Bull Math Biol (2012) 74:590–614
DOI 10.1007/s11538-011-9678-9

ORIGINAL ARTICLE

Flexible Memory Networks

Carina Curto · Anda Degeratu · Vladimir Itskov

Received: 7 November 2010 / Accepted: 11 July 2011 / Published online: 9 August 2011
© Society for Mathematical Biology 2011

Abstract Networks of neurons in some brain areas are flexible enough to encode new memories quickly. Using a standard firing rate model of recurrent networks, we develop a theory of flexible memory networks. Our main results characterize networks having the maximal number of flexible memory patterns, given a constraint graph on the network's connectivity matrix. Modulo a mild topological condition, we find a close connection between maximally flexible networks and rank 1 matrices. The topological condition is $H_1(X;\mathbb{Z})=0$, where X is the clique complex associated to the network's constraint graph; this condition is generically satisfied for large random networks that are not overly sparse. In order to prove our main results, we develop some matrix-theoretic tools and present them in a self-contained section independent of the neuroscience context.

Keywords Neural networks · Memory · Learning · Clique complex · Low rank

1 Introduction

New memories in some brain areas can be encoded quickly (Rutishauser and Mamelak 2006). It is widely believed that memories are stored via changes in the synaptic efficacies between neurons. Irrespective of the plasticity mechanism, or "learning rule," used to encode memory patterns, rapid learning is perhaps most easily accomplished if new patterns can be learned via only small changes in connection strengths between neurons. It may thus be desirable for fast-learning, flexible networks to have

C. Curto (✉) · V. Itskov
Department of Mathematics, University of Nebraska-Lincoln, Lincoln, NE, USA
e-mail: ccurto2@math.unl.edu

A. Degeratu
Max Planck Institute for Gravitational Physics, Golm, Germany

architectures that enable many memory patterns to be encoded (and unencoded) by only small perturbations of the synaptic connections. Here, by "architecture," we mean the pattern of synaptic strengths, or *weights*, assigned to directed connections between neurons. Which network architectures allow maximal flexibility for learning and unlearning new memories?

We study this question in the context of a standard firing rate model of recurrent neural networks. Building on the framework of "permitted" and "forbidden" sets first introduced in Hahnloser et al. (2003), we think of the recurrent network as a gating device that allows only a restricted set of patterns, the stored memories, to be activated by external feed-forward input. In Theorem 1.2, we establish a correspondence between the memory patterns encoded by a recurrent network and the set of stable principal submatrices of the network's effective connectivity matrix. We then make precise the notion of memory patterns that are flexible in the sense that they can be encoded (learned) and unencoded (forgotten) via only small changes to the network weights. Our main results, Theorems 1.5, 1.7, and 1.8, characterize network architectures with the maximal number of flexible memories.

1.1 Network Dynamics and Architecture

We consider a standard firing rate model (Dayan and Abbott 2001; Ermentrout and Terman 2010) with heterogeneous timescales,

$$\frac{dx_i}{dt} = -\frac{1}{\tau_i}x_i + \varphi\left(\sum_{j=1}^{n} W_{ij}x_j + b_i\right), \quad \text{for } i = 1, \ldots, n,$$

where n is the number of neurons. The real-valued function $x_i = x_i(t)$ is the firing rate of the ith neuron, b_i is the external input to the ith neuron, and W_{ij} denotes the effective strength of the recurrent connection from the jth to the ith neuron. The timescale τ_i gives the rate of recovery to rest in the absence of external or recurrent inputs. The nonlinear function $\varphi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ satisfies $\varphi(y) = 0$ whenever $y \leq 0$, and ensures that the firing rates $x_1, \ldots, x_n$ are nonnegative. Although the threshold appears to be zero for all neurons, heterogeneous thresholds can easily be incorporated into the b_is. Note that $\tau_i > 0$, while b_i and W_{ij} can take on both positive and negative values. The dynamics of the network can be described more compactly as

$$\dot{x} = -Dx + \varphi(Wx + b), \tag{1}$$

where $D \overset{\text{def}}{=} \operatorname{diag}({\tau_1}^{-1}, \ldots, {\tau_n}^{-1})$ is the diagonal matrix of inverse time constants, and φ is applied to each coordinate when the argument is a vector. Note that we do not require that the matrix W respect Dale's law,[1] as the entries are considered to be *effective* connection strengths between *principal* (excitatory) neurons. Negative weights thus reflect effectively inhibitory interactions, mediated by the presence of

[1] Dale's law states that every element in the same column of the connectivity matrix must have the same sign (Dayan and Abbott 2001). This is because neurons have either purely excitatory or purely inhibitory synapses onto other neurons.

non-specific interneurons that do not otherwise enter into the model. We always assume that the diagonal entries of the matrix $-D + W$ are strictly negative; otherwise, individual neurons may experience a run-away excitatory drive even in the absence of external or recurrent inputs. We will also assume that b_i is constant in time, though it may vary across neurons. For a given choice of nonlinearity φ, the network and its dynamics (1) are denoted by the pair of matrices (W, D).

It is worth noting here that we regard the model (1) as a description of *fast-timescale dynamics*. A more realistic network would also include stochastic fluctuations and adaptive mechanisms on a slower timescale (Abbott and Regehr 2004), so that "fixed points" of the fast-timescale dynamics only appear for short periods of time. Such fixed points will serve as our model for (transiently) activated memory patterns.

To study network flexibility, we will think of the matrix of effective connection strengths W as a sum of two components,

$$W = J + A,$$

where J corresponds to a fixed and underlying *architecture*, and A is a matrix of *perturbations* about J. While J reflects broad patterns of connection strengths that may be conserved across animals or across time, the matrix A captures individual variations, and is constantly changing as a function of the animal's learning and experience. Our main question is then the following:

Main Question (Version 1) What architectures J allow maximal flexibility for learning and unlearning new memory patterns under small perturbations A?

We will consider this as a question about the strengths, or *weights*, of the recurrent connections between neurons, rather than as a question about which neurons are connected. The pattern of allowed connections between neurons will be treated as a constraint. Indeed, networks with different *strengths* of connections, but identical connectivity patterns, may have significantly different dynamics and attractors. Moreover, in biological neural circuits, the anatomical connectivity may be difficult to modify, but the weights of synapses are known to change on relatively short timescales in response to learning and experience.

1.2 Memory Patterns as "Stable Sets" in Threshold-Linear Networks

Before addressing our main question about perturbations of network architectures, we investigate the set of memory patterns corresponding to any fixed network (W, D). The idea of stable fixed points as a model for stored memory patterns in recurrent networks dates back at least to Hopfield (1982). Following the framework of Hahnloser et al. (2003), subsets of neurons that are active at stable fixed points of (1) will serve as our model for stored memory patterns.

