Pattern Completion in Symmetric Threshold-Linear Networks

Carina Curto
ccurto@psu.edu
Department of Mathematics, Pennsylvania State University, University Park,
PA 16802, U.S.A.

Katherine Morrison
katherine.morrison@unco.edu
Department of Mathematics, Pennsylvania State University, University Park,
PA 16802, U.S.A., and School of Mathematical Sciences, University of Northern
Colorado, Greeley, CO 80639, U.S.A.

Threshold-linear networks are a common class of firing rate models that describe recurrent interactions among neurons. Unlike their linear counterparts, these networks generically possess multiple stable fixed points (steady states), making them viable candidates for memory encoding and retrieval. In this work, we characterize stable fixed points of general threshold-linear networks with constant external drive and discover constraints on the coexistence of fixed points involving different subsets of active neurons. In the case of symmetric networks, we prove the following antichain property: if a set of neurons τ is the support of a stable fixed point, then no proper subset or superset of τ can support a stable fixed point. Symmetric threshold-linear networks thus appear to be well suited for pattern completion, since the dynamics are guaranteed not to get stuck in a subset or superset of a stored pattern. We also show that for any graph G, we can construct a network whose stable fixed points correspond precisely to the maximal cliques of G. As an application, we design network decoders for place field codes and demonstrate their efficacy for error correction and pattern completion. The proofs of our main results build on the theory of permitted sets in threshold-linear networks, including recently developed connections to classical distance geometry.

1 Introduction

In this work, we study stable fixed points (equivalently: steady states, fixed-point attractors, or stable equilibria) of threshold-linear networks with constant external drive. These networks model the activity of a population of neurons with recurrent interactions, whose dynamics are governed by the system of equations:

$$\frac{dx_i}{dt} = -x_i + \left[\sum_{j=1}^{n} W_{ij}x_j + \theta \right]_+, \quad i \in [n]. \tag{1.1}$$

Here $[n] = \{1, \ldots, n\}$ is the index set for n neurons, $x_i(t)$ is the activity level (firing rate) of the ith neuron, and $\theta \in \mathbb{R}$ is a constant external drive (the same for each neuron). The real-valued $n \times n$ matrix W governs the interactions between neurons, with W_{ij} the effective connection strength from the jth to the ith neuron. The threshold-linear function $[y]_+ = \max\{0, y\}$ ensures that $x_i(t) \geq 0$ for all $t > 0$, provided $x_i(0) \geq 0$.

Equations 1.1 differ from the linear system of ordinary differential equations, $\dot{x} = (-I + W)x + \theta$, only by the nonlinearity $[\ ]_+$. This nonlinearity is quite significant, however, as it allows the network to possess multiple stable fixed points, even though the analogous linear system can have at most one. Multistability is the key feature that makes threshold-linear networks viable models of memory encoding and retrieval (Tsodyks & Sejnowski, 1995; Simmen, Treves, & Rolls, 1996; Vogels, Rajan, & Abbott, 2005; Seung & Yuste, 2012). If, for example, W is a symmetric matrix with nonpositive entries, then the system given by equation 1.1 is guaranteed to converge to a stable fixed point for any initial condition (Hahnloser, Seung, & Slotine, 2003). The threshold-linear network thus functions as a traditional attractor neural network, in the spirit of the Hopfield model (1982, 1984), with initial conditions playing the role of inputs and stable fixed points comprising outputs.

In this letter, we characterize stable fixed points of general threshold-linear networks and discover constraints on the coexistence of fixed points involving different subsets of active neurons. In the rest of this section, we give an overview of our main results, theorems 1 and 3, and explain their relevance to pattern completion. We then illustrate their power in an application. The remainder of the letter lays the foundation for proving the main results. Section 2 summarizes relevant notation and background about permitted sets and their relationship to fixed points. In section 3 we provide general conditions that must be satisfied by fixed points of equation 1.1. In section 4, we prove a key technical result using classical distance geometry. Finally, in section 5 we prove our main theorems by combining results from sections 3 and 4.

1.1 Fixed Points of Threshold-Linear Networks. A vector $x^* \in \mathbb{R}^n_{\geq 0}$ is a fixed point of equation 1.1 if, when evaluating at $x = x^*$, we obtain $dx_i/dt = 0$ for each i. The support of a fixed point is given by

$$\text{supp}(x^*) \stackrel{\text{def}}{=} \{i \in [n] \mid x_i^* > 0\}.$$

We use Greek letters, such as $\sigma, \tau \subseteq [n]$, to denote supports. A principal submatrix is obtained from a larger $n \times n$ matrix by restricting both row and

column indices to some $\sigma \subset [n]$. For example, W_σ is the $|\sigma| \times |\sigma|$ principal submatrix of connection strengths among neurons in σ. Similarly, x_σ is the vector of firing rates for only the neurons in σ.

Although a threshold-linear network may have many stable fixed points, each fixed point is completely determined by its support. To see why, observe that if x^* is a fixed point with support $\sigma \subseteq [n]$, then

$$x_\sigma^* = [W_\sigma x_\sigma^* + \theta 1_\sigma]_+ > 0,$$

where 1_σ is a column vector of all ones. This means we can drop the nonlinearity to obtain $(I - W_\sigma)x_\sigma^* = \theta 1_\sigma$. If x^* is a stable fixed point, then $I - W_\sigma$ is invertible (Curto, Degeratu, & Itskov, 2012, theorem 1.2), and hence we can solve for x_σ^* explicitly as

$$x_\sigma^* = \theta(I - W_\sigma)^{-1}1_\sigma.$$

Of course, this is a fixed point of equation 1.1 only if $x_\sigma^* > 0$ and $x_k^* = [\sum_{i \in \sigma} W_{ki}x_i^* + \theta]_+ = 0$ for all $k \notin \sigma$. If either of these conditions fails, then the fixed point with support σ does not exist. If it does exist, however, the above formula gives the precise values of the nonzero entries x_σ^* and guarantees that it is unique. Thus, in order to understand the stable fixed points of a network of the form 1.1, it suffices to characterize the possible supports.

1.2 Summary of Main Results. Given a choice of W and θ, what are the possible stable fixed point supports? In this letter, we provide a set of conditions that fully characterize these supports for general threshold-linear networks and show how the conditions simplify when W is inhibitory or symmetric (see section 3). The compatibility of the fixed-point conditions across multiple $\sigma \subseteq [n]$ enables us to obtain results about which collections $\{\sigma_i\}$ of supports can (or cannot) coexist in the same network. Our strongest results in this vein arise when specializing to symmetric W. In this case, we can build on the geometric theory of permitted sets, as introduced in (Curto, Degeratu, & Itskov, 2013) in order to greatly constrain the collection of allowed fixed point supports. Theorem 1 is our first main result:

Theorem 1 (antichain property). *Consider the threshold-linear network of the form 1.1, for a symmetric matrix W with zero diagonal. If there exists a stable fixed point with support $\tau \subseteq [n]$, then there is no stable fixed point with support σ for any $\sigma \subsetneq \tau$ or $\sigma \supsetneq \tau$.*

The proof is given in section 5 and relies critically on a technical result, proposition 3, which we state and prove in section 4 using ideas from classical distance geometry.

Theorem 1 has several immediate consequences. First, it implies that the set of possible fixed-point supports is an antichain in the Boolean lattice
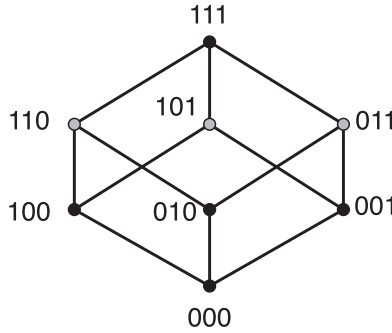
Figure 1: A Boolean lattice for $n = 3$. A maximal antichain, $\sigma_1 = \{1, 2\}, \sigma_2 = \{1, 3\}$, and $\sigma_3 = \{2, 3\}$, is shown in gray. This antichain has size $\binom{3}{\lfloor 3/2 \rfloor} = 3$, in agreement with Sperner's theorem.

$(2^{[n]}, \subseteq)$, where an antichain is defined as a set of incomparable elements in the poset (i.e., no two elements are related by $\subseteq$). Sperner's theorem (Anderson, 1987) states that the cardinality of a maximum antichain in the Boolean lattice is precisely $\binom{n}{\lfloor n/2 \rfloor}$ (see Figure 1). We thus have the corollary:

Corollary 1. *If W is symmetric with zero diagonal, then the threshold-linear network of the form 1.1 has at most $\binom{n}{\lfloor n/2 \rfloor}$ stable fixed points.*

We note, however, that this upper bound is not necessarily tight. We currently do not know whether there exist networks with $\binom{n}{\lfloor n/2 \rfloor}$ stable fixed-points for each n. More generally, it is an open question to determine which antichains in the Boolean lattice can be realized as the set of fixed-point supports for a symmetric W.

A second consequence of theorem 1 is that symmetric threshold-linear networks appear well suited for pattern completion. A network can perform pattern completion if, after initializing at a subset of a pattern, the network dynamics evolve the activity to the complete pattern. In this context, a pattern of the network is a subset of neurons corresponding to the support of a stable fixed point. Theorem 1 implies that if $\tau \subseteq [n]$ is a pattern of a symmetric network, then the network activity is guaranteed not to "get stuck" in a subpattern $\sigma \subsetneq \tau$ or a superpattern $\sigma \supsetneq \tau$.

