i \to y_i})$ for all $i, j \in C_f$.

Let $\mathbf{x}$ be the state such that $x_i = -1$ for all $i \in C_f$ and let $y_i > x_i$. If $j \in W$ and there are arcs into j in Δ, let $f_j(\mathbf{x}) = (-1)\sigma_j$ where σ_j is the sign of all the arcs into j in Δ. Otherwise, let $f_j(\mathbf{x}) = 1$. Let $f_j(\mathbf{x}^{i\to y_i}) = -f_j(\mathbf{x})$ if there is an arc from i to j in Δ. Under these assumptions,

$$\begin{aligned}\partial^{i\to y_i} f_j(\mathbf{x}) &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to y_i}) - f_j(\mathbf{x})][y_i - x_i] \\ &= \operatorname{sgn}[-2f_j(\mathbf{x})] \\ &= \operatorname{sgn}[-f_j(\mathbf{x})] \\ &= \operatorname{sgn}[-(-1)\sigma_j] \\ &= \sigma_j.\end{aligned}$$

Therefore (i, j, σ_j) is also an arc in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

If there is no arc from i to j in Δ, let $f_j(\mathbf{x}^{i\to y_i}) = f_j(\mathbf{x})$. Then $\partial^{i\to y_i} f_j(\mathbf{x}) = 0$ so there is also no arc from i to j in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

Now suppose $j \notin W$. In this case we let $f_j(\mathbf{x}) = 0$. If $(i, j, \sigma_{ij}) \in \Delta$, let $f_j(\mathbf{x}^{i\to y_i}) = (1)\sigma_{ij}$. Then

$$\partial^{i\to y_i} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i\to y_i}) - f_j(\mathbf{x})][y_i - x_i] = \operatorname{sgn}[(1)\sigma_{ij} - 0] = \sigma_{ij}.$$

Therefore $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

If there is no arc from i to j in Δ then let $f_j(\mathbf{x}^{i\to y_i}) = 0$ so that $\partial^{i\to y_i} f_j(\mathbf{x}) = 0$ and there is no arc from i to j in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$. $\square$

From here, we can strengthen the result.

Lemma 2.5.4. *Given a signed digraph Δ that contains no parallel arcs, $\zeta \in \{\pm\}^{V(\Delta)}$ and $W \subseteq V(\Delta)$ that is inward-homogeneous in $\zeta\Delta$. Then there are an activation function f and $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\Delta = \mathcal{I}_f(\mathbf{x}, \mathbf{y})$, $S_i = \{\pm 1\}$ if $i \in W$ and $S_i = \{0, \pm 1\}$ otherwise.*

Proof. By lemma 2.5.3, there are an activation function g and $\mathbf{x}, \mathbf{y} \in \mathbb{S}_g$ such that $\mathcal{I}_g(\mathbf{x}, \mathbf{y}) = \zeta\Delta$ and $S_i = \{\pm 1\}$ if $i \in W$ and $S_i = \{0, \pm 1\}$ otherwise. Using theorem 2.3.1,

$$\Delta = \zeta(\zeta\Delta) = \zeta\mathcal{I}_g(\mathbf{x}, \mathbf{y}) = \mathcal{I}_{g^\zeta}(\zeta\mathbf{x}, \zeta\mathbf{y}).$$

Now let $f = g^\zeta$. Together with the states $\zeta\mathbf{x}$ and $\zeta\mathbf{y}$, these states and f satisfy all hypotheses of the lemma. $\square$

So now we know that we can construct an activation function f such that $S_j = \{\pm 1\}$ for all j in $W \subseteq C_f$ if there is $\zeta \in \{\pm\}^{C_f}$ such that W is inward-homogeneous in $\zeta\Delta$.

For the last part of the theorem, suppose that there are an activation function f and $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$ and $|S_j| \leq 2$ for all $j \in W$ for some $W \subseteq C_f$. We will construct $\zeta \in \{\pm\}^{C_f}$ such that W is inward-homogeneous in $\zeta\Delta$.

First a useful lemma.

Lemma 2.5.5. *Given an activation function f and states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathbf{x} \leq \mathbf{y}$.*

1. *If $f_j(\mathbf{x}) \geq f_j(\mathbf{x}^{i \to y_i})$ for all $i \in C_f$, then j is negative inward-homogeneous in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.*

2. *If $f_j(\mathbf{x}) \leq f_j(\mathbf{x}^{i \to y_i})$ for all $i \in C_f$, then j is positive inward-homogeneous in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.*

Proof. If $y_i = x_i$, then $\partial^{i \to y_i} f_j(\mathbf{x}) = 0$, so suppose that $y_i > x_i$. This means that $\partial^{i \to y_i} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})]$. So if $f_j(\mathbf{x}) \geq f_j(\mathbf{x}^{i \to y_i})$, then $\partial^{i \to y_i} f_j(\mathbf{x})$ is $-$ or zero. Therefore every arc into j is negative.

Similarly if $f_j(\mathbf{x}) \leq f_j(\mathbf{x}^{i \to y_i})$, then $\partial^{i \to y_i} f_j(\mathbf{x})$ is $+$ or zero. So every arc into j is positive in that case. $\square$

The property $f_j(\mathbf{x}) \geq f_j(\mathbf{x}^{i \to y_i})$ for all $i \in C_f$, or $f_j(\mathbf{x}) \leq f_j(\mathbf{x}^{i \to y_i})$ for all $i \in C_f$ is automatically satisfied if $|S_j| \leq 2$. So the next lemma follows immediately from lemma 2.5.5

Lemma 2.5.6. *Given an activation function f and states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathbf{x} \leq \mathbf{y}$. If $|S_j| \leq 2$, then j is inward-homogeneous in $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.*

From here we are ready to show that if $|S_j| \leq 2$ for all $j \in W \subseteq C_f$, then there is $\zeta \in \{\pm\}^{C_f}$ such that j is inward-homogeneous in $\zeta\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \mathcal{I}_{f^\zeta}(\zeta\mathbf{x}, \zeta\mathbf{y})$.

Lemma 2.5.7. *Given an activation function f and states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$. Let $W \subseteq C_f$ be the set of vertices with the property that $|S_j| \leq 2$ if $j \in W$. Let $\zeta \in \{\pm\}^{C_f}$ where $\zeta_i = -$ if and only if $x_i > y_i$. Then W is inward-homogeneous in $\zeta\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.*

Proof. From the definition of ζ, $\zeta_i x_i < \zeta_i y_i$ if $\zeta_i = -$ and $\zeta_i x_i \leq \zeta_i y_i$ if $\zeta_i = +$. Therefore $\zeta\mathbf{x} \leq \zeta\mathbf{y}$. Since $|S_j| \leq 2$, for any $j \in W$ either $f_j^\zeta(\zeta\mathbf{x}) \leq f_j^\zeta(\zeta\mathbf{x}^{i \to \zeta_i y_i})$ or $f_j^\zeta(\zeta\mathbf{x}) \geq f_j(\zeta\mathbf{x}^{i \to \zeta_i y_i})$ for all $i \in C_f$. Therefore j is inward-homogeneous in $\mathcal{I}_{f^\zeta}(\zeta\mathbf{x}, \zeta\mathbf{y})$. And $\mathcal{I}_{f^\zeta}(\zeta\mathbf{x}, \zeta\mathbf{y}) = \zeta\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ by lemma 2.5.6. $\square$

So now we can construct $\zeta \in \{\pm\}^{V(\Delta)}$ as required and so conclude the proof of theorem 2.4.2. $\square$

What if you have a signed digraph Δ that contains no parallel arcs, but no Boolean activation function f exists such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$. We know now that there are non-Boolean activation functions f such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$ by 2.5.1. Among all of these activation functions, which one is the "most Boolean"? More precisely, among all such activation functions, which one has the most components such that $|S_j| = 2$? By theorem 2.5.1, there is an activation function f such that $|S_j| = 2$ for all $j \in W \subseteq V(\Delta)$ if and only if there is $\zeta \in \{\pm\}^{V(\Delta)}$ such that every vertex in W is inward-homogeneous in $\zeta\Delta$. So the maximal number of Boolean components an activation function f such that $\mathcal{I}_f(\mathbf{x}, \mathbf{y}) = \Delta$ can have is the same as the maximal size of $W \subseteq V(\Delta)$ such that, for some $\zeta \in \{\pm\}^{V(\Delta)}$, every $j \in W$ is inward-homogeneous in $\zeta\Delta$.

2.6 Interaction Graphs of Threshold Activation Functions

Continuing the theme of what kinds of signed digraphs are interaction graphs, in this section we explore the question for threshold activation functions. This is a kind of activation function has been used to model real world gene regulatory networks. See for example [3].

A Boolean activation function $f : \{\pm 1\}^{C_f} \to \{\pm 1\}^{C_f}$ is called a *threshold activation function* if for each $j \in C_f$, there is an $\mathbf{a}^j \in \mathbb{R}^{C_f}$ that defines the output of f_j in the following way. We say the jth component of f *tends active* when $f_j(\mathbf{x}) = 1$ if $\mathbf{a}^j \cdot \mathbf{x} \geq 0$ and $f_j(\mathbf{x}) = -1$ if $\mathbf{a}^j \cdot \mathbf{x} < 0$. Similarly, we say the jth component of f *tends inactive* when $f_j(\mathbf{x}) = 1$ if $\mathbf{a}^j \cdot \mathbf{x} > 0$ and $f_j(\mathbf{x}) = -1$ if $\mathbf{a}^j \cdot \mathbf{x} \leq 0$.

Interaction graphs of threshold activation functions have a convenient condition on their arc signs.

Proposition 2.6.1. *Given a threshold activation function f. If the arc $(i, j, \sigma_{ij}) \in \mathcal{I}_f$, then $\sigma_{ij} = \operatorname{sgn}(a_i^j)$ where $\mathbf{a}^j$ defines f_j.*

Proof. Suppose the arc $(i, j, \sigma_{ij}) \in \mathcal{I}_f$. Then $\sigma_{ij} = \partial^{i\to} f_j(\mathbf{x})$ for some state $\mathbf{x} \in \mathbb{S}_f$. I claim that $\operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})]$ in this case. If this is true, our result will follow because

$$\begin{aligned}
\partial^{i\to} f_j(\mathbf{x}) &= \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})][-x_i - x_i] \\
&= \operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}][-2x_i] \\
&= \operatorname{sgn}[\mathbf{a}^j \cdot (\mathbf{x}^{i\to} - \mathbf{x})][-2x_i] \\
&= \operatorname{sgn}[a_i^j(-2x_i)][-2x_i] \\
&= \operatorname{sgn}[a_i^j].
\end{aligned}$$

To see that $\operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})]$, observe that either $f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x}) < 0$, or $f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x}) > 0$ because $\partial^{i\to} f_j(\mathbf{x}) \neq 0$.

First suppose that $f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x}) < 0$. So $f_j(\mathbf{x}^{i\to}) = -1$ and $f_j(\mathbf{x}) = 1$. If f_j tends active, then $\mathbf{a}^j \cdot \mathbf{x} \geq 0$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} < 0$. If f_j tends inactive, then $\mathbf{a}^j \cdot \mathbf{x} > 0$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} \leq 0$. Either way, $\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x} < 0$. Therefore $\operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})]$.

Now suppose that $f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x}) > 0$. So $f_j(\mathbf{x}^{i\to}) = 1$ and $f_j(\mathbf{x}) = -1$. If f_j tends active, then $\mathbf{a}^j \cdot \mathbf{x} < 0$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} \geq 0$. If f_j tends inactive, then $\mathbf{a}^j \cdot \mathbf{x} \leq 0$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} > 0$. Either way, $\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x} > 0$. Therefore $\operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})]$ again.

Since $\operatorname{sgn}[\mathbf{a}^j \cdot \mathbf{x}^{i\to} - \mathbf{a}^j \cdot \mathbf{x}] = \operatorname{sgn}[f_j(\mathbf{x}^{i\to}) - f_j(\mathbf{x})]$, the result now follows from our earlier calculation. □

2.6.1 Global Interaction Graphs

Proposition 2.6.1 has an immediate corollary.

Corollary 2.6.2. *If f is a threshold activation function, then $\mathcal{I}_f$ contains no parallel arcs.*

Proof. Suppose $\mathcal{I}_f$ contains $(i, j, +)$ and $(i, j, -)$. Since $(i, j, +) \in \mathcal{I}_f$, $\operatorname{sgn}(a_i^j) = +$ by proposition 2.6.1. Similarly, since $(i, j, -) \in \mathcal{I}_f$, $\operatorname{sgn}(a_i^j) = -$, a contradiction. Therefore $(i, j, +)$ and $(i, j, -)$ cannot both be contained in $\mathcal{I}_f$. □

It turns out that this necessary condition of containing no parallel arcs is also sufficient for a signed digraph to be the interaction graph of a threshold activation function.

Theorem 2.6.3. *A signed digraph is the interaction graph of a threshold activation function f if and only if it contains no parallel arcs.*

Proof. For each $j \in \Delta$, we will construct an $\mathbf{a}^j$ that defines the component threshold activation function f_j such that the in-stars of j in both Δ and $\mathcal{I}_f$ are the same.

For $j \in C_f$, let P be the set of all vertices in Δ such that there is a positive arc from each vertex in P to j. Similarly, let N be the set of all vertices in Δ such that there is a negative arc from each vertex in N to j. There are three cases: the in-degree of j is odd, the in-degree of j is even and greater than two, or the in-degree of j is two. We will show how to construct $\mathbf{a}^j$ in each of these cases and show that the in-star of j in $\mathcal{I}_f$ contains every arc in the in-star of j in Δ.

First suppose the in-degree of j is odd. Let $a_i^j = 1$ if $i \in P$, $a_i^j = -1$ if $i \in N$ and $a_i^j = 0$ otherwise. This way if $(i, j, \sigma_{ij}) \in \Delta$, then $\operatorname{sgn}(a_i^j) = \sigma_{ij}$. Let $k \in P \cup N$. Now since $|(P \cup N)|$

is odd, $|(P \cup N) \setminus k|$ is even. Since a_i^j is 1 or -1 for all $i \in P \cup N$, as are the values of x_i for any state, there is a state $\mathbf{x} \in \mathbb{S}_f$ such that

$$\sum_{i \in (P \cup N) \setminus k} a_i^j x_i = 0.$$

You can just alternately add and subtract 1 for example. This works since $|(P \cup N) \setminus k|$ is even. Since $a_i^j = 0$ for all i outside of $P \cup N$, $\mathbf{a}^j \cdot \mathbf{x} = a_k^j x_k$ and $\mathbf{a}^j \cdot \mathbf{x}^{k\rightarrow} = -a_k^j x_k$ and these are non-zero. Therefore $\partial^{k\rightarrow} f_j(\mathbf{x}) = \operatorname{sgn}(a_k^j) = \sigma_{ij}$ by proposition 2.6.1. Therefore $(i, j, \sigma_{ij}) \in \mathcal{I}_f$ if $(i, j, \sigma_{ij}) \in \Delta$.

Now suppose that the in-degree of j is even and more than 2. We choose $m \in P \cup N$ and let $a_m^j = 2$ if $m \in P$, or $a_m^j = -2$ if $m \in N$. For all other $i \neq m$, let $a_i^j = 1$ if $i \in P$, $a_i^j = -1$ if $i \in N$ and $a_i^j = 0$ otherwise. Again this means that if $(i, j, \sigma_{ij}) \in \Delta$, then $\operatorname{sgn}(a_i^j) = \sigma_{ij}$. Now let $k \in (P \cup N)$. If $k \neq m$, then similarly to the odd case, there is a state $\mathbf{x} \in \mathbb{S}_f$ such that

$$\sum_{i \in (P \cup N) \setminus k} a_i^j x_i = 0.$$

For example, we can first let $a_m^j x_m = 2$ then subtract 1 twice. This is possible since $|(P \cup N) \setminus k| > 2$ because $P \cup N$ is even but greater than two. Doing this accounts for three elements of $(P \cup N) \setminus k$, leaving us with an even number of elements. Now we can just alternately add and subtract 1 as we did previously. Therefore $\mathbf{a}^j \cdot \mathbf{x} = a_k^j x_k$ and $\mathbf{a}^j \cdot \mathbf{x}^{k\rightarrow} = -a_k^j x_k$, as before and these are non-zero. So $\partial^{i\rightarrow} f_j(\mathbf{x}) = \operatorname{sgn}(a_i^j)$ by proposition 2.6.1. If $k = m$, then there is a state $\mathbf{x} \in \mathbb{S}_f$ such that

$$\sum_{i \in (P \cup N) \setminus m} a_i^j x_1 = 1.$$

Just alternately add and subtract 1 again, which will yield a final total of 1 since $|(P \cup N) \setminus m|$ is odd. Therefore $\mathbf{a}^j \cdot \mathbf{x} = 1 + a_m^j x_m$ and $\mathbf{a}^j \cdot \mathbf{x}^{m\rightarrow} = 1 - a_m^j x_m$. Since $a_m^j = \pm 2$, one of these is 3 and the other is -1. Therefore $f_j(\mathbf{x}) = -f_j(\mathbf{x}^{m\rightarrow})$, so $\partial^{m\rightarrow} f_j(\mathbf{x}) = \operatorname{sgn}(a_m^j)$ by proposition 2.6.1. Therefore $(i, j, \sigma_{ij}) \in \mathcal{I}_f$ if $(i, j, \sigma_{ij}) \in \Delta$.