Recall that a fixed point x^* is *asymptotically stable* if there exists an open neighborhood U of x^* such that $\lim_{t\to\infty} x(t) = x^*$ for every trajectory $x(t)$ with the initial condition $x(0) \in U$. If the fixed point x^* has only the property that for all nearby initial conditions, $x(0) \in U$, the trajectory $x(t)$ remains very close to x^* for all later times, then x^* is a *stable* point of the network dynamics. Note that every asymptotically stable point is stable, but the converse is not true. For a given firing rate vector $x \in \mathbb{R}^n_{\geq 0}$, we call the subset of active neurons the *support* of x,

$$\mathrm{supp}(x) \stackrel{\text{def}}{=} \{i \mid x_i > 0\} \subset \{1, \dots, n\}.$$

Definition 1.1 (Stable, marginal, and unstable sets) Let (W, D) be a network on n neurons with nonlinearity φ. A nonempty subset of neurons $\sigma \subset \{1, \dots, n\}$ is a *stable set* of (W, D) if there exists an asymptotically stable fixed point x^* of the dynamics (1) such that $\mathrm{supp}(x^*) = \sigma$, for at least one external input vector $b \in \mathbb{R}^n$. A *marginal set* of (W, D) is a nonempty subset of neurons σ for which there exists a stable fixed point of the dynamics (but no asymptotically stable fixed point) with support σ for at least one external input vector b, and an *unstable set* of (W, D) is a non-empty subset of neurons that is neither stable nor marginal.[2]

Stable sets are our model for memory patterns encoded by the network. For a fixed external input b, there may be one, many, or no asymptotically stable fixed points. As we range over all possible inputs, we obtain the set of stable sets of the network. Clearly, there can be at most $2^n - 1$ stable sets in a network of n neurons. In cases of interest, however, the recurrent network performs meaningful computations precisely because only a small fraction of subsets are stable (Hahnloser et al. 2003). Note that a pair of neurons in a stable set need not be connected (i.e., we may have $W_{ij} = 0$ for i, j in a stable set).

In general, it is very difficult to determine analytically the stable fixed points for a high-dimensional, nonlinear dynamical system. If the nonlinearity φ in (1) is threshold-linear, however, it is possible to use standard tools from linear systems of ordinary differential equations to obtain exact results. For this reason, we now restrict ourselves to *threshold-linear networks*, which are networks (W, D) where the nonlinearity is chosen as

$$\varphi(y) = [y]_+ \stackrel{\text{def}}{=} \begin{cases} y & \text{if } y > 0, \\ 0 & \text{if } y \leq 0. \end{cases}$$

Although sigmoids more closely match experimentally measured input-output curves for neurons, the above threshold-nonlinearity is often a good approximation when neurons are far from saturation (Dayan and Abbott 2001; Geffen et al. 2009). If we assume that encoded memory patterns are realized by neurons firing sufficiently below saturation, it is reasonable to model them as stable sets of the threshold-linear dynamics:

$$\dot{x} = -Dx + [Wx + b]_+. \tag{2}$$

In a (nondegenerate) linear system, $\dot{x} = (-D + W)x + b$, there can be at most one fixed point of the dynamics for a given input vector $b \in \mathbb{R}^n$; its stability is character-

[2]Stable and unstable sets were previously introduced in Hahnloser et al. (2003), where they were called “permitted” and “forbidden” sets, respectively.

ized by the eigenvalues of the matrix $-D + W$. Unlike linear systems, the threshold-linear network (2) can exhibit multiple fixed points for the same input vector b. It turns out, however, that stable, unstable, and marginal sets of neurons in threshold-linear networks have simple characterizations in terms of the eigenvalues of the corresponding *principal submatrices* of $-D + W$.

Given an $n \times n$ matrix A, and a subset $\sigma \subset \{1, \dots, n\}$, the *principal submatrix* A_σ is the square matrix obtained by restricting A to the index set σ; i.e., if $\sigma = \{s_1, \dots, s_k\}$, then A_σ is the $k \times k$ matrix with $(A_\sigma)_{ij} = A_{s_i s_j}$. We call a square matrix *stable* if all its eigenvalues have strictly negative real part. We call a matrix *unstable* if at least one eigenvalue has strictly positive real part, and *marginally stable* if no eigenvalue has strictly positive real part and at least one eigenvalue is purely imaginary. Marginally stable matrices are thus on the boundary between stable and unstable matrices.

We now state our characterization of stable sets in terms of the stability of principal submatrices.

Theorem 1.2 *Let (W, D) be a threshold-linear network. A subset of neurons σ is a stable set of (W, D) if and only if the principal submatrix $(-D + W)_\sigma$ is stable. Similarly, σ is a marginal set or an unstable set of (W, D) if and only if $(-D + W)_\sigma$ is marginally stable or unstable, respectively.*

In the special case of symmetric threshold-linear networks, where the matrix W is symmetric, the equivalence between stable ("permitted") sets and stable principal submatrices was shown in Hahnloser et al. (2003). We give the proof of Theorem 1.2 in Sect. 2.

1.3 G-Constrained Networks

There are two ways in which a zero-weight connection between two neurons may arise. On the one hand, there may be a lack of anatomical connectivity between the neurons. On the other hand, many synaptic connections that appear anatomically are not functional—these are referred to as *silent synapses* (Kerchner and Nicoll 2008). While the first type of zero-weight connection cannot be perturbed without major changes to the network architecture, silent synapses may become active via small modifications. In addressing our main question, we are therefore interested in characterizing maximally flexible networks where some connections are constrained to be zero, while the remaining weights (some of which may also be zero) can be modified by small perturbations of the network. The following definitions hold for general networks, not just threshold-linear ones.

Let $G = (V, E)$ be a simple graph with vertices $V = \{1, \dots, n\}$ and edges E. We say that an $n \times n$ architecture matrix J is *constrained by* the graph G if $J_{ij} = 0$ for all edges $(ij) \notin E$. By abuse of notation, we often use G to refer to the edge set E. Note that all architectures on n neurons are constrained by the complete graph $G = K_n$. If for $(ij) \in G$ the entry $J_{ij} = 0$, we say that there is a *silent connection* from neuron j to neuron i. This mirrors the phenomenon of silent synapses in the brain.

We define an *ε-perturbation* of a network architecture J to be a matrix A whose entries all satisfy $|A_{ij}| \le \varepsilon$. We say that an ε-perturbation is *consistent with G* if the matrix A satisfies $A_{ij} = 0$ for all $(ij) \notin G$. In other words, consistent ε-perturbations can only perturb entries that are not constrained to be zero (including silent connections).