In order to further illustrate the power of theorem 1, we now turn to the special case of symmetric threshold-linear networks with binary synapses. Here the connection matrix $W = W(G, \varepsilon, \delta)$ is specified by a simple graph[1] G, whose vertices correspond to neurons:

---

[1]Recall that a *simple graph* has undirected edges, no self-loops, and at most one edge between each pair of vertices.
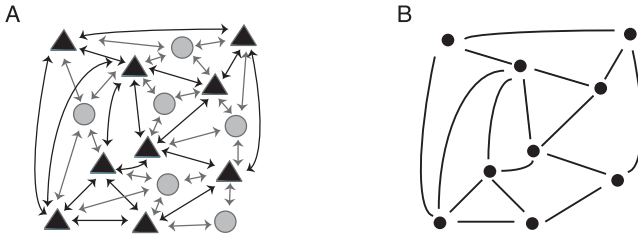
Figure 2: Excitatory connections in a sea of inhibition. (A) Pyramidal neurons (black triangles) have bidirectional synapses onto each other. Inhibitory interneurons (gray circles) produce nonspecific inhibition to all neighboring pyramidal cells. (B) The graph of the network in panel A, retaining only excitatory neurons and their connections. A missing edge corresponds to strong competitive inhibition, $-1 - \delta$, while the presence of an edge indicates attenuated inhibition, $-1 + \varepsilon$, resulting from the sum of global background inhibition and an excitatory connection.

$$W_{ij} = \begin{cases} 0 & \text{if } i = j, \\ -1 + \varepsilon & \text{if } (ij) \in G, \\ -1 - \delta & \text{if } (ij) \notin G. \end{cases} \qquad (1.2)$$

The parameters $\varepsilon, \delta \in \mathbb{R}$ satisfy $0 < \varepsilon < 1$ and $\delta > 0$, while $(ij) \in G$ indicates that there is an edge between vertices i and j. Note that since $\varepsilon < 1$, the network is inhibitory ($W_{ij} \leq 0$). One might interpret W_{ij} as the effective connection strength between neurons i and j due to competitive inhibition that is attenuated by excitation whenever $(ij) \in G$ (see Figure 2).

In the case of binary symmetric networks, combining our characterization of fixed point supports with theorem 1, we obtain our second main result. We need some standard graph-theoretic terminology. A subset of vertices σ is a *clique* of G if $(ij) \in G$ for all pairs $i, j \in \sigma$. In other words, a clique is a subset of neurons that is all-to-all connected. A clique σ is called *maximal* if it is not contained in any larger clique of G.

Theorem 2. *Let G be a simple graph, and consider the network of the form 1.1 with $W = W(G, \varepsilon, \delta)$ for any $0 < \varepsilon < 1, \delta > 0$, and $\theta > 0$. The stable fixed points of this network are in one-to-one correspondence with the maximal cliques of G. Specifically, each maximal clique σ is the support of a stable fixed point, given by*

$$x_\sigma^* = \frac{\theta}{(1 - \varepsilon)|\sigma| + \varepsilon} 1_\sigma,$$

and there are no other stable fixed points.

The proof of theorem 2, given in section 5, demonstrates that it is possible to have a network in which all stable fixed-point supports correspond to maximal stored patterns. This situation is ideal for pattern completion. We then show that this feature generalizes to a much broader class of symmetric networks (see theorem 7 in section 5).

Theorem 2 provides additional insight into the question of how many stable fixed points can be stored in a symmetric threshold-linear network of n neurons. Corollary 1 gave us an upper bound of $\binom{n}{\lfloor n/2 \rfloor}$, as a consequence of Sperner's theorem and theorem 1. As a corollary of theorem 2, we now obtain a lower bound:

Corollary 2. *For each n, there exists a symmetric threshold-linear network of the form 1.1 with $2^{\lfloor n/2 \rfloor}$ stable fixed points.*

To see how this follows from theorem 2, let $m = \lfloor n/2 \rfloor$ and consider the complete m-partite graph G, where each part has two vertices. (Note that two vertices of G have an edge between them if and only if they belong to distinct parts.) If n is odd, let the last vertex connect to all the others. It is easy to see that G has precisely 2^m maximal cliques, obtained by selecting one vertex from each part. By theorem 2, the corresponding network with $W = W(G, \varepsilon, \delta)$ has 2^m stable fixed points.

We end this section with a specific application of theorem 2, illustrating the capability of binary symmetric networks to correct noise-induced errors in combinatorial neural codes.

1.3 Application: Pattern Completion and Error Correction for Place Field Codes. The hippocampus contains a special class of neurons, called place cells, that act as position sensors for tracking the animal's location in space (O'Keefe & Dostrovsky, 1971). Place fields are the spatial receptive fields associated to place cells, and *place field codes* are the corresponding combinatorial neural codes defined from intersections of place fields (Curto, Itskov, Morrison, Roth, & Walker, 2013). In this application of theorem 2, we examine the performance of a binary symmetric network that is designed to perform pattern completion and error correction for place field codes. Our aim is to illustrate how a simple threshold-linear network could function as an effective decoder for a biologically realistic neural code. We do not wish to suggest, however, that the specific networks considered in theorem 2 are in any way accurate models of the hippocampus.

We generate place field codes from a set of circular place fields, $U_1, \ldots, U_n \subset [0, 1]^2$, in a square box environment (see Figure 3B). Each position $p \in [0, 1]^2$ corresponds to a binary codeword, $c = c_1 c_2 \cdots c_n$, where

$$c_i = \begin{cases} 1 & \text{if } p \in U_i \\ 0 & \text{if } p \notin U_i \end{cases}.$$
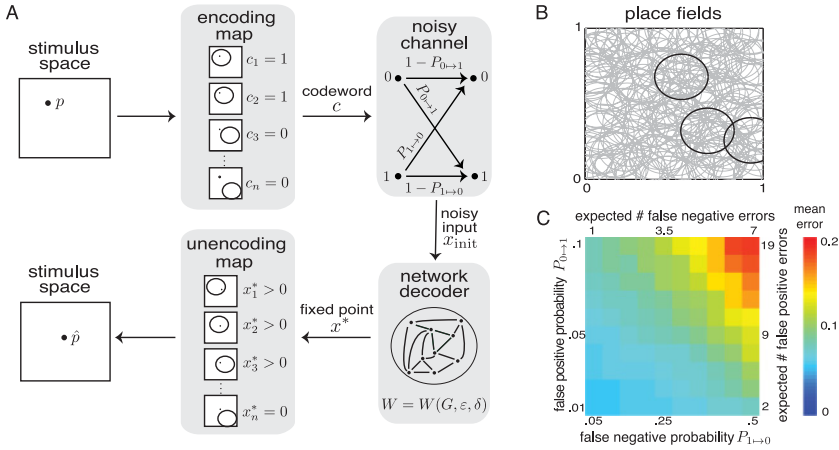
Figure 3: The binary symmetric network decoder has low mean distance error. (A) The sequence of steps in one trial of the binary symmetric network decoder. A point p is chosen uniformly at random from a two-dimensional stimulus space and is passed through an encoding map defined by the set of place fields, resulting in a codeword c. The codeword then goes through a noisy channel, and the corrupted word x_{init} becomes the input (initial condition) for the binary symmetric network decoder. The network evolves to a fixed point x^*, which is then passed to the unencoding map. The final output is an estimated point in the stimulus space, $\hat{p}$. (B) Two hundred place fields covering a 1×1 stimulus space, yielding a place field code of length 200 with an average of 14 neurons (7%) firing per codeword. Three place fields are highlighted in black for clarity. (C) Performance of the code from panel B across a range of noisy channel conditions ($P_{1 \mapsto 0}, P_{0 \mapsto 1}$). For each condition, 1000 trials were performed following the process described in panel A. The distance error was calculated as $\|p - \hat{p}\|$. Colors denote mean distance errors across 1000 trials.

The binary symmetric network corresponding to such a code has n neurons—one for each place field—and assigns the larger weight, $-1 + \varepsilon$, to connections between neurons whose place fields overlap:

$$W_{ij} = \begin{cases} 0 & \text{if } i = j, \\ -1 + \varepsilon & \text{if } U_i \cap U_j \neq \emptyset, \\ -1 - \delta & \text{if } U_i \cap U_j = \emptyset. \end{cases} \qquad (1.3)$$

This is precisely the matrix $W(G, \varepsilon, \delta)$, defined in equation 1.2, where G is the co-firing graph with edges $(ij) \in G$ if there exists a codeword c such that $c_i = c_j = 1$. It is easy to see that the network W can be learned from a simple rule where each connection W_{ij} is initially set to $-1 - \delta$, and then

potentiated to $-1 + \varepsilon$ after presentation of a codeword in which the pair of neurons i and j co-fire.

The above network can be used to correct errors induced by transmitting codewords of the place field code through a noisy channel. Figure 3A shows the basic paradigm of error correction by a network decoder. A point p in the stimulus space (the animal's square box environment) is encoded as a binary codeword via an encoding map given by the place fields $\{U_1, \ldots, U_n\}$. This codeword is then passed through a noisy channel and is received by the network decoder as a noisy initial condition. The network then evolves according to equation 1.1, until it reaches a stable fixed point x^* (see appendix A for further details). From x^* we obtain an estimated point $\hat{p}$ in the stimulus space, given by the mean of the centers of all place fields U_i such that $x_i^* > 0$. The distance error of the network decoder on a single trial is the Euclidean distance $\|p - \hat{p}\|$.

To test the performance of the binary symmetric network decoder, we randomly generated place field codes with 200 neurons and an average of about 7% of neurons firing per codeword (see appendix A for further details). Figure 3B illustrates the coverage of the 1×1 square box environment by 200 place fields. For each code, we computed W according to equation 1.3, with $\varepsilon = 0.25$ and $\delta = 0.5$ to obtain the corresponding network decoder. We then performed 1000 trials for each of 100 different noisy channel conditions. Each noise condition consisted of a false-positive probability $P_{0 \mapsto 1}$, the probability that a 0 is flipped to a 1 in a transmitted codeword, and a false-negative probability $P_{1 \mapsto 0}$, the probability of a $1 \mapsto 0$ flip (see Figure 3A, top right). Figure 3C shows the mean distance errors of the network decoder for a range of noise conditions $(P_{1 \mapsto 0}, P_{0 \mapsto 1})$. Note that because the expected number of 1s in each codeword is ≈ 14 (out of 200 bits), a value of $P_{0 \mapsto 1} = 0.1$ yields nearly 19 expected false-positive errors per codeword, while $P_{1 \mapsto 0} = 0.5$ results in only seven expected false-negative errors per codeword. Most of the considered noise conditions yielded mean distance errors of 0.1 or less. Even for very severe noise conditions with large numbers of expected errors, the mean distance error did not exceed 0.2, or about 20% of the side length of the environment. These results are representative of all place field codes tested.