If the in-degree of j is 2, then let (i_1, j, σ_1) and (i_2, j, σ_2) be the arcs into j. Let $\mathbf{a}^j \in \mathbb{R}^{C_f}$ where $a_{i_1}^j = 1$ if $\sigma_1 = +$, $a_{i_1}^j = -1$ if $\sigma_1 = -$, $a_{i_2}^j = 1$ if $\sigma_2 = +$, $a_{i_2}^j = -1$ if $\sigma_2 = -$, and all other $a_k^j = 0$. Now let $\mathbf{x} \in \mathbb{S}_f$ such that $a_{i_1}^j x_{i_1} = a_{i_2}^j x_{i_2} = 1$. Then $\mathbf{a^j} \cdot \mathbf{x} = 2$ and $\mathbf{a^j} \cdot \mathbf{x}^{i_1\rightarrow} = \mathbf{a^j} \cdot \mathbf{x}^{i_2\rightarrow} = 0$. So if f_j is the tending inactive component threshold activation function defined by $\mathbf{a}^j$, then $f_j(\mathbf{x}) = 1$ and $f_j(\mathbf{x}^{i_1\rightarrow}) = f_j(\mathbf{x}^{i_2\rightarrow}) = -1$. Therefore $\partial^{i_1\rightarrow} f_j(\mathbf{x}) = \sigma_1$ and $\partial^{i_2\rightarrow} f_j(\mathbf{x}) = \sigma_2$ by proposition 2.6.1. So (i_1, j, σ_1) and (i_2, j, σ_2) are arcs in $\mathcal{I}_f$.

Finally we have to show that every arc in $\mathcal{I}_f$ is contained in Δ. We will argue the contrapositive statement. If there is no arc from i to j in Δ, then $\mathbf{a}_i^j = 0$ by definition. By proposition 2.6.1, there is no arc from i to j in $\mathcal{I}_f$. Therefore $(i, j, \sigma_{ij}) \in \mathcal{I}_f$ if and only if $(i, j, \sigma) \in \Delta$. $\square$

It is worth pointing out that in order to have the in-degree of a component j be two, then $|a_{i_1}^j| = |a_{i_2}^j|$. To see this, just observe that if $|a_{i_1}^j| > |a_{i_2}^j|$, and $a_{i_1}^j x_{i_1} + a_{i_2}^j x_{i_2} > 0$ say, then $a_{i_1}^j x_{i_1} - a_{i_2}^j x_{i_2} > 0$ also. This means that $f_j(\mathbf{x}) = f_j(\mathbf{x}^{i_2\to})$, so $\partial^{i_2\to} f_j(\mathbf{x}) = 0$ for all states. This is why we had to invoke that the threshold activation function is tending inactive only in the in-degree two case in the previous theorem whereas this was not required for any other in-degree.

2.6.2 Local Interaction Graphs

What restrictions are there on which signed digraphs are local interaction graphs of threshold activation functions? That is, if Δ is a signed digraph, then is there a threshold activation function f and a state $\mathbf{x} \in \mathbb{S}_f$ such that $\mathcal{I}_f(\mathbf{x}) = \Delta$? It turns out that there are no restrictions on fewer than three components, save for having no parallel arcs and the condition of inward-homogeneity from theorem 2.5.1.

First consider the case where $|C_f| = 1$. Then $f(x) = x$ is a threshold activation function where $a = 1$. Also, $\partial^{1\to} f(1) = +$. So $\mathcal{I}_f(1)$ consists of a single positive loop. Similarly, $f(x) = -x$ is a threshold activation function where $a = -1$, where $\partial^{1\to} f(1) = -$. So $\mathcal{I}_f(1)$ consists of a single negative loop. And $f(x) = 0$ is a threshold activation function where $a = 0$, where $\partial^{1\to} f(x) = 0$. So $\mathcal{I}_f(1)$ contains no arcs. So every possible graph on a single vertex that does not contain parallel arcs is the local interaction graph of a threshold activation function.

Now consider the case where $|C_f| = 2$. Let us say $C_f = \{1, 2\}$. By theorem 2.5.1, there is $\zeta \in \{\pm\}^2$ such that $\zeta\Delta$ is inward-homogeneous. Let (a, b) define f_1. By proposition 2.6.1, there is no corresponding arc in $\mathcal{I}_f(\mathbf{x})$ into 1 for any state $\mathbf{x}$ if a or b is zero. So this gives us a straightforward way to handle the absence of arcs in Δ.

First suppose that the arcs into 1 in $\zeta\Delta$ are $+$. Then let $a = 1$ if there is an arc from 1 to itself in Δ and $a = 0$ otherwise. Similarly, let $b = 1$ if there is an arc from 2 to 1 in Δ and $b = 0$ otherwise. I claim that if f_1 is the tending inactive threshold activation function defined by (a, b), then the in-star of 1 in $\mathcal{I}_f(1, 1)$ is the same as that in $\zeta\Delta$. We already know what happens when either a or b is 0, no corresponding arc. If $a = 1$, then $(a, b) \cdot (1, 1) = a + b \geq 1$ and $(a, b) \cdot (-1, 1) = -a + b \leq 0$ since b is 0 or 1. Therefore $f_1(1, 1) = 1$ and $f_1(-1, 1) = -1$ since f_1 tends inactive. Therefore $\partial^{1\to} f_1(1, 1) = \mathrm{sgn}[f_1(-1, 1) - f_1(1, 1)][-1 - 1] = \mathrm{sgn}[-1 - 1][-2] = +$. Similarly, if $b = 1$, then

$(a,b)\cdot(1,1) = a+b \geq 1$ and $(a,b)\cdot(1,-1) = -a+b \leq 0$ since a is 0 or 1. Therefore $f_1(1,1) = 1$ and $f_1(1,-1) = -1$ again. So $\partial^{2\to} f_1(1,1) = \text{sgn}[f_1(1,-1) - f_1(1,1)][-1-1] = +$ also.

Now suppose that the arcs into 1 in $\zeta\Delta$ are $-$. Let $a = -1$ if there is an arc from 1 to itself in Δ and $a = 0$ otherwise. Similarly, let $b = -1$ if there is an arc from 2 to 1 in Δ and $b = 0$ otherwise. I claim that if f_1 is the active tending threshold activation function defined by (a,b), then the in-star of 1 in $\mathcal{I}_f(1,1)$ is the same as that in $\zeta\Delta$. Again, we already know what happens when either a or b is 0. If $a = -1$, then $(a,b)\cdot(1,1) = a+b \leq -1$ and $(a,b)\cdot(-1,1) = -a+b \geq 0$ since b is 0 or -1. Therefore $f_1(1,1) = -1$ and $f_1(-1,1) = 1$ since f_1 tends active. Therefore $\partial^{1\to} f_1(1,1) = \text{sgn}[f_1(-1,1) - f_1(1,1)][-1-1] = \text{sgn}[1-(-1)][-2] = -$. Similarly, if $b = -1$, then $(a,b)\cdot(1,1) = a+b \leq -1$ and $(a,b)\cdot(1,-1) = -a+b \geq 0$ since a is 0 or -1. Therefore $f_1(1,1) = -1$ and $f_1(1,-1) = 1$ again. So $\partial^{2\to} f_1(1,1) = \text{sgn}[f_1(1,-1) - f_1(1,1)][-1-1] = -$ also.

Now f_2 can be defined in exactly the same way depending on the sign of the arcs into 2 in $\zeta\Delta$. That is, if the arcs into 2 are positive in $\zeta\Delta$, then let f_2 be the tending inactive component threshold activation function where $a = 1$ if there is an arc from 2 to 1 in Δ and $a = 0$ otherwise, and $b = 1$ if there is an arc from 2 to itself in Δ and $b = 0$ otherwise. If the sign of the arcs into 2 are negative in $\zeta\Delta$, then let f_2 be the active tending component threshold activation function where $a = -1$ if there is an arc from 2 to 1 in Δ and $a = 0$ otherwise, and $b = -1$ if there is an arc from 2 to itself in Δ and $b = 0$ otherwise. Then all calculations for $\partial^{i\to} f_2(1,1)$ are the same as before.

So we have shown that $\mathcal{I}_f(1,1) = \zeta\Delta$. Therefore $\Delta = \zeta\mathcal{I}_f(\mathbf{x}) = \mathcal{I}_{f^\zeta}(\zeta\mathbf{x})$ by theorem 2.3.1. So we now use the following lemma to finish the argument, which we'll also need for the case when Δ has more than two vertices.

Lemma 2.6.4. *Let f be a threshold activation function and let $\lambda = (\phi, \zeta)$ be a switching isomorphism of f. Then f^λ is also a threshold activation function.*

Proof. I claim that f^λ is the same function as the threshold activation function g where $g_{\phi j}$ is defined by $\zeta_j(\lambda \mathbf{a}^j)$ where $\mathbf{a}^j$ defines f_j. If $\zeta_j = +$, then the tendency of $g_{\phi j}$, active or inactive, is the same as the tendency of f_j. If $\zeta_j = -$, then the tendency of $g_{\phi j}$ is the opposite of the tendency of f_j. So we just have to show that $f^\lambda = g$ which we will do by showing they always have the same output on each component.

Since we have to compare dot products in $\mathbb{S}_f$ with dot products in $\mathbb{S}_f^\lambda$, tho following equation will be useful. So for a state $\mathbf{x} \in \mathbb{S}_f$ and $\mathbf{a}^j \in \mathbb{R}^{C_f}$,

$$\begin{aligned}
\mathbf{a}^j \cdot \mathbf{x} &= \zeta\mathbf{a}^j \cdot \zeta\mathbf{x} \\
&= \phi(\zeta\mathbf{a}^j) \cdot \phi(\zeta\mathbf{x}) \\
&= \lambda\mathbf{a}^j \cdot \lambda\mathbf{x} \qquad (2.14)
\end{aligned}$$

First suppose that $\zeta_j = +$. If $f^\lambda_{\phi j}(\lambda\mathbf{x}) = 1$, then $f_j(\mathbf{x}) = 1$ by equation (2.7). If f_j tends inactive, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} > 0$ by equation (2.14). Since $\zeta_j = +$, $g_{\phi j}$ also tends inactive and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} > 0$. Therefore $g_{\phi j}(\lambda x) = 1$ also. If f_j tends active, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} \geq 0$ by equation (2.14). Since $\zeta_j = +$, $g_{\phi j}$ also tends active and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} \geq 0$. Therefore $g_{\phi j}(\lambda x) = 1$ again. If $f^\lambda_{\phi j}(\lambda\mathbf{x}) = -1$, then $f_j(\mathbf{x}) = -1$ by equation (2.7). If f_j tends inactive, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} \leq 0$ by equation (2.14). Since $\zeta_j = +$, $g_{\phi j}$ also tends inactive and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} \leq 0$. Therefore $g_{\phi j}(\lambda x) = -1$ also. If f_j tends active, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} < 0$ by equation (2.14). Since $\zeta_j = +$, $g_{\phi j}$ also tends active and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} < 0$. Therefore $g_{\phi j}(\lambda x) = -1$ again.

Now suppose that $\zeta_j = -$. If $f^\lambda_{\phi j}(\lambda\mathbf{x}) = 1$, then $f_j(\mathbf{x}) = -1$ by equation (2.7). If f_j tends inactive, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} \leq 0$ by equation (2.14). Since $\zeta_j = -$, $g_{\phi j}$ tends active and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} \geq 0$. Therefore $g_{\phi j}(\lambda x) = 1$ also. If f_j tends active, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} < 0$ by equation (2.14). Since $\zeta_j = -$, $g_{\phi j}$ tends inactive and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} > 0$. Therefore $g_{\phi j}(\lambda x) = 1$ again. If $f^\lambda_{\phi j}(\lambda\mathbf{x}) = -1$, then $f_j(\mathbf{x}) = 1$ by equation (2.7). If f_j tends inactive, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} > 0$ by equation (2.14). Since $\zeta_j = -$, $g_{\phi j}$ tends active and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} < 0$. Therefore $g_{\phi j}(\lambda x) = -1$ also. If f_j tends active, $\lambda\mathbf{a}^j \cdot \lambda\mathbf{x} \geq 0$ by equation (2.14). Since $\zeta_j = -$, $g_{\phi j}$ tends inactive and $\zeta_j(\lambda\mathbf{a}^j) \cdot \lambda\mathbf{x} \leq 0$. Therefore $g_{\phi j}(\lambda x) = -1$ again.

So we have shown that in all cases, $g_{\phi j}(\lambda\mathbf{x}) = f^\lambda_{\phi j}(\lambda\mathbf{x})$. Therefore f^λ is indeed a threshold activation function. □

So every signed digraph on two components that contains no parallel arcs and is switching equivalent to an inward-homogeneous digraph is a local interaction graph of a threshold activation function. However, the situation changes when there are more than two components.

Proposition 2.6.5. *Let f be a threshold activation function such that $n = |C_f| > 2$. Then the in-degree of any component in $\mathcal{I}_f(\mathbf{x})$ is at most $n - 1$.*

Proof. First observe that

$$\begin{aligned}\sum_{i\in C_f} \mathbf{a}^j \cdot \mathbf{x}^{i\to} &= \mathbf{a}^j \cdot \left(\sum_{i\in C_f} \mathbf{x}^{i\to}\right)\\ &= \mathbf{a}^j \cdot ([n-2]\mathbf{x})\\ &= (n-2)\mathbf{a}^j \cdot \mathbf{x} \end{aligned} \tag{2.15}$$

Now suppose there is $j \in C_f$ in $\mathcal{I}_f(\mathbf{x})$ whose in-degree is n. This means that $f_j(\mathbf{x}) \neq f_j(\mathbf{x}^{i\to})$ for all $i \in C_f$. If for all $i \in C_f$, $\mathbf{a}^j \cdot \mathbf{x}^{i\to} > 0$, $\mathbf{a}^j \cdot \mathbf{x}^{i\to} \geq 0$, $\mathbf{a}^j \cdot \mathbf{x}^{i\to} < 0$, or $\mathbf{a}^j \cdot \mathbf{x}^{i\to} \leq 0$, then by equation (2.15), $\mathbf{a}^j \cdot \mathbf{x} > 0$, $\mathbf{a}^j \cdot \mathbf{x} \geq 0$, $\mathbf{a}^j \cdot \mathbf{x} < 0$, or $\mathbf{a}^j \cdot \mathbf{x} \leq 0$ respectively. But this means that $f_j(\mathbf{x}) = f_j(\mathbf{x}^{i\to})$ for all $i \in C_f$, so there are no arcs into j in $\mathcal{I}_f(\mathbf{x})$, a contradiction. Therefore the in-degree of j cannot be n. ☐

It turns out that this is the only additional restriction on local interaction graphs of threshold activation functions.

Theorem 2.6.6. *Given a signed digraph Δ.*

1. *If Δ has 1 or 2 vertices, then it is the local interaction graph of a threshold activation function if and only if it contains no parallel arcs and is switching equivalent to an inward-homogeneous signed digraph.*

2. *If Δ has more than 2 vertices, then it is the local interaction graph of a threshold activation function if and only if it contains no parallel arcs, is switching equivalent to an inward-homogeneous signed digraph, and it's maximum in-degree is $(n-1)$.*

Proof. There is $\zeta \in \{\pm 1\}^{C_f}$ such that $\zeta\Delta$ is inward-homogeneous. For each $j \in C_f$, we will construct an $\mathbf{a}^j \in \mathbb{R}^{C_f}$ that defines the component threshold activation function f_j with the same in-star as j in $\zeta\Delta$.

Let A be the set of all vertices with arcs going to j in Δ. Let the state $\mathbf{x} = \{1\}^{C_f}$. If A is empty, then let $\mathbf{a}^j = \{0\}^{C_f}$. This way, the in-star of j is empty in $\mathcal{I}_f(\mathbf{x})$ by proposition 2.6.1. So from now on, we will assume that A is non-empty.