When considering an architecture J that is constrained by a graph G, we refer to the network as $(J, D)_G$. For a given φ, we use the following notation for the set of all G-constrained network architectures:

$$\mathcal{N}(G) \stackrel{\text{def}}{=} \{(J, D)_G\} = \{(J, D) \mid J_{ij} = 0 \text{ for all } (ij) \notin G\}.$$

Note that the set of constrained architectures is independent of the nonlinearity φ. When $G = K_n$ is the complete graph (no constraint), we will simply write $\mathcal{N}(n) \stackrel{\text{def}}{=} \mathcal{N}(K_n)$. If $G_1 \subset G_2$, then $\mathcal{N}(G_1) \subset \mathcal{N}(G_2)$. An ε-perturbation of a network $(J, D)_G$ will always be assumed to be consistent with G, and hence to stay within $\mathcal{N}(G)$.

We can now state our main question a bit more precisely.

Main Question (Version 2) For a given constraint graph G, what network architectures $(J, D)_G \in \mathcal{N}(G)$ allow the maximal number of subsets of neurons that can become both stable sets (learned/encoded) and unstable sets (forgotten/unencoded) via arbitrarily small ε-perturbations of J?

We call such subsets of neurons *flexible* memory patterns.

1.4 Flexible Memory Patterns as "Flexible Cliques"

Intuitively, a flexible memory pattern is a subset of neurons that can become both a stable set and an unstable set via only small modifications of the network's connection strengths. Ideally, these modifications should be specific enough not to change the stability of any other subsets. Moreover, we would like flexible memory patterns to correspond to subsets of neurons that are unconstrained in their connections to each other. In other words, these subsets of neurons should be all-to-all connected in the sense that all mutual connections can be perturbed, although some may be zero-weight (silent) connections. We model such memory patterns as "flexible cliques"; a precise definition is given below.

Recall that a *clique* in a graph G is a subset of vertices that are all-to-all connected, and the *clique complex* of G, denoted $X(G)$, is the set of all cliques. We will say that $\sigma \subset \{1, \ldots, n\}$ is a *stable clique* of the network $(W, D)_G$ if σ is a stable set and $\sigma \in X(G)$. Similarly, an *unstable clique* is an unstable set σ such that $\sigma \in X(G)$, and a *marginal clique* is a marginal set σ such that $\sigma \in X(G)$. Because the stability of a matrix forces one or more of its principal submatrices to be stable (see Lemma 3.15), one cannot require that a perturbation that changes the stability of a marginal clique in a threshold-linear network also preserve all other marginal cliques. For this reason, we introduce the notions of "maximally stable" and "minimally unstable" cliques. A *maximally stable clique* is a stable clique that is not properly contained in any larger stable clique; a *minimally unstable clique* is an unstable clique that does not properly contain any other unstable clique. We can now define flexible cliques.

Definition 1.3 (Flexible clique) We call a subset of neurons $\sigma \subset \{1, \ldots, n\}$ a *flexible clique* of a network architecture on n neurons, $(J, D)_G$, if for every $\varepsilon > 0$ there exist ε-perturbations A_s and A_u, consistent with G, such that σ is a maximally stable clique of $(J + A_s, D)_G$ and a minimally unstable clique of $(J + A_u, D)_G$.

Flexible cliques are our model for flexible memory patterns. All flexible cliques are marginal cliques, but the converse is not true (see Sect. 4.1). This is because the flexibility of a marginal clique depends on the relationship of this clique to other cliques in the network. In general, it is difficult to determine whether or not a marginal clique is flexible in a network with many marginal cliques. We are interested in precisely this case, as we look for properties of networks with the maximal number of flexible cliques.

1.5 Statement of the Main Results

Our main results all concern threshold-linear networks only. Consequently, from now on we assume $\mathcal{N}(n)$ and $\mathcal{N}(G)$ correspond to sets of unconstrained and G-constrained threshold-linear networks, respectively. First, we define what we mean by the "flexibility" and "rank" of a network.

Definition 1.4 (Network flexibility, rank, and completion) We define the *flexibility* of a network as the number of flexible cliques, and denote it: $\mathrm{flex}(J, D)_G$. We define the *rank* of a network $(J, D)_G$ to be the rank of the matrix $-D + J$. We say that a G-constrained network on n neurons, $(J, D)_G$, has a rank k *completion* if there exists a network $(\bar{J}, \bar{D}) \in \mathcal{N}(n)$ of rank k such that $\bar{D} = D$ and $\bar{J}_{ij} = J_{ij}$ for all $i = j$ and all distinct pairs $(ij) \in G$.

We now further refine our main question.

Main Question (Version 3) For a given constraint graph G, what threshold-linear networks $(J, D)_G \in \mathcal{N}(G)$ attain maximum flexibility?

Note that the flexibility of a G-constrained network $(J, D)_G$ is bounded by the total number of non-empty cliques in the corresponding clique complex $X(G)$, and by the fact that single neurons can not be flexible cliques because $-D + J$ has strictly negative diagonal. Thus,

$$\mathrm{flex}(J, D)_G \leq \left|X(G)\right| - n - 1,$$

where n is the number of neurons. In $\mathcal{N}(n)$, the flexibility can be at most $2^n - n - 1$. Most networks, however, have no flexible cliques.

The rank of any network $(J, D)_G$ is at least 1. For threshold-linear networks, rank 1 networks are good candidates for attaining maximum flexibility because all but the 1×1 principal submatrices are marginally stable. Indeed, we find that rank 1 networks attain the upper bound on flexibility in $\mathcal{N}(n)$, and a similar statement is true about G-constrained networks.

Theorem 1.5 *All rank* 1 *threshold-linear networks on* n *neurons are maximally flexible in* $\mathcal{N}(n)$*, and have flexibility* $2^n - n - 1$*. All* G*-constrained threshold-linear networks with a rank* 1 *completion are maximally flexible in* $\mathcal{N}(G)$*, and have flexibility* $|X(G)| - n - 1$.

The proof is given in Sect. 4.2.

Can any networks other than ones that are rank 1, or have rank 1 completions, attain maximal flexibility? The following example demonstrates that a G-constrained network can be maximally flexible without having a rank 1 completion.

Example 1.6 Let $G = (V, E)$ be a simple graph with vertices $V = \{1, 2, 3, 4\}$ and edges $E = \{(12), (23), (34), (41)\}$. G is a cycle on 4 vertices, so the clique complex $X(G)$ has no cliques of size greater than 2. Consider the threshold-linear network $(J, D)_G \in \mathcal{N}(G)$, where

$$-D + J = \begin{pmatrix} -1 & 2 & 0 & 1 \\ 1/2 & -1 & 1 & 0 \\ 0 & 1 & -1 & 1 \\ 1 & 0 & 1 & -1 \end{pmatrix}.$$

Using Lemma 4.1 (see Sect. 4.1), it is easy to see that all $\sigma \in X(G)$ such that $|\sigma| = 2$ are flexible cliques, and hence $(J, D)_G$ is maximally flexible in $\mathcal{N}(G)$. Despite this, $(J, D)_G$ does not have a rank 1 completion, since there is no rank 1 matrix that agrees with $-D + J$ on all of its nonzero entries (cf. Example 3.8 in Sect. 3.2).