2 Preliminaries

Here we introduce some notation and review some basic facts about permitted sets and fixed points, including the connection to distance geometry in the symmetric case.

2.1 Notation. Using vector notation, $x = (x_1, \ldots, x_n)^T$, we can rewrite equation 1.1 more compactly as

$$\dot{x} = -x + [Wx + \theta]_+,$$

where $\dot{x}$ denotes the time derivative dx/dt and the nonlinearity $[\ ]_+$ is applied entry-wise. Solutions to equation 1.1 have the property that $x(t) \geq 0$ for all $t > 0$, provided that $x(0) \geq 0$. We use the notation $x \geq 0$ to indicate that every entry of the vector x is nonnegative. The inequalities $>$, $<$, and $\leq$ are similarly applied entry-wise to vectors. If $\sigma = \text{supp}(x)$, then the restriction of x to its support satisfies $x_\sigma > 0$, while $x_k = 0$ for all $k \notin \sigma$.

Many of our results are most conveniently stated in terms of the auxiliary matrix,

$$A = 11^T - I + W, \tag{2.1}$$

where 11^T is the rank one matrix with entries all equal to 1—in other words,

$$W_{ij} = \begin{cases} A_{ij} & \text{for } i = j, \\ -1 + A_{ij} & \text{for } i \neq j. \end{cases}$$

Recall that for any $\sigma \subseteq [n]$, the *principal submatrix* of A obtained by restricting both rows and columns to the index set σ is denoted A_σ. When A_σ is invertible, we define the number

$$a_\sigma \overset{\text{def}}{=} \sum_{i,j \in \sigma} (A_\sigma^{-1})_{ij} = -\frac{\text{cm}(A_\sigma)}{\det(A_\sigma)}, \tag{2.2}$$

where the notation $A_\sigma^{-1} = (A_\sigma)^{-1} \neq (A^{-1})_\sigma$. The second equality is a simple consequence of Cramer's rule and connects a_σ to the Cayley-Menger determinant,

$$\text{cm}(A) \overset{\text{def}}{=} \begin{pmatrix} 0 & 1^T \\ 1 & A \end{pmatrix},$$

defined for any $n \times n$ matrix A. Here 1 denotes the $n \times 1$ column vector of all ones, and 1^T is the corresponding row vector. Note that a_σ has nice scaling properties: if $A_\sigma \mapsto tA_\sigma$, then $a_\sigma \mapsto \frac{1}{t}a_\sigma$.

The Cayley-Menger determinant is well known for its geometric meaning, which we will review below. It also appears in the determinant formula (see Curto, Degeratu et al., 2013, lemma 7),

$$\det(-11^T + A) = \det(A) + \text{cm}(A), \tag{2.3}$$

which holds for any $n \times n$ matrix A. We will make use of this formula in the proof of theorem 3.

2.2 Permitted Sets and Fixed Points. Instead of restricting to a single external drive θ for all neurons, prior work has considered a generalization of the network 1.1 of the form

$$\dot{x} = -x + [Wx + b]_+, \tag{2.4}$$

where $b \in \mathbb{R}^n$ is an arbitrary vector of external drives b_i for each neuron. In this context, Hahnloser, Seung, and collaborators (Xie, Hahnloser, & Seung, 2002; Hahnloser et al., 2003) developed the idea of *permitted sets* of the network, which are the supports of stable fixed points that can arise for some b. The set of all permitted sets of a network of the form 2.4 thus depends only on W and is denoted $\mathcal{P}(W)$. We know from prior work (Hahnloser et al., 2003; Curto, Degeratu, & Itskov, 2012) that σ is a permitted set if and only if $(-I + W)_\sigma$ is a stable matrix.[2] In other words,

$$\mathcal{P}(W) = \{\sigma \subseteq [n] \mid -11^T + A_\sigma \text{ is a stable matrix}\}, \tag{2.5}$$

since $(-I + W)_\sigma = -11^T + A_\sigma$.

In the special case considered here, where $b_i = \theta$ for all $i \in [n]$, it is clear that any stable fixed point x^* of equation 1.1 must have $\operatorname{supp}(x^*) \in \mathcal{P}(W)$. The converse, however, is not true. Although all permitted sets have corresponding fixed points for some $b \in \mathbb{R}^n$, there is no guarantee that such a fixed point exists for $b = \theta$. For example, in the case of symmetric W, it has been shown that $\sigma \in \mathcal{P}(W)$ implies $\tau \in \mathcal{P}(W)$ for all subsets $\tau \subset \sigma$. This property appears to imply that such networks cannot perform pattern completion, since subsets of permitted sets are also permitted. However, theorem 1 tells us that if a permitted set is the support of a stable fixed point of equation 1.1, none of its subsets can be a fixed-point support in the case $b = \theta$, despite being permitted.

2.3 Permitted Sets for Symmetric W. In order to prove theorem 1, we will make heavy use of the geometric theory of permitted sets that was developed in Curto, Degeratu et al. (2013). Here we review some basic facts from classical distance geometry that allow us to geometrically characterize the permitted sets of a network when W is symmetric.

An $n \times n$ matrix A is a *nondegenerate square distance matrix* if there exists a configuration of points $\{p_i\}_{i \in [n]}$ in the Euclidean space $\mathbb{R}^{n-1}$ such that $A_{ij} = ||p_i - p_j||^2$, and the convex hull of the p_is forms a full-dimensional simplex. The Cayley-Menger determinant $\operatorname{cm}(A)$ computes the volume of this simplex and can be used to detect whether a given matrix is nondegenerate square distance. In particular, if A_σ is a nondegenerate square distance matrix with $|\sigma| > 1$, then A_σ is invertible and $a_\sigma > 0$ (Curto, Degeratu

---

[2]A matrix is *stable* if all of its eigenvalues have strictly negative real parts.

et al., 2013, corollary 8). (See Curto, Degeratu et al., 2013, appendix A, for a more complete review of these and other related facts about nondegenerate square distance matrices.)

In the singleton case, $\sigma = \{i\}$ for some $i \in [n]$, the matrix $A_\sigma = [0]$ is always a nondegenerate square distance matrix with $\mathrm{cm}(A_\sigma) = -1 \neq 0$, although $\det(A_\sigma) = 0$. Because of this, it is convenient to declare $a_\sigma = a_{\{i\}} = \infty$ whenever $A_{ii} = 0$. With this convention, we can state the following geometric characterization of permitted sets for symmetric networks, first given in Curto, Degeratu et al. (2013).

Lemma 1 (*Proposition 1; Curto, Degeratu et al., 2013*). *Let W be a symmetric network with zero diagonal, let $A = 11^T - I + W$, as in equation 2.1, and suppose $\sigma \subseteq [n]$ is nonempty. Then $\sigma \in \mathcal{P}(W)$ if and only if A_σ is a nondegenerate square distance matrix with $a_\sigma > 1$. Equivalently, $-11^T + A_\sigma$ is a stable matrix if and only if A_σ is a nondegenerate square distance matrix with $a_\sigma > 1$.*

3 Fixed-Point Conditions

In this section, we derive general fixed-point conditions for networks of the form 1.1, and demonstrate simplifications of these conditions for inhibitory networks and symmetric networks.

3.1 General Fixed-Point Conditions. Consider a fixed point x^* with $\mathrm{supp}(x^*) = \sigma$. The conditions for x^* to be a fixed point of equation 1.1 are given by

$$x^* = [Wx^* + \theta]_+,$$

together with the requirement that $x_i^* > 0$ for all $i \in \sigma$, and $x_k^* = 0$ for all $k \notin \sigma$. For a single neuron $i \in [n]$, the fixed-point equation becomes

$$x_i^* = \left[\sum_{j \in \sigma} W_{ij} x_j^* + \theta \right]_+ = \left[x_i^* - \sum_{j \in \sigma} (11^T)_{ij} x_j^* + \sum_{j \in \sigma} A_{ij} x_j^* + \theta \right]_+$$

$$= \left[x_i^* - m(x^*) + (Ax^*)_i + \theta \right]_+ ,$$

where A is given by equation 2.1 and

$$m(x^*) \overset{\text{def}}{=} \sum_{i=1}^{n} x_i^* = \sum_{i \in \sigma} x_i^*$$

is the total population activity. Separating out the "on" neurons $i \in \sigma$ from the "off" neurons $k \notin \sigma$, we obtain the following lemma:

Lemma 2. *Consider a vector x^*, with $supp(x^*) = \sigma$. Then x^* is a fixed point of equation 1.1 if and only if*

$$(Ax^*)_i = m(x^*) - \theta, \quad \text{for all } i \in \sigma, \text{ and}$$
$$(Ax^*)_k \leq m(x^*) - \theta, \quad \text{for all } k \notin \sigma.$$

In particular,

$$(Ax^*)_k \leq (Ax^*)_i = (Ax^*)_j \quad \text{for all } i, j \in \sigma, k \notin \sigma.$$

As an immediate consequence, we can list conditions that must be satisfied for fixed points consisting of a single active neuron i, with firing rate $x_i^* > 0$. In this case, $m(x^*) = x_i^*$, and so by lemma 2,

$$(Ax^*)_i = A_{ii}x_i^* = x_i^* - \theta, \quad \text{and} \quad (Ax^*)_k = A_{ki}x_i^* \leq A_{ii}x_i^*, \text{ for all } k \neq i.$$

This allows us to solve for $x_i^* = \theta/(1 - A_{ii})$ and to conclude $A_{ki} \leq A_{ii}$. In order for this fixed point to be stable, we must also have $A_{ii} < 1$, and hence $\theta > 0$ because $x_i^* > 0$. We collect these observations in the following proposition, together with the case of empty support, $x^* = 0$.