First suppose every arc into j is positive in $\zeta\Delta$. Let $a_i^j = 1$ for all $i \in A$. Now since the in-degree of j is less than n, there is at least one $k \in C_f \setminus A$. Let $a_k^j = 1 - |A|$ and $a_i^j = 0$ for all other $i \in C_f \setminus (A \cup k)$. Observe that $\mathbf{a}^j \cdot \mathbf{x} = 1$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} = -1$ for all $i \in A$. This means that $f_j(\mathbf{x}) = 1$ and $f_j(\mathbf{x}^{i\to}) = -1$ for all $i \in A$. So since $a_i^j > 0$ for all $i \in A$, $\partial^{i\to} f_j(\mathbf{x}) = +$

for all $i \in A$ by proposition 2.6.1. Now $\mathbf{a}^j \cdot \mathbf{x}^{k\to} = 2|A| - 1 > 0$ since A is non-empty. So $f_j(\mathbf{x}) = f_j(\mathbf{x}^{k\to}) = 1$. And since $a_i^j = 0$ for all other i outside of A, $f_j(\mathbf{x}) = f_j(\mathbf{x}^{i\to}) = 1$ for all $i \in C_f \setminus A$. So $\partial^{i\to} f_j(\mathbf{x}) = 0$ for all $i \notin A$. Therefore $\mathcal{I}_f(\mathbf{x})$ has the same j in-star as $\zeta\Delta$.

Similarly, if every arc into j is negative in $\zeta\Delta$, then let $a_i^j = -1$ for all $i \in A$. Let $a_k^j = |A| - 1$ for some $k \in C_f \setminus A$ and $a_i^j = 0$ for all other $i \in C_f \setminus (A \cup k)$. Observe that $\mathbf{a}^j \cdot \mathbf{x} = -1$ and $\mathbf{a}^j \cdot \mathbf{x}^{i\to} = 1$ for all $i \in A$. Also, $\mathbf{a}^j \cdot \mathbf{x}^{k\to} = 1 - 2|A| < 0$. So since $a_i^j < 0$ for all $i \in A$, $\partial^{i\to} f_j(\mathbf{x}) = -$ for all $i \in A$ by proposition 2.6.1. Also, $\partial^{i\to} f_j(\mathbf{x}) = 0$ for all $i \notin A$. Therefore $\mathcal{I}_f(\mathbf{x})$ has the same j in-star as $\zeta\Delta$ again.

So we have shown that $\mathcal{I}_f(\mathbf{x}) = \zeta\Delta$. Therefore $\Delta = \zeta\mathcal{I}_f(\mathbf{x}) = \mathcal{I}_{f^\zeta}(\zeta\mathbf{x})$ by theorem 2.3.1. So by lemma 2.6.4, f^ζ is also a threshold activation function. Since $\Delta = \mathcal{I}_{f^\zeta}(\zeta\mathbf{x})$, Δ is indeed a local interaction graph of a threshold activation function.

$\square$

Chapter 3

Conjectures of René Thomas

3.1 Introduction

In the context of gene regulatory networks, René Thomas introduced two general conjectures in [7]. The first was that positive feedback in the gene regulatory network is a necessary condition for multistationarity in the dynamics of gene expression. The second was that negative feedback in the gene regulatory network is a necessary condition for stable periodicity. Signed directed graphs are often used as representations of gene regulatory networks. In fact, the main intended use of interaction graphs is to model gene regulatory networks. We interpret positive and negative feedback correspond to positive and negative cycles within an interaction graph. What we currently lack is an interpretation of multistationarity and stable periodicity.

A new directed graph called the state transition graph will be introduced in section 3.2. It will represent how gene expression levels or the values of components of an activation function can change over time. Then we will be able to define the features within the state transition graph that correspond to multistationarity and stable periodicity and relate these features to feedback in the interaction graph.

Versions of René Thomas' conjectures have been proved by several authors in different contexts. The most relevant version for our discussion is Corollary 1 in [4]. Richard and Comet's result uses a less general definition of the state transition graph then we use here however. In subsection 3.3.1, I will present and prove a new version of Thomas' first conjecture using our more general definition of the state transition graph, but with a weaker conclusion. Then I show that that Richard and Comet's result follows from this version of the Thomas' first conjecture for a special class of activation functions that produce the same kind of state transition graphs used in [4].

In [5], Adrien Richard proved a version of the second conjecture. In subsection 3.3.2, I will present a new proof of Richard's theorem.

In [5] and [2], Adrien Richard and others ask a local version of René Thomas' second conjecture. Is there a Boolean activation function f with no steady states whose local interaction graphs contain no negative cycles? Paul Ruet gives an example of such an activation function in [6] that has twelve components. In section 3.4, I will give another example with only eight components.

3.1.1 State Transition Graphs

For an activation function f, we will imagine the elements of $\mathbb{S}_f$ as the states of a discrete dynamical system where f is telling us how the states of the system can change over time. There are a few different ways to interpret this. For our sake we are interested in a dynamical system where the values of the components in C_f can only change one at a time. This is called asynchronous dynamics. More specifically while the system is in state $\mathbf{x}$, the system tends towards $f(\mathbf{x})$. Since we only allow one component to change at a time, the next state in the system will be $\mathbf{x}^{i \to f_i(\mathbf{x})}$ for some $i \in C_f$ such that $f_i(\mathbf{x}) \neq x_i$.

Now we give a graphical way to represent this asynchronous dynamical system. Given an activation function f, the *state transition graph of* f, denoted $\mathcal{S}_f$ is a directed graph derived from f whose vertex set is $\mathbb{S}_f$. If $f_i(\mathbf{x}) \neq x_i$, then the arc $(\mathbf{x}, \mathbf{x}^{i \to f_i(\mathbf{x})}) \in \mathcal{S}_f$. In particular there are no arcs in $\mathcal{S}_f$ between states that differ in more than one coordinate and there are no loops.

For example, let the activation function $f(x_1, x_2) = (1 - x_2^2, x_2(x_1 - 1)/2)$ where $\mathbb{S}_f = \{\pm 1\} \times \{0, \pm 1\}$. A picture of $\mathcal{S}_f$ is given in figure 3.1.

The focus of this chapter is relating features of the state transition graph $\mathcal{S}_f$ with features in the interaction graphs $\mathcal{I}_f$ and $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

It turns out that the state transition graphs of conjugate activation functions are related in a nice way.

Theorem 3.1.1. *If $\lambda = (\phi, \zeta)$ is a switching isomorphism of an activation function f, then $\mathcal{S}_{f^\lambda} = \phi\mathcal{S}_f$, i.e., $\mathcal{S}_f$ and $\mathcal{S}_{f^\lambda}$ are isomorphic.*

Proof. Suppose the arc $(\mathbf{x}, \mathbf{x}^{i \to f_i(\mathbf{x})}) \in \mathcal{S}_f$, i.e., $f_i(\mathbf{x}) \neq x_i$. All we need to do is show that the arc $(\lambda\mathbf{x}, \lambda\mathbf{x}^{\phi i \to f^\lambda_{\phi i}(\lambda\mathbf{x})})$ is also in $\mathcal{S}_{f^\lambda}$. This is sufficient since λ is an invertible map. Since $\lambda : \mathbb{S}_f \to \mathbb{S}_{f^\lambda}$ is a bijection, $\lambda_{\phi i} f(\mathbf{x}) \neq \lambda_{\phi i}\mathbf{x}$ if $f_i(\mathbf{x}) \neq x_i$. But $\lambda_{\phi i} f(\mathbf{x}) = f^\lambda_{\phi i}(\lambda\mathbf{x})$ by equation (2.7), so $f^\lambda_{\phi i}(\lambda\mathbf{x}) \neq \lambda x_{\phi i}$. Therefore the arc $(\lambda\mathbf{x}, \lambda\mathbf{x}^{\phi i \to f^\lambda_{\phi i}(\lambda\mathbf{x})}) \in \mathcal{S}_f$. $\square$

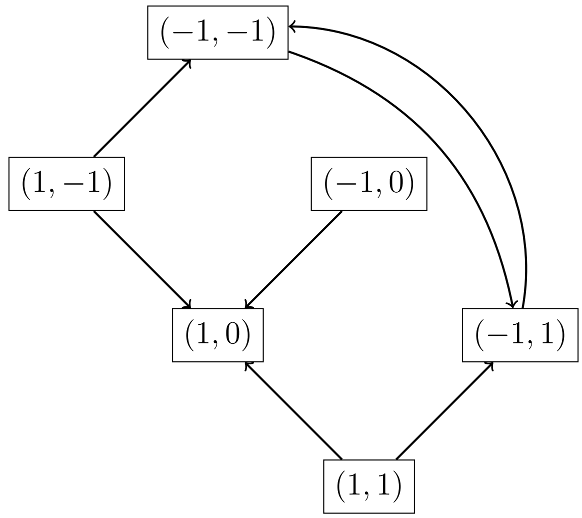

Figure 3.1: A state transition graph $\mathcal{S}_f$.

3.2 Attractors

Given a state transition graph $\mathcal{S}_f$. We call $\mathcal{A} \subseteq \mathcal{S}_f$ an *attractor of* $\mathcal{S}_f$ if $\mathcal{A}$ is a terminal strong component of $\mathcal{S}$. If $\mathcal{A}$ contains a single state, then we call $\mathcal{A}$ a *steady state attractor*, or just a steady state. Otherwise we call $\mathcal{A}$ a *periodic attractor*.

For example in figure 3.1, the attractors are $\{(1,0)\}$ and $\{(-1,-1),(-1,1)\}$. The first is a steady state and the second is a periodic attractor.

If we are interested in how the states of the dynamical system change over time, then attractors are important. If the system is in a state contained in an attractor, then all future states will also be in the attractor. Furthermore, from any state in $\mathbb{S}_f$ there is a path from that state into an attractor since attractors are terminal strong components of $\mathcal{S}_f$.

Proposition 3.2.1. *There is a dipath into an attractor from any vertex of* $\mathcal{S}_f$.

Proof. The proof is by induction on the number of strong components of $\mathcal{S}_f$.

If $\mathcal{S}_f$ contains a single connected component, then $\mathcal{S}_f$ is strongly connected itself. So the proposition is true in this case.

Now assume that $\mathcal{S}_f$ has $k > 1$ strong components and that the proposition is true for digraphs with fewer than k strong components. Let $\mathbf{x} \in \mathcal{S}_f$. If $\mathbf{x}$ is contained in an attractor, then we are done. So suppose that $\mathbf{x}$ is not contained in an attractor. By proposition 1.1.8, $\mathbf{x}$ is contained in some strong component $\mathcal{C}$ of $\mathcal{S}_f$. Since $\mathcal{C}$ is not an attractor, there is an arc out of $\mathcal{C}$ to another strong component $\mathcal{C}'$ of $\mathcal{S}_f$. Since the initial vertex of this arc is in $\mathcal{C}$, there is a dipath from $\mathbf{x}$ to $\mathbf{x}'$, the terminal vertex of the arc. Therefore there is a dipath from $\mathbf{x}$ to $\mathbf{x}'$.

Now consider $\mathcal{S}'$ the subgraph of $\mathcal{S}_f$ induced by $\mathbb{S}_f \setminus \mathcal{C}$. Any strong component of $\mathcal{S}_f$ other than $\mathcal{C}$ is also a strong component of $\mathcal{S}'$. To see this, let $\mathbf{y}, \mathbf{y}' \in \mathbb{S}_f \setminus \mathcal{C}$ be in the same strong component of $\mathbb{S}_f$. Then there is still a dipath from $\mathbf{y}$ to $\mathbf{y}'$ and a dipath from $\mathbf{y}'$ to $\mathbf{y}$ in $\mathcal{S}'$ since it is an induced subgraph.

This means that $\mathcal{S}'$ has $k-1$ strong components. So by induction, there is a dipath in $\mathcal{S}'$ from $\mathbf{x}'$ into an attractor $\mathcal{A} \subseteq \mathcal{S}'$. Combining this dipath with the dipath in $\mathcal{S}_f$ from $\mathbf{x}$ to $\mathbf{x}'$ yields a dipath from $\mathbf{x}$ to $\mathcal{A}$. All that is left to show is that $\mathcal{A}$ is also an attractor in $\mathcal{S}_f$.

As we showed earlier, every strong component of $\mathcal{S}_f$ other than $\mathcal{C}$ is also a strong component of $\mathcal{S}'$. Therefore $\mathcal{A}$ is a strong component in $\mathcal{S}_f$. Suppose that $\mathcal{A}$ is not an attractor in $\mathcal{S}_f$. Then there must be an arc from $\mathcal{A}$ into $\mathcal{C}$ since $\mathcal{A}$ is terminal in $\mathcal{S}'$. But then there would be a dipath from $\mathcal{A}$ to $\mathcal{C}$ and a dipath from $\mathcal{C}$ to $\mathcal{A}$. This means that $\mathcal{A}$ and $\mathcal{C}$ are in the same strong component of $\mathcal{S}_f$ which is not the case. Therefore there are no arcs out of $\mathcal{A}$ in $\mathcal{S}_f$ either. Therefore $\mathcal{A}$ is an attractor in $\mathcal{S}_f$.

□

It turns out that some structure of the attractors can be related to cycles in the interaction graphs.

3.2.1 Thomas' Conjectures

A gene regulatory network is a graphical representation of a system of interacting genes. Signed digraphs are often used as a model of gene regulatory networks. A positive arc from i to j in the network indicates that changing the expression level of gene i can cause the expression level of gene j to change in the same way. A negative arc from i to j in the network indicates that changing the expression level of gene i can cause a contrary change in the expression level of gene j. Interaction graphs and state transition graphs are used as a model gene regulatory networks and the dynamics of gene expression respectively. The activation functions gives us a more concrete way to think about how gene expression levels change over time since we can think of a dynamical system in state $\mathbf{x}$ tending towards state $f(\mathbf{x})$.

In the context of gene regulatory networks, the geneticist René Thomas made two conjectures.

Conjecture 3.2.2 (Conjectures of René Thomas)**.**

1. *Positive feedback is a necessary condition for multi-stablility.*

2. *Negative feedback is a necessary condition for stable periodicity.*

We interpret feedback to mean cycles within the interaction graph of an activation function f. Multi-stability we interpret to mean the presence of multiple attractors in the state transition graph of f. Stable periodicity we interpret to mean the presence of periodic attractors contained in the state transition graph of f.

3.3 Proofs of the Conjectures using Conjugate Activation Functions

For the proofs of our versions of Thomas' conjectures, we will need restricted versions of a given activation function. If f is an activation function and the states $\mathbf{s}, \mathbf{t} \in \mathbb{S}_f$, let

$$[\mathbf{s}, \mathbf{t}] = \{\mathbf{x} \in \mathbb{S}_f \mid \min(s_i, t_i) \leq x_i \leq \max(s_i, t_i)\}.$$

Also, $[s_i, t_i] = \{a \in S_i \mid \min(s_i, t_i) \leq a \leq \max(s_i, t_i)\}$ for and $i \in C_f$. Define $[\mathbf{s}, \mathbf{t}]|f$ to be the activation function with the same components as f whose state space is $[\mathbf{s}, \mathbf{t}]$, where

$$[\mathbf{s}, \mathbf{t}]|f_i(\mathbf{x}) = \begin{cases} f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ \max(s_i, t_i) & \text{if } f_i(\mathbf{x}) > \max(s_i, t_i) \\ \min(s_i, t_i) & \text{if } f_i(\mathbf{x}) < \min(s_i, t_i). \end{cases} \tag{3.1}$$

How are the interaction graphs of $[\mathbf{s}, \mathbf{t}]|f$ and f related?

Lemma 3.3.1. *Given an activation function f. Let $g = [\mathbf{s}, \mathbf{t}]|f$. Then*

1. *$\mathcal{I}_g(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$*
2. *$\mathcal{I}^*_g(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$.*

Proof. For the first item, all we need to show is that if $(i, j, \sigma_{ij}) \in \mathcal{I}_g(\mathbf{x}, \mathbf{y})$, then $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

If $(i, j, \sigma_{ij}) \in \mathcal{I}_g(\mathbf{x}, \mathbf{y})$, then $\sigma_{ij} = \partial^{i \to y_i} g_j(\mathbf{x}) = \operatorname{sgn}[g_j(\mathbf{x}^{i \to y_i}) - g_j(\mathbf{x})][y_i - x_i]$. Also, $\partial^{i \to y_i} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})][y_i - x_i]$. So if we want to see that $\partial^{i \to y_i} f_j(\mathbf{x}) = \partial^{i \to y_i} g_j(\mathbf{x})$, we need only compare the quantities $f_j(\mathbf{x}^{i \to y_i}) - f_j(\mathbf{x})$ and $g_j(\mathbf{x}^{i \to y_i}) - g_j(\mathbf{x})$.

Since $\sigma_{ij} \neq 0$, $g_j(\mathbf{x}^{i \to y_i}) \neq g_j(\mathbf{x})$. Suppose $g_j(\mathbf{x}) < g_j(\mathbf{x}^{i \to y_i})$. Since the maximum value of g_j is $\max(x_j, y_j)$, $g_j(\mathbf{x}^{i \to y_i}) \leq \max(x_j, y_j)$. So $g_j(\mathbf{x}) < \max(x_j, y_j)$. By equation (3.1), either $f_j(\mathbf{x}) = g_j(\mathbf{x})$, or $g_j(\mathbf{x}) = \min(x_j, y_j) > f_j(\mathbf{x})$. In either case $g_j(\mathbf{x}) \geq f_j(\mathbf{x})$.