It turns out, however, that examples of this kind can be eliminated by imposing a simple topological condition on the clique complex of the constraint graph G. Note that a clique complex is an abstract simplicial complex, whose homology groups can be computed using simplicial homology. The following is our main result.

Theorem 1.7 *Let* $(J, D)_G$ *be a maximally flexible threshold-linear network in* $\mathcal{N}(G)$*, and suppose that the clique complex* $X(G)$ *satisfies* $H_1(X(G); \mathbb{Z}) = 0$*. Then* $(J, D)_G$ *has a rank* 1 *completion. In particular,* $(J, D)_G$ *has no silent connections.*

The vanishing of $H_1(X(G); \mathbb{Z})$ may at first appear to be a strong condition, but in fact it is generically satisfied for large random networks that are not overly sparse. For example, it was recently shown in Kahle (2009) that if G is an Erdös–Rényi graph with edge probability p (i.e., a random graph on n vertices with independent connection probability p between any pair of vertices), then $p \geq n^{-\alpha}$ with $\alpha < 1/3$ implies that the probability of $H_1(X(G); \mathbb{Z}) = 0$ approaches 1 as $n \to \infty$. For $n = 10^4$ neurons, the first homology group of the clique complex is expected to vanish for connection probabilities as low as $p = .05$.

The proof of Theorem 1.7 is given in Sect. 4.3. For the complete graph K_n, $H_1(X(K_n), \mathbb{Z}) = 0$, thus the following result for unconstrained networks is a corollary of Theorems 1.5 and 1.7.

Theorem 1.8 *A threshold-linear network is maximally flexible in* $\mathcal{N}(n)$ *if and only if it is a rank* 1 *network.*

We give in Sect. 4.3 a separate proof of this theorem independent of homology/cohomology arguments.

1.6 Discussion

In this article, we have laid out the foundations for a theory of flexible memory networks, that is, for recurrent networks with memory patterns that can be both encoded (learned) and unencoded (forgotten) by arbitrarily small perturbations of the matrix of connection strengths between neurons. Given a constraint graph G of allowed connections, we have found, modulo a mild topological condition, that maximally flexible networks in $\mathcal{N}(G)$ correspond precisely to networks $(J, D)_G$ that have a rank 1 completion. These results may provide valuable insights for understanding fast-learning, flexible networks in the brain.

Our results are based on an analysis of the fixed point attractors of a standard firing rate model (2). We emphasize, however, that we regard this only as a model of the *fast-timescale dynamics* of a recurrent network; in a more comprehensive model, additional elements such as stochastic fluctuations, changing external inputs, and adaptive variables on a slower timescale will all lead to frequent transitions between attractors of the fast-timescale equations (2). This approach has proven particularly fruitful in modeling of hippocampal networks, where "bump attractor" models on a fast timescale form integral building blocks to more comprehensive models that have been successful in describing experimental observations from simultaneously recorded populations of neurons (Samsonovich and McNaughton 1997; McNaughton et al. 2006; Romani and Tsodyks 2010; Itskov et al. 2011).

Thus far we have only considered the extreme case of maximally flexible networks. For these networks, arbitrarily small perturbations of the synaptic weights between neurons are sufficient to encode new memory patterns. Nevertheless, larger perturbations will be necessary for these patterns to be robust in the presence of various plasticity mechanisms that are engaged during ongoing spontaneous activity. For any given learning rule and/or constraint on the rate of synaptic changes, however, maximally flexible networks have the best chance to quickly encode (or unencode) new memories as stable fixed points of the dynamics.

It may be possible to extend these results to other flexible networks. Do all rank k networks have the same flexibility? What is the general relationship between the flexibility of a network and its rank? What is the appropriate generalization of the topological condition in Theorem 1.7 when considering higher rank completions? We leave these questions to future work.

The remainder of this paper is organized as follows. In Sect. 2, we prove Theorem 1.2, making the connection between fixed points of the recurrent network dynamics and the stability of principal submatrices. In Sect. 3, we develop some matrix-theoretic results that are critical to proving our main theorems. This section is self-contained, independent of the context of neural networks. In Sect. 4, we prove our main results, Theorems 1.5, 1.7, and 1.8.

2 Stable Sets Correspond to Stable Principal Submatrices

Recall that a threshold-linear network (W, D) has dynamics described by the system,

$$\dot{x} = -Dx + [Wx + b]_+, \tag{3}$$

with $x \in \mathbb{R}^n_{\geq 0}$ the firing rate vector, D a diagonal matrix with strictly positive diagonal entries, and $-D + W$ an $n \times n$ matrix having strictly negative diagonal entries. For such networks, one is able to obtain qualitative characterizations (stable, marginal, unstable) of sets of neurons. In what follows, we consider fixed points of (3) for a fixed input vector $b \in \mathbb{R}^n$.

Suppose there exists a fixed point of (3) with all neurons firing, i.e., $x^* > 0$. Since $Dx^* > 0$, we can drop the threshold in a small neighborhood of this fixed point, where the system is described by the linear system $\dot{x} = (-D + W)x + b$. If the matrix $-D + W$ is invertible, the linear system has exactly one fixed point, $(D - W)^{-1}b$, although this fixed point may or may not be in the regime $\mathbb{R}^n_{>0}$ where all neurons are firing. Either $x^* = (D - W)^{-1}b > 0$, or we have a contradiction and there is no x^*. As is well known for linear systems, the fixed point x^* is asymptotically stable if and only if the matrix $-D + W$ is a stable matrix.