Proposition 1. *Suppose $|\sigma| \leq 1$. If $\sigma = \{i\}$, then there exists a stable fixed point x^* of equation 1.1 with support σ if and only if the following all hold:*

    i $A_{ii} < 1$ *(equivalently, $W_{ii} < 1$).*
    ii. $\theta > 0$.
    iii. $A_{ki} \leq A_{ii}$ *(equivalently, $W_{ki} \leq -1 + W_{ii}$) for all $k \neq i$.*

Moreover, if this stable fixed point exists, then it is given by $x_k^* = 0$ for all $k \neq i$ and

$$x_i^* = \frac{\theta}{1 - A_{ii}}.$$

Alternatively, if $\sigma = \emptyset$, then $x^* = 0$ is a fixed point of equation 1.1 if and only if $\theta \leq 0$. If $\theta < 0$, then this fixed point is guaranteed to be stable.

Analogous conditions can be obtained for $|\sigma| > 1$, so long as A_σ is not "fine-tuned," as defined below.

Definition 1. *We say that a principal submatrix A_σ of A is fine-tuned if either of the following is true:*

    a. $det(A_\sigma) = 0$.
    b. $\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} = 1$ *for some $k \notin \sigma$.*

Note that the second point depends on entries of the full matrix A, not just entries of A_σ.

We can now state and prove theorem 3:

Theorem 3 (*general fixed-point conditions*). *Consider the threshold-linear network 1.1, and let $A = 11^T - I + W$, as in equation 2.4. Suppose $\sigma \subseteq [n]$ is nonempty and A_σ is not fine-tuned. Then there exists a stable fixed point with support σ if and only if $\theta \neq 0$ and the following three conditions hold:*

   i. *Permitted set condition: $-11^T + A_\sigma$ is a stable matrix.*
   ii. *On-neuron conditions: There are two cases, depending on the sign of θ:*
       a. *$\theta > 0$: either $a_\sigma < 0$ and $A_\sigma^{-1}1_\sigma < 0$, or $a_\sigma > 1$ and $A_\sigma^{-1}1_\sigma > 0$.*
       b. *$\theta < 0$: $0 < a_\sigma < 1$ and $A_\sigma^{-1}1_\sigma > 0$.*
   iii. *Off-neuron conditions: For each $k \notin \sigma$, $\dfrac{\theta}{a_\sigma - 1} \sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} < \dfrac{\theta}{a_\sigma - 1}$.*

Moreover, if a stable fixed point x^* with $\mathrm{supp}(x^*) = \sigma$ exists, then it is given by

$$x_\sigma^* = \frac{\theta}{a_\sigma - 1} A_\sigma^{-1}1_\sigma,$$

with total population activity

$$m(x^*) = \frac{a_\sigma}{a_\sigma - 1}\theta.$$

Remark 1. For fixed points supported on a single neuron, $\sigma = \{i\}$, we may have $A_{ii} = 0$ so that A_σ is not invertible and is thus not covered by the theorem. This case, together with the case $\sigma = \emptyset$, is covered by proposition 1.

Proof of Theorem 3. ($\Rightarrow$) Suppose there exists a stable fixed point x^* with nonempty support, $\mathrm{supp}(x^*) = \sigma$. Then σ is a permitted set and $(-11^T + A)_\sigma$ is a stable matrix (see equation 2.5), so condition i holds. Next, observe that by lemma 2, we have $(Ax^*)_\sigma = A_\sigma x_\sigma^* = (m(x^*) - \theta)1_\sigma$. Since $\det(A_\sigma) \neq 0$ because A_σ is not fine-tuned, we can write

$$x_\sigma^* = (m(x^*) - \theta)A_\sigma^{-1}1_\sigma.$$

Summing the entries of x_σ^* we obtain

$$m(x^*) = (m(x^*) - \theta)a_\sigma. \tag{3.1}$$

Since $m(x^*) > 0$, we must have $a_\sigma \neq 0$. To see that $\theta \neq 0$, note that if $\theta = 0$, then $a_\sigma = 1$ and hence $\det(A_\sigma) + \mathrm{cm}(A_\sigma) = 0$. Using the determinant formula, equation 2.3, we see that this implies $\det(-11^T + A_\sigma) = 0$, which

contradicts the fact that $-11^T + A_\sigma$ is a stable matrix. We can thus conclude that $\theta \neq 0$, which in turn implies that $a_\sigma \neq 1$, using equation 3.1. This allows us to solve for $m(x^*)$ and x^* as

$$m(x^*) = \frac{a_\sigma}{a_\sigma - 1}\theta, \quad \text{and} \quad x_\sigma^* = \frac{\theta}{a_\sigma - 1}A_\sigma^{-1}1_\sigma,$$

yielding the desired equations for x_σ^* and $m(x^*)$.

To show that condition ii holds, we split into two cases depending on the sign of θ.

Case 1: $\theta > 0$. Because $m(x^*) > 0$, we must have $\frac{a_\sigma}{a_\sigma - 1} > 0$, which implies either $a_\sigma < 0$ or $a_\sigma > 1$. Since $x_\sigma^* > 0$, in the $a_\sigma < 0$ case, we must have $A_\sigma^{-1}1_\sigma < 0$, and in the $a_\sigma > 1$ case $A_\sigma^{-1}1_\sigma > 0$.

Case 2: $\theta < 0$. Now $m(x^*) > 0$ implies $\frac{a_\sigma}{a_\sigma - 1} < 0$, which yields $0 < a_\sigma < 1$. Since $x_\sigma^* > 0$, we must have $A_\sigma^{-1}1_\sigma > 0$. Altogether, these observations give condition ii.

Finally, we find that for $k \notin \sigma$,

$$(Ax^*)_k = \sum_{i \in \sigma} A_{ki}x_i^* = \frac{\theta}{a_\sigma - 1}\sum_{i \in \sigma} A_{ki}\sum_{j \in \sigma}(A_\sigma^{-1})_{ij} \leq m(x^*) - \theta = \frac{\theta}{a_\sigma - 1},$$

where the inequality follows because $(Ax^*)_k \leq m(x^*) - \theta$ by lemma 2. Because A_σ is not fine-tuned, this inequality must be strict, yielding condition iii.

($\Leftarrow$) Now, suppose $\theta \neq 0$, and conditions i to iii hold for a given σ, with A_σ not fine-tuned. Consider the ansatz:

$$x_\sigma^* = \frac{\theta}{a_\sigma - 1}A_\sigma^{-1}1_\sigma,$$

with $x_k^* = 0$ for $k \notin \sigma$. By condition i, σ is a permitted set and hence any fixed point with support σ is guaranteed to be stable because A_σ does not satisfy condition b of the definition of *fine-tuned* (this follows from proposition 4 of Curto, Degeratu et al., 2013). It thus suffices to check that $x_\sigma^* > 0$ and that x^* satisfies the fixed point conditions in lemma 2. Since by condition ii we assume in all cases that $\frac{\theta}{a_\sigma - 1}A_\sigma^{-1}1_\sigma > 0$, clearly $x_\sigma^* > 0$. Moreover, it is easy to see that the fixed-point conditions in lemma 2 are satisfied, using our assumption that condition iii holds and the fact that $m(x^*) - \theta = \frac{\theta}{a_\sigma - 1}$.

3.2 Inhibitory Fixed-Point Conditions. In special cases, the fixed-point conditions can be stated in simpler terms than what we had in theorem 3.

Here we consider the case where W is inhibitory, so that $W_{ij} \leq 0$ for all $i, j \in [n]$. In section 3.3, we will see a very similar simplification when W is symmetric.

Theorem 4. *Consider the threshold-linear network 1.1, with W satisfying $-1 \leq W_{ij} \leq 0$ for all $i, j \in [n]$. Let $A = 11^T - I + W$, as in equation 2.1. If $\theta \leq 0$, then $x^* = 0$ is the unique stable fixed point of equation 1.1. If $\sigma \subseteq [n]$ is nonempty and A_σ is not fine-tuned, then there exists a stable fixed point with support σ if and only if $\theta > 0$ and the following three conditions hold:*

    i. Permitted set condition: $-11^T + A_\sigma$ is a stable matrix.
    ii. On-neuron conditions: $A_\sigma^{-1} 1_\sigma > 0$ and $a_\sigma > 1$.
    iii. Off-neuron conditions: $\sum\limits_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} < 1$ for each $k \notin \sigma$.

Proof. If W is inhibitory and $\theta \leq 0$, then the only possible fixed point of equation 1.1 is $x^* = 0$, because the fixed-point equations are $x^* = [Wx^* + \theta]_+$ and $Wx^* + \theta \leq 0$ for any $x^* \geq 0$. By proposition 1, this fixed point is guaranteed to be stable for $\theta < 0$. Since W is inhibitory, however, $x^* = 0$ is also a stable fixed point for $\theta = 0$.

To show the remaining statements, we fix nonempty $\sigma \subseteq [n]$, where A_σ is not fine-tuned and apply theorem 3. It follows from the arguments above that if a stable fixed point x^* with support σ exists, then $\theta > 0$. Condition i follows directly from condition i of theorem 3. Since $\theta > 0$, condition ii.b from theorem 3 does not apply, so it suffices to consider condition ii.a, which splits into $a_\sigma < 0$ and $a_\sigma > 1$ cases. The remainder of this proof consists of showing that the $a_\sigma < 0$ case does not apply, yielding the appropriate specializations of conditions ii and iii above.

Suppose $a_\sigma < 0$. Then conditions ii and iii of theorem 3 cannot be simultaneously satisfied, since condition ii requires $A_\sigma^{-1} 1_\sigma < 0$ and condition iii simplifies to

$$\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} > 1, \quad \text{for each } k \notin \sigma.$$

To see the contradiction, observe that $(A_\sigma^{-1} 1_\sigma)_i = \sum_{j \in \sigma} (A_\sigma^{-1})_{ij} < 0$ by condition ii, and so

$$\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} = \sum_{i \in \sigma} A_{ki} \sum_{j \in \sigma} (A_\sigma^{-1})_{ij} \leq 0,$$

since $A_{ki} = W_{ki} + 1 \geq 0$, by assumption. We can thus eliminate the $a_\sigma < 0$ case from condition ii.a, as only the $a_\sigma > 1$ case can apply. Finally, for $\theta > 0$ and $a_\sigma > 1$, we see that condition iii of theorem 3 simplifies to $\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} < 1$.