Similarly, $g(\mathbf{x}^{i \to y_i}) = f(\mathbf{x}^{i \to y_i})$ or $g(\mathbf{x}^{i \to y_i}) = \max(x_i, y_i) < f(\mathbf{x}^{i \to y_i})$ by equation (3.1). So $g(\mathbf{x}^{i \to y_i}) \leq f_j(\mathbf{x}^{i \to y_i})$. Therefore $f_j(\mathbf{x}^{i \to y_i}) \geq g_j(\mathbf{x}^{i \to y_i}) > g_j(\mathbf{x}) \geq f_j(\mathbf{x})$, so $\operatorname{sgn}[f_j(\mathbf{x}^{i \to y_i}) -$

$f_j(\mathbf{x})] = \operatorname{sgn}[g_j(\mathbf{x}^{i\to y_i}) - g_j(\mathbf{x})]$. Therefore $\partial^{i\to y_i} f_j(\mathbf{x}) = \partial^{i\to y_i} g_j(\mathbf{x}) = \sigma_{ij}$, so $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{t})$.

If instead $g_j(\mathbf{x}^{i\to y_i}) < g_j(\mathbf{x})$, the same argument with all inequalities reversed shows that $\partial^{i\to y_i} f_j(\mathbf{x}) = \partial^{i\to y_i} g_j(\mathbf{x}) = \sigma_{ij}$ again. So in either case, $(i, j, \sigma_{ij}) \in \mathcal{I}_f(\mathbf{x}, \mathbf{y})$.

For the second item, $\mathcal{I}^*_g(\mathbf{x}, \mathbf{y}) = \mathcal{I}_g(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})$ and $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y}) = \mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})$ by proposition 2.2.3. By the first item of this lemma, $\mathcal{I}_g(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})$ is a subgraph of $\mathcal{I}_f(\mathbf{x}, \mathbf{x}^{\to^*\mathbf{y}})$. Therefore $\mathcal{I}^*_g(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$. □

Conjugating a restricted activation function behaves as you would expect.

Lemma 3.3.2. *Given $\lambda = (\phi, \zeta)$, a switching isomorphism of an activation function f and $\mathbf{s}, \mathbf{t} \in \mathbb{S}_f$. If $g = [\mathbf{s}, \mathbf{t}]|f$, then $g^\lambda = [\lambda\mathbf{s}, \lambda\mathbf{t}]|f^\lambda$.*

Proof. A straightforward, if tedious calculation is all that is needed to prove this. We will show the functions are equal component-wise. Let us start with the right hand side of the equation $g^\lambda = [\lambda\mathbf{s}, \lambda\mathbf{t}]|f^\lambda$. By equation (3.1),

$$[\lambda\mathbf{x}, \lambda\mathbf{y}]|f^\lambda_{\phi i}(\lambda\mathbf{x}) = \begin{cases} f^\lambda_{\phi i}(\lambda\mathbf{x}) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) \in [\lambda_{\phi i}\mathbf{s}, \lambda_{\phi i}\mathbf{t}] \\ \max(\lambda_{\phi i}\mathbf{s}, \lambda_{\phi i}\mathbf{t}) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) > \max(\lambda_{\phi i}\mathbf{s}, \lambda_{\phi i}\mathbf{t}) \\ \min(\lambda_{\phi i}\mathbf{s}, \lambda_{\phi i}\mathbf{t}) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) < \min(\lambda_{\phi i}\mathbf{s}, \lambda_{\phi i}\mathbf{t}). \end{cases}$$

Using equation (2.4),

$$[\lambda\mathbf{x}, \lambda\mathbf{y}]|f^\lambda_{\phi i}(\lambda\mathbf{x}) = \begin{cases} f^\lambda_{\phi i}(\lambda\mathbf{x}) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) \in [\zeta_i s_i, \zeta_i t_i] \\ \max(\zeta_i s_i, \zeta_i t_i) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) > \max(\zeta_i s_i, \zeta_i t_i) \\ \min(\zeta_i s_i, \zeta_i t_i) & \text{if } f^\lambda_{\phi i}(\lambda\mathbf{x}) < \min(\zeta_i s_i, \zeta_i t_i). \end{cases}$$

And by equation (2.7),

$$[\lambda\mathbf{x}, \lambda\mathbf{y}]|f^\lambda_{\phi i}(\lambda\mathbf{x}) = \begin{cases} \zeta_i f_i(\mathbf{x}) & \text{if } \zeta_i f_i(\mathbf{x}) \in [\zeta_i s_i, \zeta_i t_i] \\ \max(\zeta_i s_i, \zeta_i t_i) & \text{if } \zeta_i f_i(\mathbf{x}) > \max(\zeta_i s_i, \zeta_i t_i) \\ \min(\zeta_i s_i, \zeta_i t_i) & \text{if } \zeta_i f_i(\mathbf{x}) < \min(\zeta_i s_i, \zeta_i t_i). \end{cases}$$

For the left hand side of $g^\lambda = [\lambda\mathbf{s}, \lambda\mathbf{t}]|f^\lambda$, $g^\lambda_{\phi i}(\lambda\mathbf{x}) = \zeta_i g_i(\mathbf{x})$ by equation (2.7). Using equation (3.1),

$$\zeta_i g_i(\mathbf{x}) = \begin{cases} \zeta_i f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ \zeta_i \max(s_i, t_i) & \text{if } f_i(\mathbf{x}) > \max(s_i, t_i) \\ \zeta_i \min(s_i, t_i) & \text{if } f_i(\mathbf{x}) < \min(s_i, t_i). \end{cases}$$

If $\zeta_i = +$, then

$$
\begin{aligned}
[\lambda\mathbf{x}, \lambda\mathbf{y}]|f^{\lambda}_{\phi i}(\lambda\mathbf{x}) &= \begin{cases} \zeta_i f_i(\mathbf{x}) & \text{if } \zeta_i f_i(\mathbf{x}) \in [\zeta_i s_i, \zeta_i t_i] \\ \max(\zeta_i s_i, \zeta_i t_i) & \text{if } \zeta_i f_i(\mathbf{x}) > \max(\zeta_i s_i, \zeta_i t_i) \\ \min(\zeta_i s_i, \zeta_i t_i) & \text{if } \zeta_i f_i(\mathbf{x}) < \min(\zeta_i s_i, \zeta_i t_i). \end{cases} \\
&= \begin{cases} f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ \max(s_i, t_i) & \text{if } f_i(\mathbf{x}) > \max(s_i, t_i) \\ \min(s_i, t_i) & \text{if } f_i(\mathbf{x}) < \min(s_i, t_i). \end{cases} \\
&= g_i(\mathbf{x}) \\
&= \zeta_i g_i(\mathbf{x}) \\
&= g^{\lambda}_{\phi i}(\lambda\mathbf{x}).
\end{aligned}
$$

If $\zeta_i = -$, then

$$
\begin{aligned}
[\lambda\mathbf{x}, \lambda\mathbf{y}]|f^{\lambda}_{\phi i}(\lambda\mathbf{x}) &= \begin{cases} -f_i(\mathbf{x}) & \text{if } -f_i(\mathbf{x}) \in [-s_i, -t_i] \\ \max(-s_i, -t_i) & \text{if } -f_i(\mathbf{x}) > \max(-s_i, -t_i) \\ \min(-s_i, -t_i) & \text{if } -f_i(\mathbf{x}) < \min(-s_i, -t_i) \end{cases} \\
&= \begin{cases} -f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ \max(-s_i, -t_i) & \text{if } -f_i(\mathbf{x}) > \max(-s_i, -t_i) \\ \min(-s_i, -t_i) & \text{if } -f_i(\mathbf{x}) < \min(-s_i, -t_i) \end{cases} \\
&= \begin{cases} -f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ -\min(s_i, t_i) & \text{if } -f_i(\mathbf{x}) > -\min(s_i, t_i) \\ -\max(s_i, t_i) & \text{if } -f_i(\mathbf{x}) < -\max(s_i, t_i) \end{cases} \\
&= \begin{cases} -f_i(\mathbf{x}) & \text{if } f_i(\mathbf{x}) \in [s_i, t_i] \\ -\min(s_i, t_i) & \text{if } f_i(\mathbf{x}) < \min(s_i, t_i) \\ -\max(s_i, t_i) & \text{if } f_i(\mathbf{x}) > \max(s_i, t_i) \end{cases} \\
&= -g_i(\mathbf{x}) \\
&= \zeta_i g_i(\mathbf{x}) \\
&= g^{\lambda}_{\phi i}(\lambda\mathbf{x}).
\end{aligned}
$$

In either case, we have shown that $g^\lambda = [\lambda\mathbf{s}, \lambda\mathbf{t}]|f^\lambda$. $\square$

3.3.1 The First Conjecture

Now for the theorem that will lead to a new version of Thomas' first conjecture.

Theorem 3.3.3. *Let f be an activation function. Suppose there are distinct states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathbf{x}$ is a steady state of $g = [\mathbf{x}, \mathbf{y}]|f$ and $g(\mathbf{x}^{i \to^* y_i}) \neq \mathbf{x}$ for all $i \in W$ where $W = \{i \in C_g \mid x_i \neq y_i\}$. Then $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ contains a positive cycle.*

Proof. First we will prove a simpler version of the theorem.

Lemma 3.3.4. *Let f be an activation function. Suppose that $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$, $\mathbf{x} < \mathbf{y}$, $\mathbf{x}$ is a steady state of $g = [\mathbf{x}, \mathbf{y}]|f$ and $g(\mathbf{x}^{i \to^* y_i}) > \mathbf{x}$ for all $i \in W$ where $W = \{i \in C_g \mid x_i \neq y_i\}$. Then $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ contains a positive cycle whose vertices are in W.*

Proof. For $i \in W$, we will first show that there is a positive arc out of i to some other vertex in W.

Since $g(\mathbf{x}^{i \to^* y_i}) > \mathbf{x}$, there is some $j \in C_g$ such that $g_j(\mathbf{x}^{i \to^* y_i}) > x_j$. If $j \notin W$, then $x_j = y_j$. So $g_j(\mathbf{x}^{i \to^* y_i}) = g_j(\mathbf{x}) = x_j$ since $\mathbf{x}$ is a steady state of g. Therefore $j \in W$.

Since $y_i > x_i$,

$$\partial^{i \to^* y_i} g_j(\mathbf{x}) = \operatorname{sgn}[g_j(\mathbf{x}^{i \to^* y_i}) - g_j(\mathbf{x})][y_i - x_i] = \operatorname{sgn}[g_j(\mathbf{x}^{i \to^* y_i}) - g_j(\mathbf{x})].$$

Since $\mathbf{x}$ is a steady state of g, $\partial^{i \to^* y_i} g_j(\mathbf{x}) = \operatorname{sgn}[g_j(\mathbf{x}^{i \to^* y_i}) - x_j]$. And finally, since $g_j(\mathbf{x}^{i \to^* y_i}) > x_j$, $\partial^{i \to^* y_i} g_j(\mathbf{x}) = +$. Therefore, there is a positive arc from i to j in $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$.

Finally, $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ by lemma 3.3.1. Therefore $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ contains a positive arc from i to j since $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$ does.

Finally, Pick a vertex $i_1 \in W$. There is a positive arc out of i_1 to a new vertex i_2 in W. Since $i_2 \in W$, there is a positive arc from i_2 to $i_3 \in W$, and so on. Repeating this creates a sequence of vertices. Since W is finite, there is a first repeated vertex i_k in this sequence. Following the vertices from the first occurrence of i_k in the sequence until it repeats closes a positive cycle in $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$. $\square$

Back to the proof of theorem 3.3.3. The idea is to apply lemma 3.3.4 to f^ζ, a conjugate activation function of f. Since $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ and $\mathcal{I}_{f^\zeta}^*(\zeta\mathbf{x}, \zeta\mathbf{y})$ are switching isomorphic by theorem 2.3.1, $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ will contain a positive cycle because $\mathcal{I}_{f^\zeta}^*(\zeta\mathbf{x}, \zeta\mathbf{y})$ contains a positive cycle by lemma 3.3.4.

Let $\zeta \in \{\pm\}^{C_f}$ such that $\zeta_i = +$ if $x_i \leq y_i$ and $\zeta_i = -$ otherwise. We have to show that f^ζ together with the states $\zeta\mathbf{x}$ and $\zeta\mathbf{y}$ satisfy all three hypotheses of lemma 3.3.4.

By the definition of ζ, $\zeta_i x_i \leq \zeta_i y_i$ for all $i \in C_f$. Therefore $\zeta\mathbf{x} < \zeta\mathbf{y}$ since $\mathbf{x} \neq \mathbf{y}$. So the first hypothesis of lemma 3.3.4 is satisfied.

For the next hypothesis, we have to show that $\zeta\mathbf{x}$ is a steady state of $[\zeta\mathbf{x}, \zeta\mathbf{y}]|f^\zeta$. By lemma 3.3.2, $[\zeta\mathbf{x}, \zeta\mathbf{y}]|f^\zeta = g^\zeta$. By equation 2.6, $g^\zeta(\zeta\mathbf{x}) = \zeta g(\mathbf{x})$. Since $\mathbf{x}$ is a steady state of g, $g^\zeta(\zeta\mathbf{x}) = \zeta\mathbf{x}$. Therefore $\zeta\mathbf{x}$ is a steady state of g^ζ.

Finally, we have to show that $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) > \zeta\mathbf{x}$ for all $i \in W$. We know that $g(\mathbf{x}^{i \to^* y_i}) \neq \mathbf{x}$ for all $i \in W$ by assumption. So $\zeta\mathbf{x} \neq \zeta g(\mathbf{x}^{i \to^* y_i})$. Using equation (2.3), $\zeta g(\mathbf{x}^{i \to^* y_i}) = \zeta g(\mathbf{x}^{i \to z_i})$ where $\mathbf{z} = \mathbf{x}^{\to^* \mathbf{y}}$. Using equation (2.6), $\zeta g(\mathbf{x}^{i \to z_i}) = g^\zeta(\zeta[\mathbf{x}^{i \to z_i}])$. Then $g^\zeta(\zeta[\mathbf{x}^{i \to z_i}]) = g^\zeta([\zeta\mathbf{x}]^{i \to \zeta_i z_i})$ by lemma 2.3.2. Since $\zeta\mathbf{z} = \mathbf{z}'$ where $\mathbf{z}' = [\zeta\mathbf{x}]^{\to^* \zeta\mathbf{y}}$ by lemma 2.3.3, $z'_i = \zeta_i z_i$. So $g^\zeta([\zeta\mathbf{x}]^{i \to \zeta_i z_i}) = g^\zeta([\zeta\mathbf{x}]^{i \to z'_i}) = g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i})$ by equation (2.3). Therefore $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) = \zeta g(\mathbf{x}^{i \to^* y_i})$. Since $\zeta\mathbf{x} \neq \zeta g(\mathbf{x}^{i \to^* y_i})$, $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) \neq \zeta\mathbf{x}$. Now by the definition of ζ, $\zeta\mathbf{x}$ is the minimum state in $[\zeta\mathbf{x}, \zeta\mathbf{y}]$. Therefore $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) > \zeta\mathbf{x}$ since $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) \neq \zeta\mathbf{x}$, and $g^\zeta([\zeta\mathbf{x}]^{i \to^* \zeta_i y_i}) \in [\zeta\mathbf{x}, \zeta\mathbf{y}]$.

We can now apply lemma 3.3.4 to f^ζ and the states $\zeta\mathbf{x}$ and $\zeta\mathbf{y}$. Therefore $\mathcal{I}^*_{f^\zeta}(\zeta\mathbf{x}, \zeta\mathbf{y})$ contains a positive cycle. Since $\mathcal{I}^*_{f^\zeta}(\zeta\mathbf{x}, \zeta, \mathbf{y})$ and $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$ are switching isomorphic by theorem 2.3.1, $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$ must also contain a positive cycle. □

With this theorem, we can now prove a new version of Thomas' first conjecture.

Theorem 3.3.5 (Thomas' First conjecture)**.** *If an activation function f has at least two steady states, then there exist states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}^*_f(\mathbf{x}, \mathbf{y})$ contains a positive cycle.*

Proof. If $\mathbf{z}$ and $\mathbf{y}$ are distinct steady states of f, then $\mathbf{z}$ and $\mathbf{y}$ are also steady states of $[\mathbf{z}, \mathbf{y}]|f$. So pairs of distinct states $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x}$ is a steady state of $[\mathbf{x}, \mathbf{y}]|f$ certainly exist.

Among all such pairs, there is a pair of states that are closest. So let us assume that among all pairs of distinct states $(\mathbf{x}, \mathbf{y})$ such that $\mathbf{x}$ is a steady state of $g = [\mathbf{x}, \mathbf{y}]|f$, $|\mathbf{x}, \mathbf{y}|$ is minimum. Let $W = \{i \in C_g \mid x_i \neq y_i\}$. It turns out that with these assumptions, $g(\mathbf{x}^{i \to^* y_i}) \neq \mathbf{x}$ for all $i \in W$. So we can apply theorem 3.3.3 to get our result.