In addition to a possible fixed point with all neurons firing, the system (3) may also have fixed points corresponding to proper subsets of neurons with non-zero firing rate. Let $\sigma = \mathrm{supp}(x^*) \subset \{1, \ldots, n\}$ be the subset of neurons that are firing at the fixed point x^*, with the complement $\bar{\sigma}$ representing the remaining (silent) neurons. To describe these types of fixed points, we reorder the neurons and write

$$W = \begin{bmatrix} W_{\bar{\sigma}\bar{\sigma}} & W_{\bar{\sigma}\sigma} \\ W_{\sigma\bar{\sigma}} & W_{\sigma\sigma} \end{bmatrix}, \qquad D = \begin{bmatrix} D_{\bar{\sigma}} & 0 \\ 0 & D_{\sigma} \end{bmatrix}, \qquad x = \begin{pmatrix} x_{\bar{\sigma}} \\ x_{\sigma} \end{pmatrix}, \quad \text{and} \quad b = \begin{pmatrix} b_{\bar{\sigma}} \\ b_{\sigma} \end{pmatrix}.$$

The system (3) becomes

$$\begin{aligned} \dot{x}_{\bar{\sigma}} &= -D_{\bar{\sigma}} x_{\bar{\sigma}} + [W_{\bar{\sigma}\bar{\sigma}} x_{\bar{\sigma}} + W_{\bar{\sigma}\sigma} x_{\sigma} + b_{\bar{\sigma}}]_+, \\ \dot{x}_{\sigma} &= -D_{\sigma} x_{\sigma} + [W_{\sigma\bar{\sigma}} x_{\bar{\sigma}} + W_{\sigma\sigma} x_{\sigma} + b_{\sigma}]_+, \end{aligned}$$

and, since $x^*_{\bar{\sigma}} = 0$, the fixed point equations for x^* simplify to:

$$\begin{aligned} 0 &= [W_{\bar{\sigma}\sigma} x^*_{\sigma} + b_{\bar{\sigma}}]_+, \\ D_{\sigma} x^*_{\sigma} &= [W_{\sigma\sigma} x^*_{\sigma} + b_{\sigma}]_+. \end{aligned}$$

Since $D_{\sigma} x^*_{\sigma} > 0$, we can drop the threshold in the second equation and obtain

$$(D_{\sigma} - W_{\sigma\sigma}) x^*_{\sigma} = b_{\sigma}. \tag{4}$$

However, a solution x^* of this equation only yields a valid fixed point if $x^*_{\sigma} > 0$ and $W_{\bar{\sigma}\sigma} x^*_{\sigma} + b_{\bar{\sigma}} \leq 0$.

To analyze the stability of a fixed point x^* with $\mathrm{supp}(x^*) = \sigma$, we make the following change of coordinates. Let $\binom{y}{z} \overset{\text{def}}{=} x - x^*$, with $y \in \mathbb{R}^{|\bar{\sigma}|}_{\geq 0}$ and $z \in \mathbb{R}^{|\sigma|}_{\geq 0}$. Then x^* is a stable fixed point of (3) if and only if the origin is a stable fixed point of:

$$\dot{y} = -D_{\bar{\sigma}} y + \left[W_{\bar{\sigma}\bar{\sigma}} y + W_{\bar{\sigma}\sigma} z + (W_{\bar{\sigma}\sigma} x^*_{\sigma} + b_{\bar{\sigma}})\right]_+,$$

$$\dot{z} = -D_\sigma(z + x^*_\sigma) + \left[W_{\sigma\bar{\sigma}}\, y + W_{\sigma\sigma} z + (W_{\sigma\sigma} x^*_\sigma + b_\sigma)\right]_+.$$

The existence of the fixed point $(y = 0, z = 0)$ implies that $W_{\bar{\sigma}\sigma} x^*_\sigma + b_{\bar{\sigma}} \leq 0$ and $W_{\sigma\sigma} x^*_\sigma + b_\sigma > 0$. If we further assume that $W_{\bar{\sigma}\sigma} x^*_\sigma + b_{\bar{\sigma}} < 0$, then there exists an open neighborhood of the origin for which the sign of each of the thresholded terms is determined by the constant terms (those that do not involve y or z). In this neighborhood, we can simplify the thresholds and, using (4), the equations take the form,

$$\dot{y} = -D_{\bar{\sigma}}\, y,$$
$$\dot{z} = -D_\sigma z + W_{\sigma\bar{\sigma}}\, y + W_{\sigma\sigma} z.$$

Because the system is exactly linear in a neighborhood of the fixed point, x^* is asymptotically stable if and only if the matrix

$$M = \begin{bmatrix} -D_{\bar{\sigma}} & 0 \\ W_{\sigma\bar{\sigma}} & -D_\sigma + W_{\sigma\sigma} \end{bmatrix}$$

is stable. Similarly, x^* is stable but not asymptotically stable if and only if M is marginally stable, and x^* is an unstable fixed point if and only if M is unstable. Finally, note that the stability of M is equivalent to the stability of $-D_\sigma + W_{\sigma\sigma} = (-D + W)_\sigma$.

We collect these observations into the following characterization of fixed points in threshold-linear networks.

Proposition 2.1 *Consider the system* (3), *for a threshold-linear network* (D, W) *on* n *neurons with fixed input* b, *and let* $\sigma \subset \{1, \ldots, n\}$ *be a subset of neurons. The following statements hold*:

(i) *A point* x^* *with* $\mathrm{supp}(x^*) = \sigma$ *is a fixed point if and only if* x^*_σ *satisfies*:
  (a) $(D_\sigma - W_{\sigma\sigma})x^*_\sigma = b_\sigma$,
  (b) $x^*_\sigma > 0$, *and*
  (c) $b_{\bar{\sigma}} \leq -W_{\bar{\sigma}\sigma} x^*_\sigma$.
  In particular, if $\det(D_\sigma - W_{\sigma\sigma}) \neq 0$, *then there exists at most one fixed point with support* σ, *and it is given by* $x^*_\sigma = (D_\sigma - W_{\sigma\sigma})^{-1} b_\sigma$.

(ii) *Suppose* x^* *is a fixed point with* $\mathrm{supp}(x^*) = \sigma$. *If* $b_{\bar{\sigma}} < -W_{\bar{\sigma}\sigma} x^*_\sigma$, *then* x^* *is asymptotically stable if and only if the principal submatrix* $(-D + W)_\sigma$ *is stable. Similarly,* x^* *is stable but not asymptotically stable if and only if* $(-D + W)_\sigma$ *is marginally stable, and* x^* *is an unstable fixed point if and only if* $(-D + W)_\sigma$ *is unstable.*

Using this proposition, we can now prove Theorem 1.2.

Proof of Theorem 1.2 We begin with the first statement. ($\Rightarrow$) Let σ be a stable set of (W, D), and choose b such that there exists an asymptotically stable fixed point x^* of (3) with $\mathrm{supp}(x^*) = \sigma$. By part (i) of Proposition 2.1, it is clear that we can choose b such that $b_{\bar{\sigma}} < -W_{\bar{\sigma}\sigma} x^*_\sigma$. It then follows from part (ii) that $(-D + W)_\sigma$ is stable.

($\Leftarrow$) Now suppose that $(-D+W)_\sigma$ is stable. We construct b, the input vector in (3), so that the corresponding fixed point with support σ is asymptotically stable. Let $b_\sigma = (D-W)_\sigma 1_\sigma$, where 1_σ is the vector of all ones. Letting x^* be the firing rate vector with $\mathrm{supp}(x^*) = \sigma$ and $x^*_\sigma = 1_\sigma > 0$, we choose $b_{\bar{\sigma}}$ to satisfy $b_{\bar{\sigma}} < -W_{\bar{\sigma}\sigma} x^*_\sigma$. Note that $(D_\sigma - W_{\sigma\sigma}) x^*_\sigma = b_\sigma$. For this choice of b, it thus follows from part (i) of Proposition 2.1 that x^* is a fixed point, and by part (ii) that x^* is asymptotically stable. Hence, σ is a stable set of (W, D).