Remark 2. Like theorem 3, theorem 4 does not necessarily cover the case of fixed points supported on a single neuron, $\sigma = \{i\}$, when $A_{ii} = 0$. This case is covered by proposition 1.

3.3 Symmetric Fixed-Point Conditions.

As in the inhibitory case above, the fixed-point conditions of theorem 3 can also be stated in simpler terms when W is symmetric with zero diagonal, so that $W_{ij} = W_{ji}$ and $W_{ii} = 0$ for all $i, j \in [n]$. Note that this implies the auxiliary matrix A, given by equation 2.1, is also symmetric with zero diagonal.

Theorem 5. *Consider the threshold-linear network 1.1, for W (and hence A) symmetric with zero diagonal. If $\theta < 0$, then $x^* = 0$ is the unique stable fixed point of equation 1.1. If $\sigma \subseteq [n]$ is nonempty and A_σ is not fine-tuned, then there exists a stable fixed point with support σ if and only if $\theta > 0$ and the following three conditions hold:*

    i. *Permitted set condition:* $-11^T + A_\sigma$ *is a stable matrix.*
    ii. *On-neuron conditions:* $A_\sigma^{-1} 1_\sigma > 0.$
    iii. *Off-neuron conditions:* $\sum\limits_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} < 1$ *for each $k \notin \sigma$.*

Moreover, if a stable fixed point x^* with $\mathrm{supp}(x^*) = \sigma$ exists, then $a_\sigma > 1$.

Proof. The theorem consists of three statements, which we prove in reverse order. First, recall from lemma 1 that if $-11^T + A_\sigma$ is a stable matrix, then $a_\sigma > 1$, yielding the final statement. We now turn to the second statement and show that it is a special case of theorem 3. Note that we can reduce to the $a_\sigma > 1$ case in condition ii of theorem 3, which applies only if $\theta > 0$. Since we must have $a_\sigma > 1$ and $\theta > 0$ in order for the off-neuron conditions to be relevant, we see that condition iii above is the correct simplification of condition iii in theorem 3. Finally, note that for $\theta < 0$, we cannot have a stable fixed point supported on a nonempty σ, by the arguments above. By proposition 1, the fixed point $x^* = 0$ with empty support exists and is stable in this case, giving us the first statement.

Theorem 5 does not cover the singleton case, $\sigma = \{i\}$, since in this case, the matrix A_σ is not invertible (because $A_{ii} = 0$) and is thus fine-tuned. The following variant of theorem 5, which we will use in our proof of theorem 1, does include the singleton case. It is a consequence of theorem 5, lemma 1, proposition 1, and the proof of theorem 3 (in order to drop the fine-tuned hypothesis). Note that proposition 2 is not an "if and only if" statement; the conditions here are only the necessary consequences of the existence of a stable fixed point.

Proposition 6. *Consider the threshold-linear network 1.1, for W (and hence A) symmetric with zero diagonal. Let $\sigma \subseteq [n]$ be nonempty. If there exists a stable fixed point with support σ, then $\theta > 0$ and the following three conditions hold:*

    i. A_σ is a nondegenerate square distance matrix and $a_\sigma > 1$.

   ii. $A_\sigma^{-1} 1_\sigma > 0$ if $|\sigma| > 1$.

  iii. If $|\sigma| > 1$, then $\sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij} \leq 1$ for each $k \notin \sigma$.

       If $\sigma = \{i\}$, then $A_{ki} \leq 0$ for all $k \neq i$.

Proof. Suppose σ supports a stable fixed point. Then σ must be a permitted set, and so condition i follows directly from lemma 1. (Recall our convention from section 2.3 that $a_\sigma = \infty$ if $\sigma = \{i\}$ and $A_{ii} = 0$.) Next, observe that the $\theta > 0$ requirement in theorem 5 holds even if A_σ is fine-tuned, because it required invertibility of A_σ only in the proof of theorem 3. Since, by condition i, A_σ is a nondegenerate square distance matrix, the only instance when it is not invertible is when $|\sigma| = 1$. In this case, however, proposition 1 implies that $\theta > 0$.

For the remaining conditions, we split into two cases: $|\sigma| > 1$ and $|\sigma| = 1$. If $|\sigma| > 1$, then A_σ is invertible, and so $A_\sigma^{-1} 1_\sigma > 0$ (see the proof of theorem 3), yielding condition ii. In the proof of the forward direction of theorem 3, we see that condition iii holds without the strict inequality even when A_σ is fine-tuned, and so the simplified version of this condition in theorem 5 also holds, without the strict inequality, for all A_σ. This gives us the first part of condition iii.

If $\sigma = \{i\}$, proposition 1 provides the on-neuron condition $\theta > 0$, which is automatically satisfied, so there is no further addition to condition ii. It also gives us the off-neuron condition $A_{ki} \leq A_{ii}$ for all $k \neq i$, which is the rest of condition iii, since $A_{ii} = 0$.

4 Some Geometric Lemmas and Proposition 3

In addition to proposition 2, the other main ingredient we will need to prove theorem 1 is the following technical result about nondegenerate square distance matrices:

Proposition 3. *Let A be an $n \times n$ matrix, and let $\tau \subseteq [n]$ with $|\tau| > 1$. If A_τ is a nondegenerate square distance matrix satisfying $A_\tau^{-1} 1_\tau > 0$, then for any $\sigma \subsetneq \tau$ with $|\sigma| > 1$, there exists $k \in \tau \setminus \sigma$ such that $\sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij} > 1$.*

Proposition 3 is key to the proof of theorem 1 because supports of stable fixed points in the symmetric case correspond to nondegenerate square distance matrices A_σ (see lemma 1). Recalling proposition 2, we see that proposition 3 implies an incompatibility between the fixed-point conditions for nested pairs of supports $\sigma, \tau \subseteq [n]$, with $\sigma \subsetneq \tau$. Specifically, if τ satisfies both the permitted set condition i and the on-neuron conditions ii in proposition 2, then proposition 3 implies that the off-neuron conditions iii cannot hold for a proper subset σ.

The remainder of this section is devoted to the proof of proposition 3. Note that the proposition assumes $|\sigma|, |\tau| > 1$, so that the symmetric matrices A_σ and A_τ are invertible. We will maintain this assumption throughout this section.

First, we need four geometric lemmas about nondegenerate square distance matrices. These lemmas rely on observations involving moments of inertia and center of mass for a particular mass configuration associated with the square distance matrix. Let $q_1, \ldots, q_n \in \mathbb{R}^d$ be a configuration of n points, with masses $y_1, \ldots, y_n$ assigned to each point, respectively. Recall that the center of mass of this configuration is given by

$$q_{\mathrm{cm}} = \frac{1}{m(y)} \sum_{i=1}^n y_i q_i,$$

where $m(y) = \sum_{i=1}^n y_i$ is the total mass. The moment of inertia about any point $q \in \mathbb{R}^d$ is

$$I_q(y) = \sum_{i=1}^n \|q - q_i\|^2 y_i.$$

The classical parallel axis theorem states that the moment of inertia depends on the distance between the chosen point q and the center of mass q_{cm}.

Proposition 4 (*parallel axis theorem*). *Let $y_1, \ldots, y_n \in \mathbb{R}$ denote the masses assigned to the points $q_1, \ldots, q_n \in \mathbb{R}^d$. For any $q \in \mathbb{R}^d$,*

$$I_q(y) = m(y)\|q - q_{\mathrm{cm}}\|^2 + I_{q_{\mathrm{cm}}}(y).$$

This result is well known and is a staple of undergraduate physics. It has also been called Appolonius's formula in Euclidean geometry (Berger, 1987). For completeness, we provide a proof in appendix B as a corollary of a novel and more general result. From our proof, it is clear that proposition 4 is valid for general dimension d and general mass assignments, including negative masses.

We will exploit proposition 4 by applying it to point configurations corresponding to nondegenerate square distance matrices, A_σ. Let $\{q_i\}_{i\in\sigma}$ be a representing point configuration of A_σ in $\mathbb{R}^{|\sigma|-1}$, so that

$$A_{ij} = \|q_i - q_j\|^2 \quad \text{for all } i, j \in \sigma.$$

For this point configuration, let $\{y_i\}_{i\in\sigma}$ be an assignment of masses to each point (not necessarily nonnegative), given by

$$y_i = (A_\sigma^{-1} 1_\sigma)_i \quad \text{for each } i \in \sigma.$$

Recall that because A_σ is a nondegenerate square distance matrix and $|\sigma| > 1$, A_σ is invertible, and hence these masses are always well defined. The total mass is

$$m(y) = \sum_{i \in \sigma} y_i = \sum_{i,j \in \sigma} (A_\sigma^{-1})_{ij} = a_\sigma.$$

Finally, note that because $\{q_i\}_{i \in \sigma}$ is a nondegenerate point configuration, there exists a unique equidistant point, $p_\sigma \in \mathbb{R}^{|\sigma|-1}$, so that the distances $\|q_i - p_\sigma\|$ are the same for all $i \in \sigma$ (Berger, 1987). The following lemma shows that the equidistant point coincides with the center of mass for this mass configuration:

Lemma 3. *Let A_σ be a nondegenerate square distance matrix, and assign masses $y_\sigma = A_\sigma^{-1} 1_\sigma$ to a representing point configuration $\{q_i\}_{i \in \sigma}$, as described above. Then $q_{\mathrm{cm}} = p_\sigma$.*

Proof. First, observe that the moment of inertia about any of the points q_i is given by

$$I_{q_i}(y) = \sum_{j \in \sigma} \|q_i - q_j\|^2 y_j = (A_\sigma y_\sigma)_i = (A_\sigma A_\sigma^{-1} 1_\sigma)_i = 1, \quad \text{for each } i \in \sigma.$$

On the other hand, the parallel axis theorem tells us that $I_{q_i}(y) = m(y)\|q_i - q_{\mathrm{cm}}\|^2 + I_{q_{\mathrm{cm}}}(y)$, so we must have that $\|q_i - q_{\mathrm{cm}}\|^2 = \|q_j - q_{\mathrm{cm}}\|^2$ for all $i, j \in \sigma$. Clearly, these equations imply that $q_{\mathrm{cm}} = p_\sigma$.