Let $\mathbf{z} = \mathbf{x}^{i \to^* y_i}$ and $g' = [\mathbf{z}, \mathbf{y}]|f$. It is possible that $\mathbf{z} = \mathbf{y}$. In this case, $g(\mathbf{x}^{i \to^* y_i}) = g(\mathbf{y}) = \mathbf{y}$ since $\mathbf{y}$ is a steady state of g. So $g(\mathbf{x}^{i \to^* y_i}) \neq \mathbf{x}$ for all $i \in W$ since $\mathbf{x}$ and $\mathbf{y}$ only differ in the ith component. So theorem 3.3.3 applies in this case.

Now suppose that $\mathbf{z} \neq \mathbf{y}$, i.e., $\mathbf{z}$ and $\mathbf{y}$ are distinct. If $x_i < y_i$, then by equation (3.1), $f(\mathbf{x}^{i \to^* y_i}) \leq x_i$ since $g_i(\mathbf{z}) = x_i$. So $f_i(\mathbf{z}) < z_i$. Similarly if $x_i > y_i$, then $f_i(\mathbf{z}) > x_i$. In either case, $g_i'(\mathbf{z}) = z_i$ by equation (3.1). For $j \in W$ different from i, again using equation (3.1),

$$
\begin{aligned}
g_j'(\mathbf{z}) &= \begin{cases} f_j(\mathbf{z}) & \text{if } f_j(\mathbf{z}) \in [z_j, y_j] \\ \max(z_j, y_j) & \text{if } f_j(\mathbf{z}) > \max(z_j, y_j) \\ \min(z_j, y_j) & \text{if } f_j(\mathbf{z}) < \min(z_j, y_j) \end{cases} \\
&= \begin{cases} f_j(\mathbf{z}) & \text{if } f_j(\mathbf{z}) \in [x_j, y_j] \\ \max(x_j, y_j) & \text{if } f_j(\mathbf{z}) > \max(x_j, y_j) \\ \min(x_j, y_j) & \text{if } f_j(\mathbf{z}) < \min(x_j, y_j) \end{cases} \\
&= g_j(\mathbf{z}) \\
&= x_j.
\end{aligned}
$$

Therefore $\mathbf{z}$ is a steady state of g'. However, $|\mathbf{z}, \mathbf{y}| = |\mathbf{x}^{i \to^* y_i}, \mathbf{y}| < |\mathbf{x}, \mathbf{y}|$. This contradicts our assumption that among all distinct pairs of states $(\mathbf{x}, \mathbf{y})$ such that $\mathbf{x}$ is a steady state of $g = [\mathbf{x}, \mathbf{y}]|f$, $|\mathbf{x}, \mathbf{y}|$ is minimal. Therefore $g(\mathbf{x}^{i \to^* y_i}) \neq \mathbf{x}$ for all $i \in W$. Now we can apply theorem 3.3.3 to f, $\mathbf{x}$ and $\mathbf{y}$ to get our result. □

Call an activation function f *unitary* if for all states $\mathbf{x} \in \mathbb{S}_f$ and $i \in C_f$, either $f_i(\mathbf{x}) \gtrdot x_i$, $f_i(\mathbf{x}) \lessdot x_i$ or $f_i(\mathbf{x}) = x_i$. So if f is unitary and $(\mathbf{x}, \mathbf{x}^{i \to a}) \in \mathcal{S}_f$, then $a \gtrdot x_i$, or $a \lessdot x_i$ since $a = f_i(\mathbf{x})$ and $f_i(\mathbf{x}) \gtrdot x_i$ or $f_i(\mathbf{x}) \lessdot x_i$. So the state transition graphs of unitary activation functions are the same as those used in [4]. If we restrict to unitary activation functions, then corollary 1 in [4] follows from corollary 3.3.5.

Corollary 3.3.6. *Given a unitary activation function f. If $\mathcal{S}_f$ contains at least two attractors, then there are states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ contains a positive cycle.*

Proof. Let $\mathcal{A}$ and $\mathcal{B}$ be attractors in $\mathcal{S}_f$. Let $\mathbf{s} \in \mathcal{A}$ and $\mathbf{t} \in \mathcal{B}$ be as close as possible, i.e., $|\mathbf{s}, \mathbf{t}|$ is minimum. Let $g = [\mathbf{s}, \mathbf{t}]|f$. I claim that $\mathbf{s}$ and $\mathbf{t}$ are steady states of g. If this is true, then we can apply corollary 3.3.5 to g. Our result will then follow since $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ by lemma 3.3.1.

First, we will show that $\mathbf{s}$ is a steady state of g. The situation of $\mathbf{s}$ and $\mathbf{t}$ is symmetric, so the same argument shows $\mathbf{t}$ is also a steady state of g.

Suppose that $g_i(\mathbf{s}) \neq s_i$ for some $i \in C_g$. This means that $s_i \neq t_i$ since $g_i(\mathbf{s}) = s_i$ by equation (3.1) otherwise. Now $g_i(\mathbf{s}) \in [s_i, t_i]$ by equation (3.1) also. Suppose that $s_i < t_i$. Since f is unitary, $f_i(\mathbf{s}) \gtrdot s_i$ since $g_i(\mathbf{s}) = s_i$ otherwise. Similarly, if $s_i > t_i$, then $f_i(\mathbf{s}) \lessdot s_i$. In either

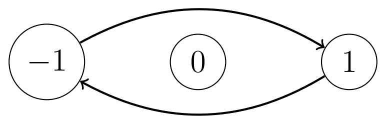

Figure 3.2: $\mathcal{S}_{-x}$ has two attractors, but $\mathcal{I}_{-x}$ contains no positive cycle.

case, $f_i(\mathbf{s}) \in [s_i, t_i]$. Therefore $f_i(\mathbf{s}) = g_i(\mathbf{s})$ by equation (3.1). Since $g_i(\mathbf{s}) \neq s_i$, $f_i(\mathbf{s}) \neq s_i$ also. Therefore the arc $(\mathbf{s}, \mathbf{s}^{i \to f_i(\mathbf{s})}) \in \mathcal{S}_f$. So the state $\mathbf{s}^{i \to f_i(\mathbf{s})} \in \mathcal{A}$. But $|\mathbf{s}^{i \to f_i(\mathbf{s})}, \mathbf{t}| < |\mathbf{s}, \mathbf{t}|$, a contradiction since $\mathbf{s}$ and $\mathbf{t}$ were chosen to be closest. Therefore $g_i(\mathbf{s}) = s_i$ for all $i \in C_g$, i.e., $\mathbf{s}$ is a steady state of g.

By symmetry, the same argument shows that $\mathbf{t}$ is a steady state of g. So g has at least two steady states. Therefore, there exist states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_g$ such that $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$ contains a positive cycle by corollary 3.3.5. Finally, $\mathcal{I}_g^*(\mathbf{x}, \mathbf{y})$ is a subgraph of $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ by lemma 3.3.1, so $\mathcal{I}_f^*(\mathbf{x}, \mathbf{y})$ also contains a positive cycle. □

You may ask whether unitary activation functions are required for corollary 3.3.6 to be true. It turns out that this version of Thomas' first conjecture is not true in general. To see this, let $f(x) = -x$ where $\mathbb{S}_f = \{0, \pm 1\}$. Then $\mathcal{S}_f$ contains the two attractors $\{-1, 1\}$ and $\{0\}$. This can be seen in figure 3.2. But

$$\partial^{i \to a} f(x) = \operatorname{sgn}[f(a) - f(x)][a - x] = \operatorname{sgn}[-a + x][a - x] = -\operatorname{sgn}[a - x]^2 = -.$$

So $\mathcal{I}_f$ consists of a single vertex and a single negative loop. So no local interaction graph of f can contain a positive cycle. However, this does not contradict corollary 3.3.5 since $\mathcal{S}_f$ does not contain multiple steady states.

3.3.2 The Second Conjecture

Theorem 1 in [5] proves that a negative cycle in the interaction graph $\mathcal{I}_f$ is a necessary condition for $\mathcal{S}_f$ to contain a periodic attractor. What follows is my own new proof of the theorem using techniques I have developed.

Theorem 3.3.7. *If $\mathcal{S}_f$ contains a periodic attractor, then $\mathcal{I}_f$ contains a negative cycle.*

Proof. I will prove the contrapositive statement.

Lemma 3.3.8. *If $\mathcal{I}_f$ is cycle-balanced, then every attractor in $\mathcal{S}_f$ is a steady state.*

Proof. We will prove several weaker intermediate versions of the lemma on our way to the final version. We begin with a result about positive interaction graphs.

Lemma 3.3.9. *Given an activation function f such that every arc in $\mathcal{I}_f$ is positive. If $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ and $\mathbf{x} \leq \mathbf{y}$, then $f(\mathbf{x}) \leq f(\mathbf{y})$, i.e., f is isotone.*

Proof. We proceed by induction on $|\mathbf{x}, \mathbf{y}|$.

If $|\mathbf{x}, \mathbf{y}| = 1$ then $\mathbf{y} = \mathbf{x}^{i \to a}$ for some $i \in C_f$. Since every arc in $\mathcal{I}_f$ is positive, $\partial^{i \to a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to a}) - f_j(\mathbf{x})][a - x_i] \geq 0$ for any $j \in C_f$. Now $\mathbf{y} > \mathbf{x}$, so $a > x_i$. Therefore $\partial^{i \to a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{i \to a}) - f_j(\mathbf{x})]$, so $f_j(\mathbf{x}^{i \to a}) \geq f_j(\mathbf{x})$. This is true for all $j \in C_f$, so $f(\mathbf{x}^{i \to a}) \geq f(\mathbf{x})$ i.e., $f(\mathbf{y}) \geq f(\mathbf{x})$.

Now suppose $|\mathbf{x}, \mathbf{y}| = k$ and that the lemma is true when the distance is less than k. Say $\mathbf{x}$ and $\mathbf{y}$ differ in the ith coordinate. Since $\mathbf{y} > \mathbf{x}$, $y_i > x_i$. But $|\mathbf{y}, \mathbf{x}^{i \to y_i}| = k - 1$ since $|\mathbf{x}, \mathbf{y}| = k$. By the induction hypothesis, $f(\mathbf{y}) \geq f(\mathbf{x}^{i \to y_i})$. Since $f(\mathbf{x}^{i \to y_i}) \geq f(\mathbf{x})$ by the first step of the induction, it follows that $f(\mathbf{y}) \geq f(\mathbf{x})$. $\square$

We are now ready to prove the weakest version of lemma 3.3.8.

Lemma 3.3.10. *If every arc in $\mathcal{I}_f$ is positive, then every attractor in $\mathcal{S}_f$ is a steady state.*

Proof. Let $\mathcal{A}$ be an attractor in $\mathcal{S}_f$ and let $\mathbf{x}$ be a maximal state in $\mathcal{A}$. I claim that $f(\mathbf{x}) \leq \mathbf{x}$. To see this, suppose that $f_j(\mathbf{x}) > x_j$ for some $j \in C_f$. So $(\mathbf{x}, \mathbf{x}^{j \to f_j(\mathbf{x})}) \in \mathcal{S}_f$. This means that $\mathbf{x}^{j \to f_j(\mathbf{x})} \in \mathcal{A}$. But $\mathbf{x}^{j \to f_j(\mathbf{x})} > \mathbf{x}$. Since $\mathbf{x}$ is a maximal state in $\mathcal{A}$, $\mathbf{x}^{j \to f_j(\mathbf{x})}$ cannot be contained in $\mathcal{A}$. Therefore $f_j(\mathbf{x}) \leq x_j$. Since this is true for each component of f, it follows that $f(\mathbf{x}) \leq \mathbf{x}$.

Now suppose that $\mathbf{x}$ is not a steady state. So $f(\mathbf{x}) < \mathbf{x}$. In particular, $f_i(\mathbf{x}) < x_i$ for some $i \in C_f$. This means that there is an arc $(\mathbf{x}, \mathbf{x}') \in \mathcal{S}_f$ where $\mathbf{x}' = \mathbf{x}^{i \to f_i(\mathbf{x})}$ and $\mathbf{x}' \in \mathcal{A}$.

I claim that there is no directed path from $\mathbf{x}'$ back to $\mathbf{x}$, a contradiction since $\mathbf{x} \in \mathcal{A}$. To prove the claim, I will show that if $\mathbf{y}$ is any state such that there is a directed path from $\mathbf{x}'$ to $\mathbf{y}$ in $\mathcal{S}_f$, then $y_i < x_i$.

Let $\mathbf{y}$ and $\mathbf{z}$ be consecutive states in a dipath in $\mathcal{S}_f$ that starts at $\mathbf{x}'$ where $\mathbf{y} < \mathbf{x}$ and $y_i < x_i$ in particular. I claim that this implies that $\mathbf{z} < \mathbf{x}$ and $z_i < x_i$ also. Since $\mathbf{y} < \mathbf{x}$, $f(\mathbf{y}) \leq f(\mathbf{x})$ by lemma 3.3.9. This means $f_i(\mathbf{y}) \leq f_i(\mathbf{x}) < x_i$.

The arc $(\mathbf{y}, \mathbf{z}) \in \mathcal{S}_f$, so $\mathbf{z} = \mathbf{y}^{j \to f_j(\mathbf{y})}$ for some $j \in C_f$. If $j = i$, then $z_i = f_i(\mathbf{y}) < x_i$. If $j \neq i$, then $z_i = y_i < x_i$. Either way $z_i < x_i$. Also, since $\mathbf{y} < \mathbf{x}$ and $f_j(\mathbf{y}) \leq f_j(\mathbf{x}) \leq x_j$, $\mathbf{z} = \mathbf{y}^{j \to f_j(\mathbf{y})} \leq \mathbf{x}$. So $\mathbf{z} < \mathbf{x}$ and $z_i < x_i$ in particular, establishing the claim.

Now for the proof of this lemma, we are assuming that $\mathbf{x}' < \mathbf{x}$ and $x'_i < x_i$. So for any state $\mathbf{y}$ such that there is a directed path from $\mathbf{x}'$ to $\mathbf{y}$ in $\mathcal{S}_f$, it follows by the argument above that $y_i < x_i$ since this property holds for consecutive states in such a dipath. Therefore there is no directed path from $\mathbf{x}'$ to $\mathbf{x}$. This cannot happen since $\mathbf{x} \in \mathcal{A}$. Therefore $f(\mathbf{x}) = \mathbf{x}$, so $\mathcal{A} = \{\mathbf{x}\}$ is a steady state. $\square$

This lemma can be immediately improved.

Lemma 3.3.11. *If $\mathcal{I}_f$ is balanced, then every attractor in $\mathcal{S}_f$ is a steady state.*

Proof. By theorem 1.3.2, there is ζ such that every arc in $\zeta\mathcal{I}_f$ is positive. By corollary 2.3.1, $\zeta\mathcal{I}_f = \mathcal{I}_{f\zeta}$. Therefore every attractor in $\mathcal{S}_{f\zeta}$ is a steady state by lemma 3.3.10. By theorem 3.1.1, $\mathcal{S}_{f\zeta}$ and $\mathcal{S}_f$ are isomorphic, so every attractor in $\mathcal{S}_f$ is also a steady state. $\square$

We are now ready to tackle the final version of lemma 3.3.8. There are two cases, $\mathcal{I}_f$ is strongly connected, or not. In the former case, we can use the following result.

Lemma 3.3.12 ([1, Corollary 13.11a])**.** *A strongly connected signed digraph is cycle-balanced if and only if it is balanced.*

So if $\mathcal{I}_f$ is strongly connected, it is balanced by lemma 3.3.12 since it is cycle-balanced. Then all attractors in $\mathcal{S}_f$ are steady states by lemma 3.3.11.

Suppose $\mathcal{I}_f$ is not strongly connected. Let f be an activation function whose interaction graph is cycle-balanced and not strongly connected, and such that $\mathcal{S}_f$ contains periodic attractors. In addition, suppose that among all activation functions with these properties, $|\mathbb{S}_f|$ is minimum. We will show that from any state in an attractor $\mathcal{A}$ of f, there is a dipath to a steady state of f. This will mean that the attractor must have been a steady state to begin with.