Similar arguments using Proposition 2.1 can be used to show that σ is a marginal or unstable set of (W, D) if and only if $(-D+W)_\sigma$ is marginally stable or unstable, respectively. □

As a result of Theorem 1.2, we see that in order to investigate stable, unstable, or marginal sets of neurons in threshold-linear networks we need to understand the stability of principal submatrices.

3 Matrix-Theoretic Results

In this section, we prove results concerning real matrices with strictly negative entries on the diagonal. These results will be critical for Sect. 4, where we prove our main theorems regarding maximally flexible networks. This section is self-contained, however, and the results can be understood independently of the context of neural networks.

Throughout this section, A is an $n \times n$ matrix with real coefficients and strictly negative entries on the diagonal. The matrix $\mathcal{E}_A = (\varepsilon_{ij})$ is the sign matrix associated to A; this is a matrix whose entries $\varepsilon_{ij} \in \{\pm 1, 0\}$ are the signs of the corresponding entries of A. To the matrix A, we also associate the graph G_A, which we call the *connectivity graph* of A. It is the simple graph that includes each edge (ij) unless $A_{ij} = A_{ji} = 0$. Let $X(G_A)$ be the clique complex associated to the graph G_A; we call this the *clique complex associated to* the matrix A. Note that a clique complex is an abstract simplicial complex. For any simplicial complex X and abelian group $\mathcal{G}$, we denote the associated simplicial homology and cohomology groups as $H_i(X;\mathcal{G})$ and $H^i(X;\mathcal{G})$, respectively. Finally, mirroring Definition 1.4, we call an $n \times n$ matrix $\bar{A}$ a *completion* of A if $\bar{A}_{ij} = A_{ij}$ for all $i = j$ and all distinct pairs $(ij) \in G_A$.

3.1 Bipartite Matrices

Bipartite matrices play an important role in Sect. 3.2.

Definition 3.1 (Bipartite matrix) We say that a real-valued, $n \times n$ matrix A is a *bipartite matrix* if the index set $\{1, \dots, n\}$ can be partitioned into two disjoint sets, σ and $\bar{\sigma}$, such that:

1. if $i \in \sigma$ and $j \in \bar{\sigma}$, both $A_{ij} \geq 0$ and $A_{ji} \geq 0$.
2. if $i, j \in \sigma$ or $i, j \in \bar{\sigma}$, both $A_{ij} \leq 0$ and $A_{ji} \leq 0$.

This definition is equivalent to the condition that there exists a permutation of the indices $\{1, \ldots, n\}$ such that the sign pattern of A takes on the block-form:

$$\mathcal{E}_A = \left(\begin{array}{c|c} - & + \\ \hline + & - \end{array}\right), \tag{5}$$

where "$+$" indicates a submatrix with all nonnegative entries, and "$-$" a submatrix with all nonpositive entries. From this observation, it is easy to see that all rank 1 matrices with negative diagonal are bipartite.

Lemma 3.2 *Let A be a real $n \times n$ matrix with* $\operatorname{rank} A = 1$ *and $A_{ii} < 0$ for all $i = 1, \ldots, n$. Then there exists a permutation of the indices such that the sign pattern of A is of the form* (5). *In particular, A is a bipartite matrix.*

The following result gives a sufficient condition for bipartiteness of a matrix in terms of the associated clique complex.

Lemma 3.3 (Bipartite lemma) *Let A be a real-valued $n \times n$ matrix with strictly negative diagonal, sign matrix $\mathcal{E}_A = (\varepsilon_{ij})$, connectivity graph G_A, and clique complex $X(G_A)$. Suppose that*

(i) $\varepsilon_{ij}\varepsilon_{ji} = 1$, *whenever* $(ij) \in X(G_A)$,
(ii) $\varepsilon_{ij}\varepsilon_{jk}\varepsilon_{ki} = -1$, *whenever* $(ijk) \in X(G_A)$, *and*
(iii) $H^1(X(G_A); \mathbb{Z}_2) = 0$.

Then A is a bipartite matrix.

Proof It is convenient to think of $\mathbb{Z}_2 = \{\pm 1\}$, the multiplicative group with two elements. Consider the cochain complex, with $X = X(G_A)$,

$$\mathcal{C}^0(X; \mathbb{Z}_2) \xrightarrow{\delta_0} \mathcal{C}^1(X; \mathbb{Z}_2) \xrightarrow{\delta_1} \mathcal{C}^2(X; \mathbb{Z}_2) \xrightarrow{\delta_2} \cdots \xrightarrow{\delta_{n-1}} \mathcal{C}^n(X; \mathbb{Z}_2) \xrightarrow{\delta_n} 0. \tag{6}$$

The maps are the usual coboundary operators. For example, $\delta_0(\{v_i\}) = \{e_{ij}\}$, where $e_{ij} = v_j v_i^{-1} = v_i v_j$. Similarly, $\delta_1(\{e_{ij}\}) = \{f_{ijk}\}$, where $f_{ijk} = e_{jk}e_{ik}^{-1}e_{ij} = e_{ij}e_{jk}e_{ki}$.

By (i), we have $\{\varepsilon_{ij}\} \in \mathcal{C}^1(X; \mathbb{Z}_2)$, while (ii) implies that $\{-\varepsilon_{ij}\} \in \operatorname{Ker} \delta_1$ (we need the minus sign because kernel elements map to $+1$). Using (iii), we conclude that $\{-\varepsilon_{ij}\} \in \operatorname{Im} \delta_0$. Hence, there exists a vertex labeling $\{\nu_i\} \in \mathcal{C}^0(X; \mathbb{Z}_2)$ such that $-\varepsilon_{ij} = \nu_i \nu_j$ whenever $(ij) \in X$. Let $\sigma = \{i \mid \nu_i = +1\}$ and $\bar{\sigma} = \{i \mid \nu_i = -1\}$, with $\sigma \cup \bar{\sigma} = \{1, \ldots, n\}$. The sign of an edge, $\varepsilon_{ij} = -\nu_i \nu_j$, can only be positive if $i \in \sigma$ and $j \in \bar{\sigma}$ or if $i \in \bar{\sigma}$ and $j \in \sigma$, and ε_{ij} can only be negative if $i, j \in \sigma$ or $i, j \in \bar{\sigma}$. This proves that the matrix A is bipartite. □

Example 3.4 To see why we need the cohomology condition in cases where there are zero entries, consider the matrix

$$A = \begin{pmatrix} -1 & 1 & 0 & 0 & 1 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 \\ 1 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

The graph G_A is a cycle on 5 vertices, and A is not a bipartite matrix. Nevertheless, A satisfies conditions (i) and (ii).