For a given point configuration $\{q_i\}_{i \in \sigma}$, we denote the interior of the convex hull by

$$(\mathrm{conv}\{q_i\}_{i \in \sigma})^o.$$

The next lemma states that the equidistant point p_σ is inside the convex hull of its corresponding point configuration if and only all entries of $A_\sigma^{-1} 1_\sigma$ are strictly positive.

Lemma 4. *Let A_σ be a nondegenerate square distance matrix. Then $A_\sigma^{-1} 1_\sigma > 0$ if and only if $p_\sigma \in (\mathrm{conv}\{q_i\}_{i \in \sigma})^o$ for any representing point configuration $\{q_i\}_{i \in \sigma}$.*

Proof. Let $\{q_i\}_{i \in \sigma}$ be a representing point configuration for A_σ, and assign masses $y_\sigma = A_\sigma^{-1} 1_\sigma$, as in lemma 3, so that $q_{\mathrm{cm}} = p_\sigma$. Next, observe that $q_{\mathrm{cm}} \in \mathrm{conv}\{q_i\}_{i \in \sigma}$ if and only if the masses are all nonnegative. Similarly, $q_{\mathrm{cm}} \in (\mathrm{conv}\{q_i\}_{i \in \sigma})^o$ if and only if the masses are all strictly positive. It thus follows that $p_\sigma \in (\mathrm{conv}\{q_i\}_{i \in \sigma})^o$ if and only if $A_\sigma^{-1} 1_\sigma > 0$.

Lemma 5. *Let A_τ be a nondegenerate square distance matrix with representing point configuration $\{q_i\}_{i\in\tau}$. Suppose $\sigma \subsetneq \tau$, with $|\sigma| > 1$, and let p_σ denote the unique equidistant point in the affine subspace defined by the points $\{q_i\}_{i\in\sigma}$. Then for any $k \in \tau \setminus \sigma$, we have $\sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij} \leq 1$ if and only if $\|q_k - p_\sigma\| \leq \|q_i - p_\sigma\|$, for $i \in \sigma$.*

Proof. First, assign masses $y_\sigma = A_\sigma^{-1}1_\sigma$ to the subset of points $\{q_i\}_{i\in\sigma}$, and observe that

$$\sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij} = \sum_{i\in\sigma} A_{ki}(A_\sigma^{-1}1_\sigma)_i = \sum_{i\in\sigma} \|q_k - q_i\|^2 y_i = I_{q_k}(y).$$

Next, recall from the proof of lemma 3 that for each $i \in \sigma$, $I_{q_i}(y) = 1$. Using proposition 4, we see that $I_{q_k}(y) \leq 1$ if and only if $\|q_k - q_{cm}\| \leq \|q_i - q_{cm}\|$. On the other hand, lemma 3 tells us that $q_{cm} = p_\sigma$, yielding the desired result.

By definition of p_σ, the distances $\|q_i - p_\sigma\|$ for $i \in \sigma$ in the statement of lemma 5 are all equal. This distance to the equidistant point, $\rho_\sigma = \|q_i - p_\sigma\|$, has in fact a well-known formula in terms of the matrix A_σ (see Curto, Degeratu et al., 2013, appendix C):

$$\rho_\sigma^2 = -\frac{1}{2}\frac{\det(A_\sigma)}{cm(A_\sigma)} = \frac{1}{2a_\sigma}. \tag{4.1}$$

To state the next lemma, we define

$$B_{\rho_\sigma}(p_\sigma) \stackrel{\text{def}}{=} \{q \mid \|q - p_\sigma\| \leq \rho_\sigma\},$$

which is the closed ball of radius ρ_σ centered at p_σ.

Lemma 6. *Let A_τ be a nondegenerate square distance matrix with representing point configuration $\{q_i\}_{i\in\tau}$. If $p_\tau \in (conv\{q_i\}_{i\in\tau})^o$ and $\sigma \subsetneq \tau$ with $|\sigma| > 1$, then there exists $k \in \tau \setminus \sigma$ such that $q_k \notin B_{\rho_\sigma}(p_\sigma)$.*

Proof. Note that since A_τ is a nondegenerate square distance matrix, so is A_σ. Let p_τ and p_σ be the corresponding unique equidistant points, both embedded in $\mathbb{R}^{|\tau|-1}$ (so p_σ lies in the affine subspace spanned by $\{q_i\}_{i\in\sigma}$). Let H_σ denote the hyperplane in $\mathbb{R}^{|\tau|-1}$ that contains p_σ and is perpendicular to the line $\overline{p_\sigma p_\tau}$ (see Figure 4). Note that because p_τ is also equidistant to all $\{q_i\}_{i\in\sigma}$, the line $\overline{p_\sigma p_\tau}$ is perpendicular to the affine subspace defined by $\{q_i\}_{i\in\sigma}$; hence, H_σ contains all points $\{q_i\}_{i\in\sigma}$. Denote by H_σ^+ the closed half-space that contains p_τ, and let H_σ^- denote the opposite half-space, so that $H_\sigma^+ \cap H_\sigma^- = H_\sigma$. Because $p_\tau \in H_\sigma^+$ and $p_\tau \in (conv\{q_i\}_{i\in\tau})^o$, we must have at least one $k \in \tau \setminus \sigma$ with $q_k \in H_\sigma^+ \setminus H_\sigma$. This implies $q_k \notin B_{\rho_\sigma}(p_\sigma)$, since the
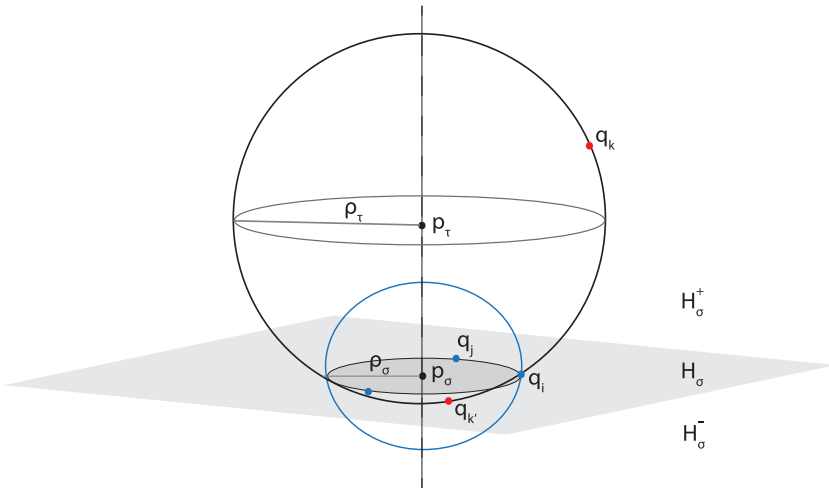
Figure 4: Picture for the proof of lemma 6.

only way a point can lie both on the large sphere of radius ρ_τ and in the ball $B_{\rho_\sigma}(p_\sigma)$ is if it lies in H_σ^- (see $q_{k'}$ in Figure 4).

Finally, we are ready to prove proposition 3.

Proof of Proposition 3. Suppose A_τ is a nondegenerate square distance matrix, satisfying $A_\tau^{-1}1_\tau > 0$. Let $\{q_i\}_{i \in \tau}$ be a representing point configuration. By lemma 4, we know that $p_\tau \in (\text{conv}\{q_i\}_{i \in \tau})^o$. Now applying lemma 6, we find there exists $k \in \tau \setminus \sigma$ such that $q_k \notin B_{\rho_\sigma}(p_\sigma)$. This means $\|q_k - p_\sigma\| > \|q_i - p_\sigma\|$ for $i \in \sigma$, which in turn implies that $\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} > 1$, by lemma 5.

5 Proof of Theorems 1 and 2

In this section we prove theorems 1 and 2. Following these proofs, we state and prove a related result, theorem 7.

The proof of theorem 1 follows from propositions 2 and 3, together with a simple lemma (below). Recall from section 2.2 that for a given connectivity matrix W, $\mathcal{P}(W)$ denotes the set of all permitted sets of the corresponding threshold-linear network. We will use the notation $\mathcal{P}_{\max}(W)$ to denote the maximal permitted sets with respect to inclusion. In other words, if $\sigma \in \mathcal{P}_{\max}(W)$, then $\tau \notin \mathcal{P}(W)$ for any $\tau \supsetneq \sigma$.

Lemma 7. *Consider the threshold-linear network 1.1, for W symmetric with zero diagonal. If there exists a stable fixed point x^* with support $\sigma = \{i\}$, then $\sigma \in \mathcal{P}_{\max}(W)$.*

Proof. Suppose $\sigma = \{i\}$ is not a maximal permitted set, so that $\sigma \subsetneq \tau$ for some $\tau \in \mathcal{P}(W)$. Since A_τ must be a nondegenerate square distance matrix and $|\tau| \geq 2$, it follows that $A_{ki} > 0$ for all $k \in \tau \setminus \{i\}$. But this contradicts condition iii of proposition 2, so we can conclude that $\sigma \in \mathcal{P}_{\max}(W)$.