We will use a special kind of restricted activation function for the remainder of the proof. Given an activation function f, $\mathbf{x} \in \mathbb{S}_f$ and $W \subseteq C_f$. Let $\mathbf{s} \in \mathbb{S}_f$ such that $s_i = \min S_i$ for $i \in W$ and $s_i = x_i$ otherwise. Let $\mathbf{t} \in \mathbb{S}_f$ such that $t_i = \max S_i$ for $i \in W$ and $t_i = x_i$ otherwise. Similar to equation (3.1), define

$$[W, \mathbf{x}]|f := [\mathbf{s}, \mathbf{t}]|f. \tag{3.2}$$

Lemma 3.3.13. *Given an activation function f, $\mathbf{x} \in \mathbb{S}_f$ and $W \subseteq C_f$. If $g = [W, \mathbf{x}]|f$, then $\mathcal{S}_g \subseteq \mathcal{S}_f$.*

Proof. Suppose the arc $(\mathbf{y}, \mathbf{y}^{i\to a}) \in \mathcal{S}_g$, i.e., $g_i(\mathbf{y}) = a$. Note that $i \notin W$, then $a = x_i = y_i$ since there is only one possible output of g_i in this case. So there would be no arc if this were the case. Therefore $i \in W$. By equation (3.1), $g_i(\mathbf{y}) = f_i(\mathbf{y})$ since s_i and t_i are the minimum and maximum values of S_i respectively. Therefore $f_i(\mathbf{y}) = a$, so $(\mathbf{y}, \mathbf{y}^{i\to a}) \in \mathcal{S}_f$ also. □

Now let $\mathcal{A}$ be an attractor of f, $\mathbf{x}$ be a state in $\mathcal{A}$, and let W be an initial strong component of $\mathcal{I}_f$. Let $g = [W, \mathbf{x}]|f$. By lemma 3.3.1, $\mathcal{I}_g$ is a subgraph of $\mathcal{I}_f$. For this particular activation function g, we can actually say more than this.

First, observe that for all $j \in W$, f_j depends only on inputs in W. To see this, let $j \in W$. If $k \notin W$, there is no arc from k to j in $\mathcal{I}_f$ since W is an initial strong component. Therefore

$$\partial^{k\to a} f_j(\mathbf{x}) = \operatorname{sgn}[f_j(\mathbf{x}^{k\to a}) - f_j(\mathbf{x})][a - x_k] = 0.$$

If $x_k \neq a$, $f_j(\mathbf{x}^{k\to a}) - f_j(\mathbf{x}) = 0$, so $f_j(\mathbf{x}^{k\to a}) = f_j(\mathbf{x})$. So f_j does not depend on inputs outside of W.

Now suppose $(i, j, \sigma_{ij}) \in \mathcal{I}_f$ where $i, j \in W$. So there is $\mathbf{y} \in \mathbb{S}_f$ such that $\partial^{i\to a} f_j(\mathbf{y}) = \sigma_{ij}$. Let $\mathbf{z} \in \mathbb{S}_f$ such that $z_k = y_k$ for all $k \in W$ and $z_k = x_k$ otherwise. Since f_j only depends on inputs from W,

$$\begin{aligned}
\partial^{i\to a} f_j(\mathbf{y}) &= \operatorname{sgn}[f_j(\mathbf{y}^{i\to a}) - f_j(\mathbf{y})][y_i - a] \\
&= \operatorname{sgn}[f_j(\mathbf{z}^{i\to a}) - f_j(\mathbf{z})][z_i - a] \\
&= \operatorname{sgn}[g_j(\mathbf{z}^{i\to a}) - g_j(\mathbf{z})][z_i - a] \\
&= \partial^{i\to a} g_j(\mathbf{z}).
\end{aligned}$$

Therefore $\mathcal{I}_g$ contains every arc between vertices in W that is in $\mathcal{I}_f$. So the subgraph of $\mathcal{I}_g$ on W is actually the induced subgraph of $\mathcal{I}_f$ on W.

Since W is a strong component of $\mathcal{I}_f$, the subgraph of $\mathcal{I}_g$ on W is strongly connected. Since $\mathcal{I}_g$ is cycle-balanced, this strong component of $\mathcal{I}_g$ is balanced by 3.3.12. Since the components in $C_g \setminus W$ can take only a single value, there are no arcs to or from these vertices in $\mathcal{I}_g$, i.e., they are all isolated vertices. Therefore $\mathcal{I}_g$ is balanced. So all attractors in $\mathcal{S}_g$ are steady states by lemma 3.3.11. Therefore, there is a dipath P_1 from $\mathbf{x}$ to the steady state $\mathbf{x}'$ in $\mathcal{S}_g$ by lemma 3.2.1.

Now let $g' = [C_f \setminus W, \mathbf{x}']|f$. By lemma 3.3.1 again, $\mathcal{I}_{g'}$ is a subgraph of $\mathcal{I}_f$. So $\mathcal{I}_{g'}$ is cycle-balanced. Now if $\mathcal{I}_{g'}$ is strongly connected, then it is balanced by lemma 3.3.12. So all attractors of g' are steady states by lemma 3.3.11. If $\mathcal{I}_{g'}$ is not strongly connected, then g' is an activation function whose interaction graph is cycle-balanced and not strongly connected such that $|\mathbb{S}_{g'}| <$

$|\mathbb{S}_f|$. Therefore all attractors of g' are steady states by the minimality of f. Either way, there is a dipath P_2 in $\mathcal{S}_{g'}$ from $\mathbf{x}'$ to $\mathbf{x}''$, a steady state of g'.

I claim that $\mathbf{x}''$ is a steady state of f. To see this, Let $k \in W$. Since there is only one possible value for the components of g' in W, $x''_k = x'_k$. Now f_k depends only on inputs in W, so $f_k(\mathbf{x}'') = f_k(\mathbf{x}')$. By the definition of g, $f_k(\mathbf{x}') = g_k(\mathbf{x}')$. Since $\mathbf{x}'$ is a steady state of g, $g_k(\mathbf{x}') = x'_k = x''_k$. Therefore $f_k(\mathbf{x}'') = x''_k$.

If instead $k \notin W$, then $g'_k(\mathbf{x}'') = x''_k$ since $\mathbf{x}''$ is a steady state of g'. But $g'_k(\mathbf{x}'') = f_k(\mathbf{x})$ by the definition of g, so $f_k(\mathbf{x}'') = x''_k$ in this case also. Therefore $\mathbf{x}''$ is indeed a steady state of f.

Finally, by lemma 3.3.13, P_1 and P_2 are also contained in $\mathcal{S}_f$. Therefore $P_1 \cup P_2 \subset \mathcal{S}_f$ is a dipath from $\mathbf{x}$ to $\mathbf{x}''$, a steady state of f. Remember that we assumed $\mathbf{x}$ is contained in the attractor $\mathcal{A} \subseteq \mathcal{S}_f$. The only way it is possible for there to be a path from a state in an attractor to a steady state is if $\mathcal{A}$ is a steady state to begin with. Therefore all attractors in $\mathcal{S}_f$ are steady states.

$\square$

This completes the proof of the contrapositive statement of theorem 3.3.7.

$\square$

3.4 The Local Version of the Second Conjecture

3.4.1 A Boolean Counterexample

In [5], Adrian Richard proves the following.

Theorem 3.4.1 ([5, Corollary 1])**.** *If an activation function f has no steady states, then $\mathcal{I}_f$ contains a negative cycle.*

Adrian Richard then asks a natural follow-up question. If an activation function f has no steady states, are there states $\mathbf{x}, \mathbf{y} \in \mathbb{S}_f$ such that the local interaction graph $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ contains a negative cycle? Put another way, if every local interaction graph of f is cycle-balanced, must f have a steady state? A non-Boolean counterexample to this question is given in example 6 in [5]. However the question was left open in the boolean case, i.e., when each component of f can takes only two values. The boolean version of the question is also asked in [2]. In [6], Paul Ruet gives a Boolean counterexample on twelve components. I will now present another Boolean activation function that has no steady states and for which every local interaction graph is cycle-balanced, but on only eight components. I suspect that there are no counter examples on fewer components, but I have not been able to prove this yet.

For this activation function f, $C_f = [8] := \{1, 2, \ldots, 8\}$, $S_i = \{\pm 1\}$ and $\mathbb{S}_f = \{\pm 1\}^8$. Since the activation function is boolean and has eight components, it has 256 local interaction graphs. Deriving each of these would be impractical so we will develop some tools to streamline the presentation.

Given an activation function f and λ, a signed permutation of C_f such that $\lambda\mathbb{S}_f = \mathbb{S}_f$. We say that λ *commutes with* f if $f(\lambda\mathbf{x}) = \lambda f(\mathbf{x})$ for all states $\mathbf{x} \in \mathbb{S}_f$. Note that $\lambda\mathbf{x} \in \mathbb{S}_f$ since $\lambda\mathbb{S}_f = \mathbb{S}_f$. So $f(\lambda\mathbf{x})$ makes sense here.

Lemma 3.4.2. *Let f be an activation function and let λ be a signed permutation of C_f such that $\lambda\mathbb{S}_f = \mathbb{S}_f$. If f commutes with λ, then $\mathcal{I}_f(\lambda\mathbf{x}, \lambda\mathbf{y}) = \lambda\mathcal{I}_f(\mathbf{x}, \mathbf{y})$, i.e., $\mathcal{I}_f(\mathbf{x}, \mathbf{y})$ and $\mathcal{I}_f(\lambda\mathbf{x}, \lambda\mathbf{y})$ are switching isomorphic.*

Proof. We use the conjugate activation function f^λ. Under these hypotheses, $f = f^\lambda$. To see this we use equation 2.6.

$$\begin{aligned} f(\mathbf{x}) &= f(\lambda\lambda^{-1}\mathbf{x}) \\ &= \lambda f(\lambda^{-1}\mathbf{x}) \\ &= f^\lambda(\lambda\lambda^{-1}\mathbf{x}) \\ &= f^\lambda(\mathbf{x}). \end{aligned}$$

Note that $f(\lambda^{-1}\mathbf{x})$ makes sense since $\lambda^{-1}\mathbf{x} \in \mathbb{S}_f$ because $\lambda\mathbb{S}_f = \mathbb{S}_f$. Since $f = f^\lambda$, by lemma 2.3.1,

$$\begin{aligned} \mathcal{I}_f(\lambda\mathbf{x}, \lambda\mathbf{y}) &= \mathcal{I}_{f^\lambda}(\lambda\mathbf{x}, \lambda\mathbf{y}) \\ &= \lambda\mathcal{I}_f(\mathbf{x}, \mathbf{y}). \end{aligned}$$

So $\mathcal{I}_f(\lambda\mathbf{x}, \lambda\mathbf{y})$ and $\mathcal{I}_f(\mathbf{x}, \lambda\mathbf{y})$ are switching isomorphic. ☐

So in the presentation of this counterexample, we will have a group Γ of signed permutations of $[8]$ such that every element of Γ commutes with our activation function f. If $\mathcal{I}_f(\mathbf{x})$ is cycle-balanced, then the local interaction graph of any state in the orbit of $\mathbf{x}$ is also cycle-balanced by lemma 3.4.2. So we need only pick a set of distinct orbit representatives and check whether the local interaction graphs at each representative state is cycle-balanced. The action of Γ on $\{\pm 1\}^n$ will have eight orbits. So we will only have to check eight local interaction graphs instead of 256.

Before we can put this plan into action, we have to address how multiplication works in the signed permutation group based on our definitions. Recall definition (2.5) which defines how a switching isomorphism of an activation function f transforms states of f. The set of signed permutations of f forms a group with group product being composition. But how does this composition work? Let $\lambda_1 = (\zeta^1, \pi_1)$ and $\lambda_2 = (\zeta^2, \pi_2)$ be signed permutations of an activation function f. Using equation (2.5) several times,

$$\begin{aligned}
\lambda_2(\lambda_1 \mathbf{x}) &= \lambda_2(\pi_1[\zeta^1 \mathbf{x}]) \\
&= \pi_2(\zeta^2 \pi_1[\zeta^1 \mathbf{x}]) \\
&= \pi_2(\pi_1 \pi_1^{-1})(\zeta^2 \pi_1[\zeta^1 \mathbf{x}]) \\
&= \pi_2 \pi_1[\pi_1^{-1}(\zeta^2 \pi_1[\zeta^1 \mathbf{x}])] \\
&= \pi_2 \pi_1[(\pi_1^{-1}, \zeta^2)\pi_1(\zeta^1 \mathbf{x})] \\
&= \pi_2 \pi_1[(\pi_1^{-1}\zeta^2)(\pi_1^{-1}\pi_1[\zeta^1 \mathbf{x}])] \\
&= \pi_2 \pi_1[(\pi_1^{-1}\zeta^2)\zeta^1 \mathbf{x}] \\
&= (\pi_2 \pi_1, (\pi_1^{-1}\zeta^2)\zeta^1)\mathbf{x}.
\end{aligned}$$

So we have shown that for elements of the signed permutation group,

$$\lambda_2 \lambda_1 = (\pi_2, \zeta^2)(\pi_1, \zeta^1) = (\pi_2 \pi_1, (\pi_1^{-1}\zeta^2)\zeta^1). \tag{3.3}$$

The identity element of the group is $e = (\pi^0, \{+\}^{C_f})$ where π^0 is the identity permutation. What about inverses? Using equation (3.3),

$$\begin{aligned}
(\pi^{-1}, \pi\zeta)(\pi, \zeta) &= (\pi^{-1}\pi, [\pi^{-1}\pi\zeta]\zeta) \\
&= (\pi^0, \zeta\zeta) \\
&= e.
\end{aligned}$$

Similarly $(\pi, \zeta)(\pi^{-1}, \pi\zeta,) = e$, so $(\zeta, \pi)^{-1} = (\pi\zeta, \pi^{-1})$.

Here is a useful formula for powers in this group.

Lemma 3.4.3. *If $\lambda = (\pi, \zeta,)$ is a signed permutation, then*

$$\lambda^n = \left(\pi^n, \prod_{i=1}^{n}(\pi^{1-i}\zeta)\right).$$

Proof. The proof is by induction. The formula is valid for $n = 1$ since

$$\begin{aligned}\lambda^1 &= (\pi^1, \prod_{i=1}^{1} \pi^{1-i}\zeta) \\ &= (\pi, \pi^0\zeta) \\ &= (\pi, \zeta) \\ &= \lambda.\end{aligned}$$

Now suppose that the formula is true for $n < k$. Using equation (3.3),

$$\begin{aligned}\lambda^k &= \lambda\lambda^{k-1} \\ &= (\pi, \zeta)\left(\pi^{k-1}, \prod_{i=1}^{k-1} \pi^{1-i}\zeta\right) \\ &= \left(\pi\pi^{k-1}, \pi^{-(k-1)}\zeta \prod_{i=1}^{k-1} \pi^{1-i}\zeta\right) \\ &= \left(\pi^k, \pi^{1-k}\zeta \prod_{i=1}^{k-1} \pi^{1-i}\zeta\right) \\ &= \left(\pi^k, \prod_{i=1}^{k} \pi^{1-i}\zeta\right).\end{aligned}$$

Therefore the formula is also true for k. So the formula is valid for all n by induction. $\square$

Let $\Gamma = \langle \lambda, \zeta \rangle$, the group generated by $\zeta = (-, +, -, +, -, +, -, +)$ and $\lambda = [\pi, \zeta'] = [(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8), (+, +, +, +, +, +, +, -)]$. Note that for π we are using the cycle notation for permutations. The group Γ will commute with our activation function. So it demands some detailed analysis. We will show that this group has order 32 and its action on $\{\pm 1\}^8$ is free.

Let us first show that the order of Γ is 32. By equation (2.4), $\pi_j \mathbf{x} = x_{\pi^{-1}j}$. It follows that $\pi(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8) = (x_8, x_1, x_2, x_3, x_4, x_5, x_6, x_7)$. For powers of λ, using lemma 3.4.3,

$$\begin{aligned}\lambda^2 &= [\pi^2, (\pi^{-1}\zeta')\zeta'] \\ &= [\pi^2, (+, +, +, +, +, +, -, +)\zeta'] \\ &= [\pi^2, (+, +, +, +, +, +, -, -)].\end{aligned}$$

Similarly, for the remaining powers,

$$\lambda^3 = [\pi^3, (+, +, +, +, +, -, -, -)]$$
$$\lambda^4 = [\pi^4, (+, +, +, +, -, -, -, -)]$$
$$\lambda^5 = [\pi^5, (+, +, +, -, -, -, -, -)]$$
$$\lambda^6 = [\pi^6, (+, +, -, -, -, -, -, -)]$$
$$\lambda^7 = [\pi^7, (+, -, -, -, -, -, -, -)]$$
$$\lambda^8 = (-, -, -, -, -, -, -, -)$$

So λ^{16} is the identity, i.e the order of λ is 16. By equation (3.3),

$$\begin{aligned}
\zeta\lambda &= [\pi, (\pi^{-1}\zeta)\zeta'] \\
&= [\pi, \zeta'](\pi^{-1}\zeta) \\
&= \lambda(+, -, +, -, +, -, +, -) \\
&= \lambda(-, -, -, -, -, -, -, -)(-, +, -, +, -, +, -, +) \\
&= \lambda\lambda^8\zeta \\
&= \lambda^9\zeta.
\end{aligned}$$

So each $\gamma \in \Gamma$ can be written in the form $\lambda^n\zeta^m$ for $n \in [16]$ and $m = 0, 1$ since ζ has order 2. Therefore $|\Gamma| = 32$.