When the matrix A has no zero entries, however, this kind of example cannot occur. The clique complex of the complete graph, $X(K_n)$, is contractible, so $H^1(X(K_n); \mathbb{Z}_2) = 0$ and we obtain the following corollary.

Corollary 3.5 *Let A be a real-valued, $n \times n$ matrix with strictly negative diagonal and sign matrix $\mathcal{E}_A = (\varepsilon_{ij})$ with all entries ε_{ij} nonzero. Suppose that* (i) $\varepsilon_{ij}\varepsilon_{ji} = 1$, *for all* (ij), *and* (ii) $\varepsilon_{ij}\varepsilon_{jk}\varepsilon_{ki} = -1$, *for all distinct triples* (ijk). *Then A is a bipartite matrix.*

3.2 Vanishing Principal Minors and Rank

Recall that, given an $n \times n$ matrix A and a subset $\sigma \subset \{1, \dots, n\}$, the *principal submatrix* A_σ is the square matrix obtained by restricting A to the index set σ. The determinant of a principal submatrix is called a *principal minor*.

It is well known that if the rank of A is k, then all principal minors larger than $k \times k$ must vanish. The converse is not true. For example, consider a strictly upper-triangular $n \times n$ matrix. It can have rank up to $n - 1$, yet each and every principal minor vanishes. In the case of matrices with strictly negative diagonal entries, however, we do have a kind of converse; this is the subject of Lemma 3.7 and Proposition 3.9, the main result in this section. We begin with a simple technical lemma that will be used throughout.

Lemma 3.6 *Let A be a real-valued $n \times n$ matrix with strictly negative diagonal entries and clique complex $X(G_A)$. Suppose that all 2×2 and 3×3 principal minors corresponding to cliques in $X(G_A)$ vanish. Then the matrix A has symmetric sign matrix $\mathcal{E}_A$, and its entries satisfy*

$$A_{ii}A_{jj} = A_{ij}A_{ji}, \quad \text{for } (ij) \in X(G_A), \quad \text{and} \tag{7}$$

$$A_{ii}A_{jj}A_{kk} = A_{ij}A_{jk}A_{ki}, \quad \text{for } (ijk) \in X(G_A). \tag{8}$$

Proof The first set of relations (7) is obvious, and ensures that $\mathcal{E}_A$ is symmetric and $A_{ij} \neq 0$ for all $(ij) \in X(G_A)$. This, together with the vanishing of 3×3 principal minors, yields:

$$2A_{ii}A_{jj}A_{kk} = A_{ij}A_{jk}A_{ki} + A_{kj}A_{ji}A_{ik}, \quad \text{for } (ijk) \in X(G_A).$$

Given a triple $(ijk) \in X(G_A)$, denote $x = A_{ii}A_{jj}A_{kk}$ and $y = A_{ij}A_{jk}A_{ki}$, and note that $y \neq 0$. Using again (7), we can write $A_{kj}A_{ji}A_{ik} = x^2/y$. The above 3×3 condition becomes $2xy = y^2 + x^2$, and we conclude that $x = y$, which yields (8). □

We can now show that the vanishing of all 2×2 and 3×3 principal minors suffices to guarantee that A has rank 1.

Lemma 3.7 (Rank 1) *Let A be a real-valued, $n \times n$ matrix with strictly negative diagonal entries such that all 2×2 and 3×3 principal minors vanish. Then G_A is the complete graph, A is a bipartite matrix, and* rank $A = 1$.

Proof Since all 2×2 principal minors vanish, it follows that for each $i \neq j$ we have $A_{ij}A_{ji} = A_{ii}A_{jj} \neq 0$, and G_A is therefore the complete graph. By Lemma 3.6, the relations (7) and (8) are satisfied for all pairs and triples of distinct indices. Anchoring ourselves on the first row and column, we find that any entry of the matrix can be written as

$$A_{ij} = \frac{A_{ii}A_{jj}A_{11}}{A_{j1}A_{1i}} = \frac{A_{ii}A_{1j}}{A_{1i}} = \frac{A_{i1}A_{1j}}{A_{11}}.$$

Let $u = (A_{11}, A_{21}, \ldots, A_{n1})^T$ be the first column vector of A and $v = (A_{11}, A_{12}, \ldots, A_{1n})$ the first row vector. Then $A = (A_{11})^{-1}uv$, which is manifestly rank 1. It follows from Lemma 3.2 that A is bipartite. □

Can we generalize this result for matrices with zeroes, i.e., for matrices A such that G_A is not the complete graph? Here, we are looking for conditions that ensure the matrix A has a rank 1 *completion*, where the entries with zeroes are treated as "unknown" entries that can be completed to any value. In this case, we can require only that all 2×2 and 3×3 principal minors in the clique complex $X(G_A)$ vanish. The following example shows that such a requirement is insufficient to guarantee the existence of a rank 1 completion.

Example 3.8 Consider the matrix

$$A = \begin{pmatrix} -1 & a & 0 & 1/d \\ 1/a & -1 & b & 0 \\ 0 & 1/b & -1 & c \\ d & 0 & 1/c & -1 \end{pmatrix}.$$

This matrix has $G_A = (V, E)$, where $V = \{1, 2, 3, 4\}$ and $E = \{(12), (23), (34), (41)\}$. G_A is a cycle on 4 vertices, and the clique complex $X(G_A) = G_A$ since there are no 2-dimensional faces. Note that all 2×2 principal minors corresponding to 2-cliques in $X(G_A)$ vanish, and there are no 3×3 ones to check. Does this matrix have a rank 1 completion? Generically, the answer is "No." In fact, it is easy to see that a rank 1 completion exists if and only if $abcd = 1$.

The intuition we gain from this example is that there is a topological obstruction to a matrix having a rank 1 completion. It is the presence of a closed but hollow cycle in $X(G_A)$ that prevents A from having a rank 1 completion. In fact, if we added or removed an edge from the graph G_A in Example 3.8, we would have a rank 1 completion without any further condition other than the vanishing of 2×2 and 3×3 principal minors $\det A_\sigma$ for $\sigma \in X(G_A)$. The following proposition gives topological conditions that guarantee the existence of a rank 1 completion. Note that a condition ensuring that A is bipartite is needed to show that A (as opposed to only $|A|$) has a rank 1 completion.