Recall that proposition 3 implies that if τ satisfies conditions i and ii of proposition 2, then no $\sigma \subsetneq \tau$ with $|\sigma| > 1$ can possibly satisfy condition iii. Lemma 7 allows us to extend this observation to the case of $|\sigma| = 1$, as singletons can support stable fixed points only if they are maximal permitted sets. Proposition 2 also tells us that the existence of a stable fixed point with nonempty support τ implies $\theta > 0$, ruling out the possibility of the stable fixed point with empty support, $x^* = 0$. Thus, the above consequence of proposition 3 can be extended to all $\sigma \subsetneq \tau$. We are now ready to prove theorem 1.

Proof of Theorem 1. Suppose there exists a stable fixed point with support τ. We will first show that for $\sigma \subsetneq \tau$, σ cannot support a stable fixed point. We may assume τ is nonempty (otherwise it has no proper subsets). It follows from proposition 2 that $\theta > 0$. We consider three cases: $\sigma = \emptyset$, $|\sigma| = 1$, and $|\sigma| \geq 2$.

Suppose $\sigma = \emptyset$. Since $\theta > 0$, we know that σ cannot support a stable fixed point, by proposition 1. Next, suppose $|\sigma| = 1$. Here we can also conclude that σ cannot support a stable fixed point, by lemma 7. (Note that $\sigma \notin \mathcal{P}_{\max}(W)$ because $\sigma \subsetneq \tau$, and we must have $\tau \in \mathcal{P}(W)$.) Finally, suppose $|\sigma| \geq 2$. Since both $|\sigma|, |\tau| > 1$, we can apply proposition 3. All hypotheses are satisfied because τ is the nonempty support of a stable fixed point, and so by proposition 2, we know that A_τ is a nondegenerate square distance matrix with $A_\tau^{-1}1_\tau > 0$. It thus follows from proposition 3 that σ cannot satisfy condition iii of proposition 2, so σ cannot be a stable fixed point support.

Finally, observe that for any proper superset $\sigma \supsetneq \tau$, if σ is the support of a stable fixed point, then the above logic implies that τ cannot support a stable fixed point, contradicting the hypothesis. Thus, there is no stable fixed point with support σ for any $\sigma \subsetneq \tau$ or $\sigma \supsetneq \tau$.

We now turn to the proof of theorem 2. We will use the following simple lemma:

Lemma 8. *Let G be a simple graph, and consider the binary symmetric network $W = W(G, \varepsilon, \delta)$, for some $0 < \varepsilon < 1$ and $\delta > 0$, as in equation 1.2. Let $A = 11^T - I + W$, as in equation 2.1. If σ is a clique of G with $|\sigma| > 1$, then*

    i. $-11^T + A_\sigma$ is stable, and
    ii. $A_\sigma^{-1}1_\sigma = \frac{1}{\varepsilon(|\sigma|-1)}1_\sigma$ (and thus $a_\sigma = \frac{|\sigma|}{\varepsilon(|\sigma|-1)}$).

Proof. We first prove condition i, and then condition ii.

Condition i. Observe that if σ is a clique and $|\sigma| > 1$, then $-11^T + A_\sigma$ is a matrix with all diagonal entries equal to -1 and all off-diagonal entries equal to $-1 + \varepsilon$. Thus, $-11^T + A_\sigma = (-1 + \varepsilon)11^T - \varepsilon I_\sigma$. The term $(-1 + \varepsilon)11^T$ is a rank one matrix, whose only nonzero eigenvalue is $|\sigma|(-1 + \varepsilon)$, corresponding to the eigenvector 1_σ. Thus, the eigenvalues of $-11^T + A_\sigma$ are $|\sigma|(-1 + \varepsilon) - \varepsilon$ and $-\varepsilon$. The eigenvalue $|\sigma|(-1 + \varepsilon) - \varepsilon$ is negative precisely when $\varepsilon < |\sigma|/(|\sigma| - 1)$. This always holds since $|\sigma| > 1$ and $\varepsilon < 1$. We conclude that all eigenvalues of $-11^T + A_\sigma$ are negative, and hence $-11^T + A_\sigma$ is stable.

Condition ii. Notice that since σ is a clique, $A_\sigma = \varepsilon(11^T - I)_\sigma$. It is easy to check that A_σ^{-1} satisfies,

$$
A_\sigma^{-1} = \frac{1}{\varepsilon(|\sigma| - 1)}
\begin{bmatrix}
-|\sigma| + 2 & 1 & 1 & \cdots & 1 \\
1 & -|\sigma| + 2 & 1 & \cdots & 1 \\
1 & 1 & -|\sigma| + 2 & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & 1 & \cdots & -|\sigma| + 2
\end{bmatrix},
$$

from which it immediately follows that $A_\sigma^{-1}1_\sigma = \frac{1}{\varepsilon(|\sigma|-1)}1_\sigma$.

Proof of Theorem 2. First, we show that non-cliques cannot support stable fixed points because they are not permitted sets. Then we show that all maximal cliques do support stable fixed points and derive the corresponding expression for x_σ^*. Finally, we invoke theorem 1 to conclude that there can be no other stable fixed points.

Let $\sigma \subseteq [n]$ be a subset that is not a clique in G. In the case $\sigma = \emptyset$, we know from proposition 1 that there is no $x^* = 0$ fixed point because $\theta > 0$. Suppose, then, that σ is nonempty. Since single vertices are always cliques, we know that $|\sigma| \geq 2$ and there exists at least one pair $i, j \in \sigma$ such that $(ij) \notin G$. The corresponding principal submatrix $(-11^T + A_\sigma)_{\{ij\}}$ is given by $\begin{bmatrix} -1 & -1-\delta \\ -1-\delta & -1 \end{bmatrix}$. Since both the determinant and trace of this submatrix are negative, it must have a positive eigenvalue and is thus unstable. By the Cauchy interlacing theorem (e.g., Horn & Johnson, 1985; Curto, Degeratu, & Itskov, 2013, appendix A), $-11^T + A_\sigma$ must also have a positive eigenvalue, rendering it unstable. It follows from equation 2.5 that σ is not a permitted set, so there is no stable fixed point with support σ.

Next, we show that all maximal cliques support stable fixed points. We split into two cases: $|\sigma| = 1$ and $|\sigma| > 1$. Let $\sigma = \{i\}$ be a maximal clique consisting of a single vertex, i. Since $A_{ii} = 0 < 1$ and $\theta > 0$, conditions i and ii of proposition 1 are always satisfied. Since $\{i\}$ is a maximal clique, $(ik) \notin G$ for all $k \neq i$, and so $A_{ki} = -\delta < 0 = A_{ii}$ for all $k \neq i$. Thus, condition iii of proposition 1 is also satisfied. It follows that $\sigma = \{i\}$ supports a stable fixed

point, given by $x_i^* = \theta$ and $x_k^* = 0$ for all $k \neq i$, in agreement with the desired formula for x_σ^* with $|\sigma| = 1$.

Now suppose σ is a maximal clique with $|\sigma| > 1$. Note that A_σ is not fine-tuned, so theorem 5 applies. By lemma 8, $-11^T + A_\sigma$ is stable and $A_\sigma^{-1} 1_\sigma = \frac{1}{\varepsilon(|\sigma|-1)} 1_\sigma > 0$. Thus, conditions i and ii of theorem 5 are satisfied, and $\theta > 0$ by assumption. It remains only to show that condition iii holds. Since σ is a maximal clique, for all $k \notin \sigma$ there exists some $i_k \in \sigma$ such that $(i_k k) \notin G$, and thus $A_{ki_k} = -\delta$. We obtain:

$$\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} = \sum_{i \in \sigma} A_{ki} \sum_{j \in \sigma} (A_\sigma^{-1})_{ij} = \frac{1}{\varepsilon(|\sigma|-1)} \sum_{i \in \sigma} A_{ki}$$

$$= \frac{-\delta + \sum_{i \in \sigma \setminus \{i_k\}} A_{ki}}{\varepsilon(|\sigma|-1)},$$

where we have used part ii of lemma 8 to evaluate the row sums of A_σ^{-1}. Since $A_{ki} \leq \varepsilon$ for each $i \in \sigma$, it follows that $\sum_{i \in \sigma \setminus \{i_k\}} A_{ki} \leq \varepsilon(|\sigma|-1)$, and so condition iii is satisfied:

$$\sum_{i,j \in \sigma} A_{ki}(A_\sigma^{-1})_{ij} \leq \frac{-\delta}{\varepsilon(|\sigma|-1)} + 1 < 1.$$

We conclude that there exists a stable fixed point for each maximal clique σ. Using the formula for x_σ^* from theorem 3, together with the expressions for $A_\sigma^{-1} 1_\sigma$ and a_σ from lemma 8, we obtain the desired equation for the fixed point:

$$x_\sigma^* = \frac{\theta}{a_\sigma - 1} A_\sigma^{-1} 1_\sigma = \frac{\theta}{(1-\varepsilon)|\sigma| + \varepsilon} 1_\sigma.$$

Finally, observe that since any non-maximal clique is necessarily a proper subset of a maximal clique, theorem 1 guarantees that non-maximal cliques cannot support stable fixed points. Thus, the supports of stable fixed points correspond precisely to maximal cliques in G.

In theorem 2, we saw that all fixed-point supports corresponded to maximal permitted sets, as these were the maximal cliques of the underlying graph G. This situation is ideal for pattern completion, as it guarantees that only maximal patterns can be returned as outputs of the network. Our final result shows that this phenomenon generalizes to a broader class of symmetric networks, provided all maximal permitted sets $\tau \in \mathcal{P}_{\max}(W)$ satisfy the on-neuron conditions, $A_\tau^{-1} 1_\tau > 0$, from theorem 5.

Theorem 7. *Let W be symmetric with zero diagonal, and suppose that $\theta > 0$ and $A_\tau^{-1} 1_\tau > 0$ for each $\tau \in \mathcal{P}_{\max}(W)$ with $|\tau| > 1$. If x^* is a stable fixed point of equation 1.1, then $\text{supp}(x^*) \in \mathcal{P}_{\max}(W)$.*

Proof. Let $\sigma = \text{supp}(x^*)$ be the support of a stable fixed point x^*. Observe that $\sigma \neq \emptyset$, since proposition 1 states that $x^* = 0$ can be a fixed point only when $\theta \leq 0$. If $\sigma = \{i\}$ for some $i \in [n]$, then by lemma 7, we have that $\{i\} \in \mathcal{P}_{\max}(W)$.