It turns out that the action of Γ on $\mathbb{S}_f$ is free. So when Γ acts on $\mathbb{S}_f$, each orbit of Γ will contains thirty two states. Since $|\mathbb{S}_f| = 256 = 32 \times 8$, there are eight orbits in total. We will need a set of distinct representative states for the orbits of Γ. I claim that the following states will forms a set of distinct representatives.

$$
\begin{aligned}
\mathbf{x}_1 &= (1,1,1,1,1,1,1,1)\\
\mathbf{x}_2 &= (1,-1,1,1,1,1,1,1)\\
\mathbf{x}_3 &= (1,1,-1,1,1,1,1,1)\\
\mathbf{x}_4 &= (1,1,1,-1,1,1,1,1)\\
\mathbf{x}_5 &= (1,1,1,1,-1,1,1,1)\\
\mathbf{x}_6 &= (1,1,1,1,1,-1,1,1)\\
\mathbf{x}_7 &= (1,1,-1,-1,1,1,1,1)\\
\mathbf{x}_8 &= (1,1,-1,1,1,-1,1,1)
\end{aligned}
$$

To show that $\{\mathbf{x}_1,\dots,\mathbf{x}_8\}$ forms a set of distinct representatives, we just look at each set of states $\{\gamma\mathbf{x}_i \mid \gamma \in \Gamma,\ i \in [8]\}$ and confirm that these sets partition $\mathbb{S}_f$ into eight blocks, each containing 32 states. This will also show that this group action is free. The orbit of $\mathbf{x}_1$ is given in table 3.4.1. It occupies half a page of space, so four pages are needed to print each orbit. So to confirm that we have a set of distinct orbit representatives and that the action of Γ on $\mathbb{S}_f$ is free, we just to check that each table contains 32 distinct states and that there is no overlap between the tables. Easy! But in the interest of the readers eyes and the Amazon rain forest, or server space more likely, let us try to be more concise.

For $\mathbf{x} \in \mathbb{S}_f$, since $\zeta = (-,+,-,+,-,+,-,+) \in \Gamma$, either $x_7 = 1$ or $(\zeta\mathbf{x})_7 = 1$. Similarly $\zeta\lambda^8 = (+,-,+,-,+,-,+,-) \in \Gamma$, so either $x_8 = 1$ or $(\zeta\lambda^8\mathbf{x})_8 = 1$. So each state in $\mathbb{S}_f$ is equivalent to a state $\mathbf{x}$ such that $x_7 = x_8 = 1$ under the action of Γ. So we do not need to include states whose seventh of eighth coordinates are negative in the table.

To save even more space, we will use the following function. For $\mathbf{x} \in \mathbb{S}_f$, let $N(\mathbf{x}) = \{i \mid x_i = -1\}$. If we exclude states whose seventh or eight coordinate is negative, $N(\mathbf{x}) \subseteq [6]$. Table 3.2 contains each orbit of Γ where $N(\lambda^n\mathbf{x})$ is represented by $N(\lambda^n\mathbf{x})$, $N(\zeta\lambda^n\mathbf{x})$, $N(\lambda^{n+8}\mathbf{x})$ or $N(\zeta\lambda^{n+8}\mathbf{x})$, depending on which one is a subset of $[6]$. As you can see, the table contains every subset of $[6]$ and there is no overlap between the rows. This justifies that $\{\mathbf{x}_1,\dots,\mathbf{x}_8\}$ forms a set of distinct representatives and that the action of Γ on $\mathbb{S}_f$ is free.

Now we are ready to define the activation function f itself. The first defining property of f is that it commutes with Γ. Now every state $\mathbf{y} = \gamma\mathbf{x}_i$ for some $\gamma \in \Gamma$ and orbit representative $\mathbf{x}_i$. Since f commutes with Γ, $f(\mathbf{y}) = f[\gamma\mathbf{x}_i] = \gamma f(\mathbf{x}_i)$. So to completely define f, we only need to specify the outputs of f on each orbit representative. This will determine the output of f on every

n	$\lambda^n \mathbf{x}_1$	$\lambda^n \zeta \mathbf{x}_1$
0	$(1,1,1,1,1,1,1,1)$	$(-1,1,-1,1,-1,1,-1,1)$
1	$(-1,1,1,1,1,1,1,1)$	$(-1,-1,1,-1,1,-1,1,-1)$
2	$(-1,-1,1,1,1,1,1,1)$	$(1,-1,-1,1,-1,1,-1,1)$
3	$(-1,-1,-1,1,1,1,1,1)$	$(-1,1,-1,-1,1,-1,1,-1)$
4	$(-1,-1,-1,-1,1,1,1,1)$	$(1,-1,1,-1,-1,1,-1,1)$
5	$(-1,-1,-1,-1,-1,1,1,1)$	$(-1,1,-1,1,-1,-1,1,-1)$
6	$(-1,-1,-1,-1,-1,-1,1,1)$	$(1,-1,1,-1,1,-1,-1,1)$
7	$(-1,-1,-1,-1,-1,-1,-1,1)$	$(-1,1,-1,1,-1,1,-1,-1)$
8	$(-1,-1,-1,-1,-1,-1,-1,-1)$	$(1,-1,1,-1,1,-1,1,-1)$
9	$(1,-1,-1,-1,-1,-1,-1,-1)$	$(1,1,-1,1,-1,1,-1,1)$
10	$(1,1,-1,-1,-1,-1,-1,-1)$	$(-1,1,1,-1,1,-1,1,-1)$
11	$(1,1,1,-1,-1,-1,-1,-1)$	$(1,-1,1,1,-1,1,-1,1)$
12	$(1,1,1,1,-1,-1,-1,-1)$	$(-1,1,-1,1,1,-1,1,-1)$
13	$(1,1,1,1,1,-1,-1,-1)$	$(1,-1,1,-1,1,1,-1,1)$
14	$(1,1,1,1,1,1,-1,-1)$	$(-1,1,-1,1,-1,1,1,-1)$
15	$(1,1,1,1,1,1,1,-1)$	$(1,-1,1,-1,1,-1,1,1)$

Table 3.1: Orbit of $\mathbf{x}_1$

i	$N(\mathbf{x}_i)$	$N(\lambda \mathbf{x}_i)$	$N\left(\lambda^2 \mathbf{x}_i\right)$	$N(\lambda^3 \mathbf{x}_i)$	$N(\lambda^4 \mathbf{x}_i)$	$N(\lambda^5 \mathbf{x}_i)$	$N(\lambda^6 \mathbf{x})$	$N(\lambda^7 \mathbf{x}_i)$
1	$\{\}$	$\{1\}$	$\{1,2\}$	$\{1,2,3\}$	$\{1,2,3,4\}$	$\{1,2,3,4,5\}$	$\{1,2,3,4,5,6\}$	$\{2,4,6\}$
2	$\{2\}$	$\{1,3\}$	$\{1,2,4\}$	$\{1,2,3,5\}$	$\{1,2,3,4,6\}$	$\{2,4\}$	$\{1,3,5\}$	$\{1,2,4,6\}$
3	$\{3\}$	$\{1,4\}$	$\{1,2,5\}$	$\{1,2,3,6\}$	$\{2,4,5\}$	$\{1,3,5,6\}$	$\{2,3,4,5,6\}$	$\{4,6\}$
4	$\{4\}$	$\{1,5\}$	$\{1,2,6\}$	$\{2,5\}$	$\{1,3,6\}$	$\{2,3,4,5\}$	$\{1,3,4,5,6\}$	$\{2,3,4,6\}$
5	$\{5\}$	$\{1,6\}$	$\{2,3,5\}$	$\{1,3,4,6\}$	$\{2,3,4\}$	$\{1,3,4,5\}$	$\{1,2,4,5,6\}$	$\{2,6\}$
6	$\{6\}$	$\{3,5\}$	$\{1,4,6\}$	$\{2,3\}$	$\{1,3,4\}$	$\{1,2,4,5\}$	$\{1,2,3,5,6\}$	$\{2,4,5,6\}$
7	$\{3,4\}$	$\{1,4,5\}$	$\{1,2,5,6\}$	$\{2,5,6\}$	$\{5,6\}$	$\{3,5,6\}$	$\{3,4,5,6\}$	$\{3,4,6\}$
8	$\{3,6\}$	$\{3,4,5\}$	$\{1,4,5,6\}$	$\{2,3,6\}$	$\{4,5\}$	$\{1,5,6\}$	$\{2,3,5,6\}$	$\{4,5,6\}$

Table 3.2: Orbits of Γ

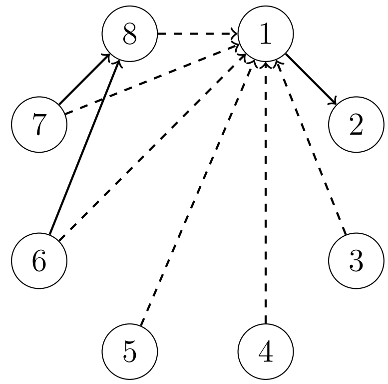

Figure 3.3: $\mathcal{I}_f(\mathbf{x}_1)$

state via the group action.

Here are the outputs of f for each orbit representative.

$$\begin{aligned} f(\mathbf{x}_1) = f(\mathbf{x}_2) &= (-1, 1, 1, 1, 1, 1, 1, 1) \\ f(\mathbf{x}_3) = f(\mathbf{x}_4) = f(\mathbf{x}_5) &= \mathbf{x}_1 \\ f(\mathbf{x}_6) &= (1, 1, 1, 1, 1, 1, 1, -1) \\ f(\mathbf{x}_7) &= \mathbf{x}_4 \\ f(\mathbf{x}_8) &= (-1, 1, -1, 1, 1, -1, 1, 1). \end{aligned}$$

For any $\gamma \in \Gamma$, $\mathcal{I}_f(\mathbf{x})$ and $\mathcal{I}_f[\gamma\mathbf{x}]$ are switching isomorphic by lemma 3.4.2. Therefore, if $\mathbf{x}$ and $\mathbf{y}$ are in the same orbit of Γ, $\mathcal{I}_f(\mathbf{x})$ and $\mathcal{I}_f(\mathbf{y})$ are switching isomorphic. Also, since $f(\mathbf{x}) \neq \mathbf{x}$ for each orbit representative, $f[\gamma\mathbf{x}] = \gamma f(\mathbf{x}) \neq \gamma(\mathbf{x})$ since the action of Γ is free. Therefore f has no steady states, as required.

Next we need to show that $\mathcal{I}_f(\mathbf{x}_i)$ is cycle-balanced for each $i \in [8]$. I will generate $\mathcal{I}_f(\mathbf{x}_1)$ in detail so we can see how the process goes. We will need to know $f(\mathbf{x}_1^{i\to})$ for each $i \in [8]$. First, notice that $\mathbf{x}_1^{i\to} = \mathbf{x}_i$ for $i \in [2, 6]$. So we know the output of f for each of these states from above. Now $\mathbf{x}_1^{1\to} = \lambda\mathbf{x}_1$, so $f(\mathbf{x}_1^{1\to}) = \lambda f(\mathbf{x}_1) = (-1, -1, 1, 1, 1, 1, 1, 1)$. Next, $\mathbf{x}_1^{7\to} = \lambda^{-2}\mathbf{x}_2$, so $f(\mathbf{x}_1^{7\to}) = \lambda^{-2} f(\mathbf{x}_2) = (1, 1, 1, 1, 1, 1, 1, -1)$. Finally, $\mathbf{x}_1^{8\to} = \lambda^{-1}\mathbf{x}_1$, so $f(\mathbf{x}_1^{8\to}) = \lambda^{-1} f(\mathbf{x}_1) = \mathbf{x}_1$. This is all the data needed to generate the adjacency matrix of $\mathcal{I}_f(\mathbf{x}_1)$. A picture of $\mathcal{I}_f(\mathbf{x}_1)$ is given in figure 3.3. It contains no cycles, so it is cycle-balanced.

For the local interaction graph of each other orbit representative, we proceed in the same way. To find $f(\mathbf{x}_i^{j\to})$, we just need to express $\mathbf{x}_i^{j\to}$ as $\gamma\mathbf{x}_k$. Then $f(\mathbf{x}_i^{j\to}) = \gamma f(\mathbf{x}_k)$.

$$\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_1)] = \begin{bmatrix} \partial^{1\to} f(\mathbf{x}_1) \\ \partial^{2\to} f(\mathbf{x}_1) \\ \partial^{3\to} f(\mathbf{x}_1) \\ \partial^{4\to} f(\mathbf{x}_1) \\ \partial^{5\to} f(\mathbf{x}_1) \\ \partial^{6\to} f(\mathbf{x}_1) \\ \partial^{7\to} f(\mathbf{x}_1) \\ \partial^{8\to} f(\mathbf{x}_1) \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_1^{1\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{2\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{3\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{4\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{5\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{6\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{7\to}) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_1^{8\to}) - f(\mathbf{x}_1)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda\mathbf{x}_1) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_2) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_3) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_4) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_5) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\mathbf{x}_6) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\lambda^{-2}\mathbf{x}_2) - f(\mathbf{x}_1)] \\ -\operatorname{sgn}[f(\lambda^{-1}\mathbf{x}_1) - f(\mathbf{x}_1)] \end{bmatrix}$$

$$= \begin{bmatrix} -\operatorname{sgn}[(-1,-1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(-1,1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,-1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,-1) - (-1,1,1,1,1,1,1,1)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,1) - (-1,1,1,1,1,1,1,1)] \end{bmatrix}$$

$$= \begin{bmatrix} -\operatorname{sgn}[(0,-2,0,0,0,0,0,0)] \\ -\operatorname{sgn}[(0,0,0,0,0,0,0,0)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,0)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,0)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,0)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,-2)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,-2)] \\ -\operatorname{sgn}[(2,0,0,0,0,0,0,0)] \end{bmatrix} = \begin{bmatrix} 0 & + & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & + \\ - & 0 & 0 & 0 & 0 & 0 & 0 & + \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$
\begin{aligned}
\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_2)] &= \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_2^{1\to}) - f(\mathbf{x}_2)] \\ \operatorname{sgn}[f(\mathbf{x}_2^{2\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{3\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{4\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{5\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{6\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{7\to}) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\mathbf{x}_2^{8\to}) - f(\mathbf{x}_2)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda^2\mathbf{x}_1) - f(\mathbf{x}_2)] \\ \operatorname{sgn}[f(\mathbf{x}_1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\lambda^3\mathbf{x}_6) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\zeta\lambda^5\mathbf{x}_2) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\zeta\lambda^3\mathbf{x}_4) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\zeta\lambda^7\mathbf{x}_5) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\zeta\lambda^3\mathbf{x}_2) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[f(\lambda^{-1}\mathbf{x}_3) - f(\mathbf{x}_2)] \end{bmatrix} \\
&= \begin{bmatrix} -\operatorname{sgn}[(-1,-1,-1,1,1,1,1,1) - f(\mathbf{x}_2)] \\ \operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(-1,-1,1,1,1,1,1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(1,-1,1,-1,1,-1,-1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(1,-1,1,1,-1,1,-1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(1,-1,1,-1,1,-1,1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(1,-1,1,-1,-1,1,-1,1) - f(\mathbf{x}_2)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,-1) - f(\mathbf{x}_2)] \end{bmatrix} \\
&= \begin{bmatrix} 0 & + & + & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & + & 0 & 0 & 0 & 0 & 0 & 0 \\ - & + & 0 & + & 0 & + & + & 0 \\ - & + & 0 & 0 & + & 0 & + & 0 \\ - & + & 0 & + & 0 & + & 0 & 0 \\ - & + & 0 & + & + & 0 & + & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & + \end{bmatrix}
\end{aligned}
$$

$$
\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_3)] = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_3^{1\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{2\rightarrow}) - f(\mathbf{x}_3)] \\ \operatorname{sgn}[f(\mathbf{x}_3^{3\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{4\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{5\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{6\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{7\rightarrow}) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_3^{8\rightarrow}) - f(\mathbf{x}_3)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda\mathbf{x}_2) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\lambda^3\mathbf{x}_6) - f(\mathbf{x}_3)] \\ \operatorname{sgn}[f(\mathbf{x}_1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_7) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_6) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\mathbf{x}_8) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_4) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[f(\lambda^{-1}\mathbf{x}_4) - f(\mathbf{x}_3)] \end{bmatrix}
$$

$$
= \begin{bmatrix} -\operatorname{sgn}[(-1,-1,1,1,1,1,1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(-1,-1,1,1,1,1,1,1) - f(\mathbf{x}_3)] \\ \operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(1,1,1,-1,1,1,1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(-1,1,-1,1,-1,1,-1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(-1,1,-1,1,1,-1,1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(1,1,-1,1,-1,1,-1,1) - f(\mathbf{x}_3)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,1,-1) - f(\mathbf{x}_3)] \end{bmatrix}
$$

$$
= \begin{bmatrix} + & + & 0 & 0 & 0 & 0 & 0 & 0 \\ + & + & 0 & 0 & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ + & 0 & + & 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & 0 & + & 0 & 0 \\ 0 & 0 & + & 0 & + & 0 & + & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \end{bmatrix}
$$