Next, consider $|\sigma| > 1$. Since $\sigma \in \mathcal{P}(W)$, there exists some $\tau \in \mathcal{P}_{\max}(W)$ such that $\sigma \subseteq \tau$. Since τ is a permitted set, A_τ is a nondegenerate square distance matrix (see lemma 1). By hypothesis, $A_\tau^{-1} 1_\tau > 0$, so proposition 3 applies. If $\sigma \subsetneq \tau$ is a proper subset, then there exists $k \in \tau \setminus \sigma$ that violates condition iii of theorem 5, contradicting the assumption that σ is a stable fixed point support. It follows that $\sigma = \tau$, and thus $\sigma \in \mathcal{P}_{\max}(W)$.

Appendix A: Additional Details for the Simulations in Section 1.3 _____

A.1 Network Implementation. We solved the system of differential equations 1.1, with $\theta = 1$, using a standard Matlab ode solver for a length of time corresponding to $50\tau_L$, where τ_L is the leak time constant associated with each neuron. This length of time was sufficient for the network to numerically stabilize at a fixed point x^*. Note that τ_L is omitted from our equations because we have set $\tau_L = 1$, so that time is measured in units of τ_L.

A.2 Generation of 2D Place Field Codes. Two-dimensional place field codes were generated following the same methods as in Curto, Itskov et al. (2013). Two hundred place field centers were randomly chosen from a 1×1 square box environment, with each place field a disk of radius 0.15. This produced place field codes with an average of 7% of neurons firing per codeword. As described in Curto, Itskov et al. (2013), the place field centers were initially chosen randomly from uncovered regions of the stimulus space until complete coverage was achieved. The remaining place field centers were then chosen uniformly at random from the full space. Here we introduced one modification to the procedure in Curto, Itskov et al. (2013): our place field centers were generated 50 at a time, repeating the process from the beginning for the four sets of 50 neurons in order to guarantee that every point in the stimulus space was covered by a minimum of four place fields. This ensured that all codewords had at least four 1s (out of 200 bits).

Appendix B: A Generalization of the Parallel Axis Theorem _____

Here we present a new, more general version of the parallel axis theorem (see proposition 4), which was used in section 4.

Proposition 5. *Let A be an $n \times n$ matrix and $y \in \mathbb{R}^n$ a vector such that $m = m(y) = \sum_{i=1}^n y_i \neq 0$. Then there exists a unique $\lambda = (\lambda_1, \dots, \lambda_n)$ such that*

$$Ay = \Lambda y, \quad \text{where} \quad \Lambda_{ij} = \lambda_i + \lambda_j.$$

Explicitly,

$$\lambda_i = \frac{(Ay)_i}{m} - \frac{y \cdot (Ay)}{2m^2}. \tag{B.1}$$

Proof. First, observe that $Ay = \Lambda y$ implies

$$(Ay)_i = (\Lambda y)_i = \sum_j (\lambda_i + \lambda_j) y_j = m\lambda_i + \lambda \cdot y.$$

Since $m \neq 0$, $\lambda_i = \frac{(Ay)_i - \lambda \cdot y}{m}$. Taking the dot product with y yields

$$\lambda \cdot y = \frac{1}{m} y \cdot (Ay) - \frac{1}{m}(\lambda \cdot y) \sum_i y_i = \frac{1}{m} y \cdot (Ay) - \lambda \cdot y,$$

allowing us to solve for $\lambda \cdot y = \frac{1}{2m} y \cdot (Ay)$. Plugging this into the above expression for λ_i yields the desired result.

We can now obtain the classical parallel axis theorem as a special case of proposition 5. Consider the case where A is the square distance matrix for a configuration of points $q_1, \ldots, q_n \in \mathbb{R}^d$, and $y = (y_1, \ldots, y_n)^T$ is a vector of masses, one for each point, whose sum $m = \sum_{i=1}^n y_i$ is nonzero. In this case, $A_{ij} = \|q_i - q_j\|^2$, the center of mass is $q_{cm} = \frac{1}{m} \sum_{i=1}^n q_i y_i$, and it is not difficult to check that

$$\lambda_i = \frac{(Ay)_i}{m} - \frac{y \cdot (Ay)}{2m^2} = \|q_i - q_{cm}\|^2.$$

(Without loss of generality, choose $q_{cm} = 0$ and use this fact to cancel terms of the form $\sum_i q_i y_i$ that appear when you rewrite $A_{ij} = (q_i - q_j) \cdot (q_i - q_j)$ and expand.) Recall that the moment of inertia of such a mass configuration about a point $q \in \mathbb{R}^d$ is given by $I_q(y) = \sum_{j=1}^n \|q - q_j\|^2 y_j$. If $q = q_i$ for some $i \in [n]$, then we have

$$I_{q_i}(y) = \sum_{j=1}^n \|q_i - q_j\|^2 y_j = (Ay)_i.$$

We can now prove proposition 4, which states that

$$I_q(y) = m\|q - q_{cm}\|^2 + I_{q_{cm}}(y).$$

Proof of Proposition 4. Without loss of generality, we can assume that $q = q_i$ for some $i \in [n]$. (If not, add the point q to the collection $\{q_i\}$ and assign it a mass of 0.) As observed above, $I_q(y) = I_{q_i}(y) = (Ay)_i$, and recall from the proof of proposition 5 that $(Ay)_i = m\lambda_i + \lambda \cdot y$. Since $\lambda_i = \|q_i - q_{cm}\|^2$, and thus $\lambda \cdot y = I_{q_{cm}}(y)$, substituting these expressions into the equation for $(Ay)_i$ immediately yields the desired result.

To see why proposition 5 may be useful more generally, consider the situation where $y \in \mathbb{R}^n$ is proportional to the vector of firing rates at a fixed point of a threshold-linear network 1.1. Specifically, suppose y has support σ, and $y_\sigma = A_\sigma^{-1}1_\sigma$. In this case, $m = \sum_{i\in\sigma}(A_\sigma^{-1}1_\sigma)_i = a_\sigma$, $(Ay)_i = 1$ for all $i \in \sigma$, and $y \cdot (Ay) = \sum_{i\in\sigma} y_i = m = a_\sigma$. It follows that

$$\lambda_i = \frac{1}{2a_\sigma} \quad \text{for all} \quad i \in \sigma.$$

On the other hand, for $k \notin \sigma$, we have $(Ay)_k = \sum_{i\in\sigma} A_{ki}y_i = \sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij}$, so that

$$\lambda_k = \frac{1}{a_\sigma}\left(\sum_{i,j\in\sigma} A_{ki}(A_\sigma^{-1})_{ij} - \frac{1}{2}\right) \quad \text{for all} \quad k \notin \sigma.$$

Assuming $a_\sigma > 0$, the off-neuron conditions of theorems 4 and 5 are satisfied if and only if $\lambda_k < \frac{1}{2a_\sigma} = \lambda_i$. Thus, λ_i for $i \in \sigma$ generalizes the quantity

$$\|q_i - q_{cm}\|^2 = \|q_i - p_\sigma\|^2 = \rho_\sigma^2 = \frac{1}{2a_\sigma},$$

which appeared in section 4 for A a nondegenerate square distance matrix (see equation 4.1). Similarly, for $k \notin \sigma$, the condition $\lambda_k < \lambda_i$ generalizes the requirement $\|q_k - p_\sigma\| \le \|q_i - p_\sigma\|$ from lemma 5. This suggests that it may be possible to generalize proposition 3, and thus theorem 1, beyond the symmetric case.

References

Anderson, I. (1987). *Combinatorics of finite sets*. New York: Oxford University Press.
Berger, M. (1987). *Geometry I*. New York: Springer Science & Business Media.

Curto, C., Degeratu, A., & Itskov, V. (2012). Flexible memory networks. *Bull. Math. Biol.*, *74*(3), 590–614.

Curto, C., Degeratu, A., & Itskov, V. (2013). Encoding binary neural codes in networks of threshold-linear neurons. *Neural Comput.*, *25*, 2858–2903.

Curto, C., Itskov, V., Morrison, K., Roth, Z., & Walker, J. L. (2013). Combinatorial neural codes from a mathematical coding theory perspective. *Neural Comput.*, *25*(7), 1891–1925.

Hahnloser, R. H., Seung, H., & Slotine, J. (2003). Permitted and forbidden sets in symmetric threshold-linear networks. *Neural Comput.*, *15*(3), 621–638.

Hopfield, J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.*, *79*(8), 2554–2558.

Hopfield, J. (1984). Neurons with graded response have collective computational properties like those of two-sate neurons. *Proc. Natl. Acad. Sci.*, *81*, 3088–3092.

Horn, R., & Johnson, C. (1985). *Matrix analysis*. Cambridge: Cambridge University Press.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, *34*(1), 171–175.

Seung, H., & Yuste, R. (2012). Appendix E: Neural networks. In E. Kandel, J. Schwartz, T. Jessell, S. Siegelbaum, & A. Hudspeth (Eds.), *Principles of Neural Science* (5th ed., pp. 1581–1600). New York: McGraw-Hill Education/Medical.

Simmen, M., Treves, A., & Rolls, E. (1996). Pattern retrieval in threshold-linear associative nets. *Network-Comp. Neural*, *7*, 109–122.

Tsodyks, M., & Sejnowski, T. (1995). Associative memory and hippocampal place cells. *Int. J. Neural Syst.*, *6*, 81–86.

Vogels, T., Rajan, K., & Abbott, L. (2005). Neural network dynamics. *Annu. Rev. Neurosci*, *28*, 357–376.

Xie, X., Hahnloser, R. H., & Seung, H. (2002). Selectively grouping neurons in recurrent networks of lateral inhibition. *Neural Comput.*, *14*, 2627–2646.