$$\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_4)] = \begin{bmatrix} -\,\mathrm{sgn}[f(\mathbf{x}_4^{1\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{2\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{3\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{4\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{5\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{6\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{7\to}) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_4^{8\to}) - f(\mathbf{x}_4)] \end{bmatrix} = \begin{bmatrix} -\,\mathrm{sgn}[f(\lambda \mathbf{x}_3) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\zeta\lambda^5 \mathbf{x}_2) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\mathbf{x}_7) - f(\mathbf{x}_4)] \\ \mathrm{sgn}[f(\mathbf{x}_1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\zeta\lambda^4 \mathbf{x}_8) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\zeta\lambda^7 \mathbf{x}_3) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\zeta\lambda^5 \mathbf{x}_5) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[f(\lambda^{-1} \mathbf{x}_5) - f(\mathbf{x}_4)] \end{bmatrix}$$

$$= \begin{bmatrix} -\,\mathrm{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,-1,1,-1,1,-1,-1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,1,1,-1,1,1,1,1) - f(\mathbf{x}_4)] \\ \mathrm{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,1,1,-1,1,1,1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,-1,1,-1,1,-1,1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,-1,1,-1,1,1,-1,1) - f(\mathbf{x}_4)] \\ -\,\mathrm{sgn}[(1,1,1,1,1,1,1,-1) - f(\mathbf{x}_4)] \end{bmatrix}$$

$$= \begin{bmatrix} + & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & + & 0 & + & 0 & + & + & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ - & 0 & 0 & + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ 0 & + & 0 & + & 0 & + & 0 & 0 \\ 0 & + & 0 & + & 0 & 0 & + & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \end{bmatrix}$$

$$\operatorname{Adj}[\mathcal{I}_f(\mathbf{x}_5)] = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_5^{1\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{2\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{3\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{4\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{5\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{6\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{7\to}) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\mathbf{x}_5^{8\to}) - f(\mathbf{x}_5)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda\mathbf{x}_4) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\zeta\lambda^3\mathbf{x}_4) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_6) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\zeta\lambda^4\mathbf{x}_8) - f(\mathbf{x}_5)] \\ \operatorname{sgn}[f(\mathbf{x}_1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\lambda^{-4}\mathbf{x}_7) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_2) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[f(\lambda^{-1}\mathbf{x}_6) - f(\mathbf{x}_5)] \end{bmatrix}$$

$$= \begin{bmatrix} -\operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(1,-1,1,1,-1,1,-1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(-1,1,-1,1,-1,1,-1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(1,1,1,-1,1,1,1,1) - f(\mathbf{x}_5)] \\ \operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(1,1,1,1,-1,-1,-1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(1,-1,-1,1,-1,1,-1,1) - f(\mathbf{x}_5)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,-1,-1) - f(\mathbf{x}_5)] \end{bmatrix}$$

$$= \begin{bmatrix} + & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & + & 0 & 0 & + & 0 & + & 0 \\ + & 0 & + & 0 & + & 0 & + & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & + & + & + & 0 \\ 0 & + & + & 0 & + & 0 & + & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & + & + \end{bmatrix}$$

$$\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_6)] = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_6^{1\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{2\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{3\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{4\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{5\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{6\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{7\to}) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_6^{8\to}) - f(\mathbf{x}_6)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda\mathbf{x}_5) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\zeta\lambda^7\mathbf{x}_5) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\mathbf{x}_8) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\zeta\lambda^7\mathbf{x}_3) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\lambda^{-4}\mathbf{x}_7) - f(\mathbf{x}_6)] \\ \operatorname{sgn}[f(\mathbf{x}_1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\lambda^{-3}\mathbf{x}_3) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[f(\lambda^{-3}\mathbf{x}_2) - f(\mathbf{x}_6)] \end{bmatrix}$$

$$= \begin{bmatrix} -\operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(1,-1,1,-1,1,-1,1,1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(-1,1,-1,1,1,-1,1,1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(1,-1,1,-1,1,-1,1,1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(1,1,1,1,-1,-1,-1,1) - f(\mathbf{x}_6)] \\ \operatorname{sgn}[(-1,1,1,1,1,1,1,1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(1,1,1,1,1,-1,-1,-1) - f(\mathbf{x}_6)] \\ -\operatorname{sgn}[(1,1,1,1,1,1,-1,-1) - f(\mathbf{x}_6)] \end{bmatrix}$$

$$= \begin{bmatrix} + & 0 & 0 & 0 & 0 & 0 & 0 & - \\ 0 & + & 0 & + & 0 & + & 0 & - \\ + & 0 & + & 0 & 0 & + & 0 & - \\ 0 & + & 0 & + & 0 & + & 0 & - \\ 0 & 0 & 0 & 0 & + & + & + & - \\ - & 0 & 0 & 0 & 0 & 0 & 0 & + \\ 0 & 0 & 0 & 0 & 0 & + & + & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & + & 0 \end{bmatrix}$$

$$\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_7)] = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_7^{1\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{2\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{3\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{4\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{5\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{6\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{7\to}) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\mathbf{x}_7^{8\to}) - f(\mathbf{x}_7)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\lambda^4\mathbf{x}_6) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\lambda^4\mathbf{x}_5) - f(\mathbf{x}_7)] \\ \operatorname{sgn}[f(\mathbf{x}_4) - f(\mathbf{x}_7)] \\ \operatorname{sgn}[f(\mathbf{x}_3) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_8) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\zeta\lambda^7\mathbf{x}_7) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\zeta\lambda\mathbf{x}_7) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[f(\zeta\lambda^{-5}\mathbf{x}_8) - f(\mathbf{x}_7)] \end{bmatrix}$$

$$= \begin{bmatrix} -\operatorname{sgn}[(-1,-1,-1,1,1,1,1,1) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[(-1,-1,-1,-1,1,1,1,1) - f(\mathbf{x}_7)] \\ \operatorname{sgn}[(1,1,1,1,1,1,1,1) - f(\mathbf{x}_7)] \\ \operatorname{sgn}[(1,1,1,1,1,1,1,1) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[(1,-1,-1,-1,-1,1,1,1) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[(1,-1,-1,-1,1,-1,1,1) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[(-1,1,-1,1,1,1,-1,1) - f(\mathbf{x}_7)] \\ -\operatorname{sgn}[(1,1,-1,1,1,1,1,-1) - f(\mathbf{x}_7)] \end{bmatrix}$$

$$= \begin{bmatrix} + & + & + & - & 0 & 0 & 0 & 0 \\ + & + & + & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & + & 0 & 0 & 0 & 0 \\ 0 & + & + & 0 & + & 0 & 0 & 0 \\ 0 & + & + & 0 & 0 & + & 0 & 0 \\ 0 & 0 & + & - & 0 & 0 & + & 0 \\ 0 & 0 & + & - & 0 & 0 & 0 & + \end{bmatrix}$$

$$
\mathrm{Adj}[\mathcal{I}_f(\mathbf{x}_8)] = \begin{bmatrix} -\operatorname{sgn}[f(\mathbf{x}_8^{1\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{2\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{3\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{4\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{5\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{6\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{7\to}) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\mathbf{x}_8^{8\to}) - f(\mathbf{x}_8)] \end{bmatrix} = \begin{bmatrix} -\operatorname{sgn}[f(\zeta\lambda^{-4}\mathbf{x}_4) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\lambda^{3}\mathbf{x}_8) - f(\mathbf{x}_8)] \\ \operatorname{sgn}[f(\mathbf{x}_6) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\zeta\lambda^{7}\mathbf{x}_7) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\zeta\lambda^{-3}\mathbf{x}_7) - f(\mathbf{x}_8)] \\ \operatorname{sgn}[f(\mathbf{x}_3) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\lambda^{-3}\mathbf{x}_8) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[f(\zeta\lambda^{-4}\mathbf{x}_5) - f(\mathbf{x}_8)] \end{bmatrix}
$$

$$
= \begin{bmatrix} -\operatorname{sgn}[(-1,1,-1,1,1,-1,1,1) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[(1,-1,-1,-1,1,-1,1,1) - f(\mathbf{x}_8)] \\ \operatorname{sgn}[(1,1,1,1,1,1,1,-1) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[(1,-1,-1,-1,1,-1,1,1) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[(1,1,-1,1,-1,-1,1,-1) - f(\mathbf{x}_8)] \\ \operatorname{sgn}[(1,1,1,1,1,1,1,1) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[(1,1,-1,1,1,1,-1,1) - f(\mathbf{x}_8)] \\ -\operatorname{sgn}[(-1,1,-1,1,1,-1,1,-1) - f(\mathbf{x}_8)] \end{bmatrix}
$$

$$
= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ - & + & 0 & + & 0 & 0 & 0 & 0 \\ + & 0 & + & 0 & 0 & + & 0 & - \\ - & + & 0 & + & 0 & 0 & 0 & 0 \\ - & 0 & 0 & 0 & + & 0 & 0 & + \\ + & 0 & + & 0 & 0 & + & 0 & 0 \\ - & 0 & 0 & 0 & 0 & - & + & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & + \end{bmatrix}
$$

Pictures of the local interaction graphs for the orbit representative of Γ are given in figures 3.4.1 and 3.4.1. Each of these interaction graphs is cycle-balanced. Since every local interaction graph is switching isomorphic to one of these, every local interaction graph is cycle-balanced. So f is indeed an activation function whose local interaction graphs contain no negative cycles that has no steady-states.

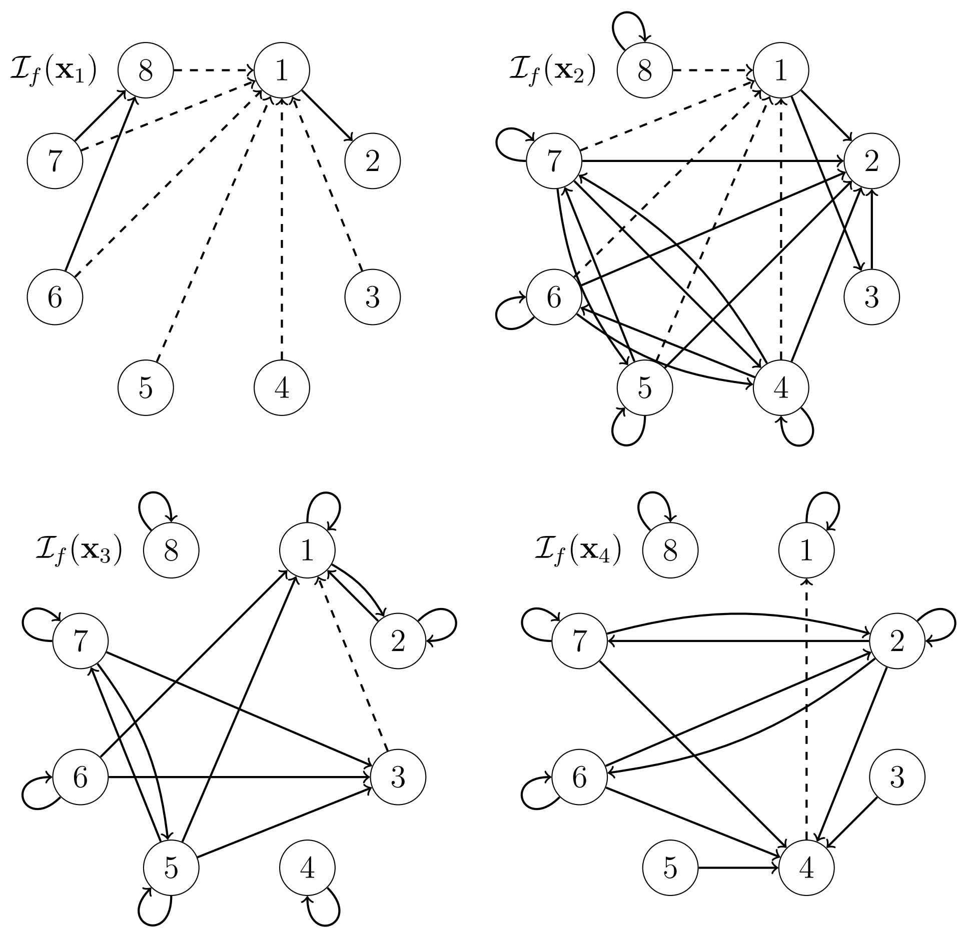

Figure 3.4: Local interaction graphs of the first four orbit representatives.

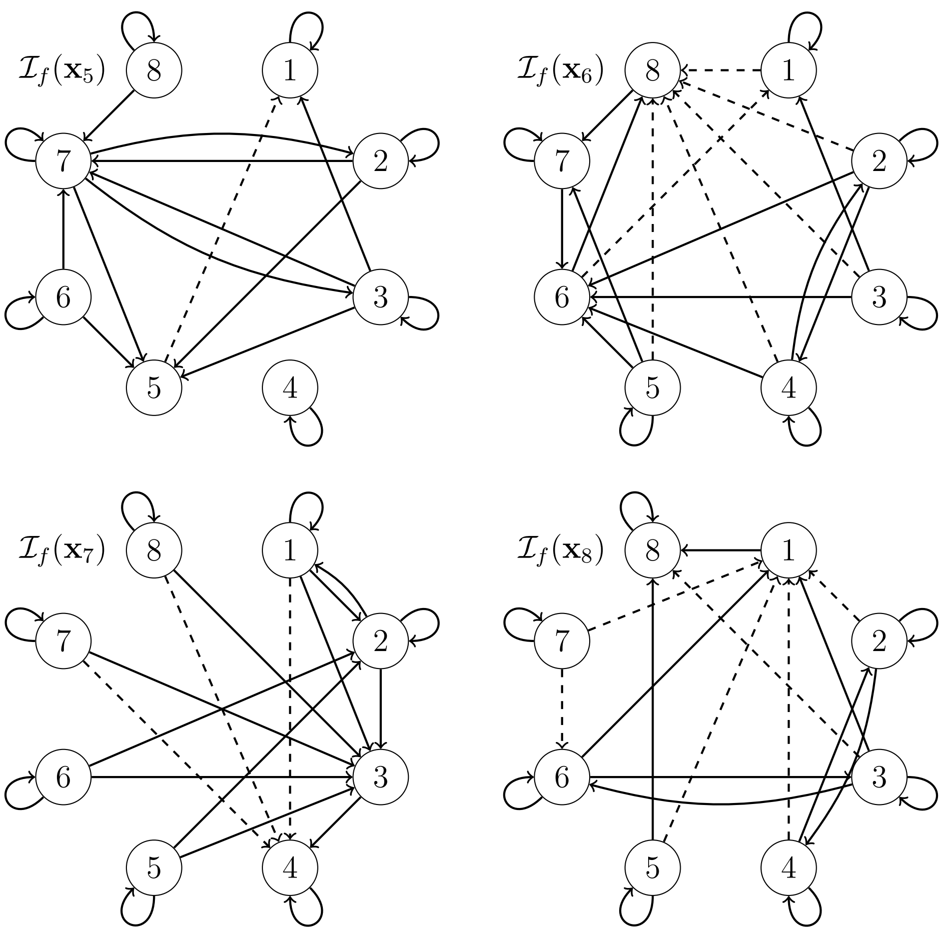

Figure 3.5: Local interaction graphs of last four orbit representatives.

Bibliography

[1] Frank Harary, Robert Z. Norman and Dorwin Cartwright,
Structural Models: An Introduction to the Theory of Directed Graphs.
Wiley, New York, 1965.

[2] Jean-Paul Comet, Mathilde Noual, Adrien Richard, Julio Aracena, Laurence Calzone, Jacques Demongeot, Marcelle Kaufman, Aurélien Naldi, El Houssine Snoussi, Denis Thieffry,
On circuit functionality in boolean networks.
Bulletin of Mathematical Biology **75** (2013) 906–919. MR3070304

[3] Akihiko Nakajima, Takako Isshiki, Kunihiko Kaneko, Shuji Ishihara,
Robustness under functional constraint: The genetic network for temporal expression in drosophila neurogenesis.
PLoS Computational Biology **6**(4) (2010), e1000760.

[4] Adrien Richard, Jean-Paul Comet,
Necessary conditions for multistationarity in discrete dynamical systems.
Discrete Applied Mathematics **155** (2007) 2403–2413. MR2365052 (2008m:37032).

[5] Adrien Richard,
Negative circuits and sustained oscillations in asynchronous automata networks.
Advances in Applied Mathematics **44** (2010) 378–392. MR2600786 (2011a:92007).

[6] Paul Ruet,
Negative local feedbacks in boolean networks.
Discrete Applied Mathematics **221** (2017) 1–17. MR3612582

[7] René Thomas,
Boolean formalization of genetic control circuits.
Journal of Theoretical Biology **42**(3) (1973), 563–585.

[8] Thomas Zaslavsky,
Signed graphs.
Discrete Applied Mathematics **4** (1982) 47–74. MR0676405