Proposition 3.9 *Let A be a real-valued $n\times n$ matrix with strictly negative diagonal and clique complex $X(G_A)$. Let $|A|$ denote the matrix of absolute values of A. Suppose that* $\det A_\sigma = 0$ *for all $\sigma \in X(G_A)$ such that $|\sigma| = 2$ or 3. Then*

(a) $H^1(X(G_A);\mathbb{Z}_2) = 0 \implies$ *A is a bipartite matrix.*
(b) $H^1(X(G_A);\mathbb{R}) = 0 \implies$ *$|A|$ has a rank 1 completion.*
(c) $H^1(X(G_A);\mathbb{R}) = H^1(X(G_A);\mathbb{Z}_2) = 0 \implies$ *A has a rank 1 completion.*

Proof Let $X = X(G_A)$. First, observe that we satisfy the conditions of Lemma 3.6, and so we have relations (7) and (8).

(a) Let $\mathcal{E}_A = (\varepsilon_{ij})$, with $\varepsilon_{ij} \in \{\pm 1, 0\}$, be the sign matrix of A. Relations (7) and (8) imply

$$\varepsilon_{ij}\varepsilon_{ji} = 1, \quad \text{for } (ij)\in X, \quad \text{and}$$
$$\varepsilon_{ij}\varepsilon_{jk}\varepsilon_{ki} = -1, \quad \text{for } (ijk)\in X.$$

Since we also have $H^1(X;\mathbb{Z}_2) = 0$, it follows from Lemma 3.3 that A is a bipartite matrix.

(b) For every A_{ij} that is nonzero, introduce the following (real) variables:

$$L_{ij} := \ln\left(\frac{|A_{ij}|}{\sqrt{A_{ii}A_{jj}}}\right).$$

In these variables, the relations (7) and (8) are equivalent to antisymmetry and cocycle conditions on the L_{ij}:

$$L_{ij} + L_{ji} = 0, \quad \text{for } (ij)\in X, \quad \text{and} \tag{9}$$
$$L_{ij} + L_{jk} + L_{ki} = 0, \quad \text{for } (ijk)\in X. \tag{10}$$

Now consider the cochain complex

$$\mathcal{C}^0(X;\mathbb{R}) \xrightarrow{\delta_0} \mathcal{C}^1(X;\mathbb{R}) \xrightarrow{\delta_1} \mathcal{C}^2(X;\mathbb{R}) \xrightarrow{\delta_2} \cdots \xrightarrow{\delta_{n-1}} \mathcal{C}^n(X;\mathbb{R}) \xrightarrow{\delta_n} 0, \tag{11}$$

where $C^k(X;\mathbb{R})$ is the group of k-cochains with coefficients in $\mathbb{R}$. $C^0(X;\mathbb{R})$ corresponds to vertex-labelings, $C^1(X;\mathbb{R})$ is the set of edge-labelings, etc. As usual, the coboundary operators are $\delta_k(\{f_{i_0,\dots,i_k}\}) = \{g_{i_0,\dots,i_{k+1}}\}$, where

$$g_{i_0,\dots,i_{k+1}} = \sum_{j=0}^{k+1}(-1)^j f_{i_0,\dots,\widehat{i_j},\dots,i_{k+1}},$$

and $\delta_{k+1}\circ\delta_k = 0$. By assumption, $H^1(X;\mathbb{R}) = 0$, so $\operatorname{Im}\delta_0 = \operatorname{Ker}\delta_1$.

Let $L = (L_{ij})$, for $(ij) \in X$. Observe that (9) implies that $L \in \mathcal{C}^1(X; \mathbb{R})$, while the cocycle condition (10) implies that $L \in \operatorname{Ker} \delta_1$. It follows that $L \in \operatorname{Im} \delta_0$, so there exists an $a \in \mathbb{R}^n \cong \mathcal{C}^0(X; \mathbb{R})$ such that $L_{ij} = a_j - a_i$. This implies that for each $A_{ij} \neq 0$,

$$|A_{ij}| = \sqrt{A_{ii} A_{jj}} e^{L_{ij}} = \sqrt{|A_{ii}|} e^{-a_i} \sqrt{|A_{jj}|} e^{a_j}.$$

Let $u, v \in \mathbb{R}^n$ with $u_i = \sqrt{|A_{ii}|} e^{-a_i}$ and $v_j = \sqrt{|A_{jj}|} e^{a_j}$. Since $|A|$ and uv^T agree on all nonzero entries of $|A|$, the matrix $\overline{|A|} = uv^T$ is a rank 1 completion of $|A|$.

(c) Recall from the proof of part (a) that ε_{ij} is the sign of A_{ij}, so that $A_{ij} = \varepsilon_{ij} |A_{ij}|$ for each entry of A. Following the proof of Lemma 3.3, there exists a vertex labeling $\{\nu_i\} \in \mathbb{C}^0(X; \mathbb{Z}_2)$, with $\nu_i \in \{\pm 1\}$, such that $\varepsilon_{ij} = -\nu_i \nu_j$ whenever $(ij) \in X$. Choose $u, v \in \mathbb{R}^n$ as in the proof of part (b), so that $|A_{ij}| = u_i v_j$ whenever $(ij) \in X$. Now consider $\tilde{u}, \tilde{v} \in \mathbb{R}^n$ where $\tilde{u}_i = -\nu_i u_i$ and $\tilde{v}_j = \nu_j v_j$. It follows that $A_{ij} = \tilde{u}_i \tilde{v}_j$ whenever $A_{ij} \neq 0$. The matrix $\bar{A} = \tilde{u} \tilde{v}^T$ is thus a rank 1 completion of A. □

Remark 3.10 Note that Lemma 3.7 follows easily from Proposition 3.9, since the clique complex of the complete graph $X(K_n)$ is contractible, so the conditions $H^1(X(K_n); \mathbb{Z}_2) = 0$ and $H^1(X(K_n); \mathbb{R}) = 0$ are trivially satisfied. In Theorem 1.7, for simplicity we use instead the somewhat stronger condition $H_1(X(G); \mathbb{Z}) = 0$. If $H_1(X(G); \mathbb{Z}) = 0$, then $H^1(X(G); \mathbb{Z}_2) = H^1(X(G); \mathbb{R}) = 0$; this follows from the following well-known observation.

Lemma 3.11 *Let X be a simplicial complex. Assume that $H_1(X; \mathbb{Z}) = 0$. Then $H^1(X, \mathcal{G}) = 0$, for every abelian group $\mathcal{G}$.*

Proof This is a consequence of the Universal Coefficients Theorem (Hatcher 2002), which for an abelian group $\mathcal{G}$ yields the short exact sequence

$$0 \longrightarrow \operatorname{Ext}\bigl(H_{q-1}(X, \mathbb{Z}), \mathcal{G}\bigr) \longrightarrow H^q(X, \mathcal{G}) \longrightarrow \operatorname{Hom}\bigl(H_q(X, \mathbb{Z}), \mathcal{G}\bigr) \longrightarrow 0$$

for all $q \geq 1$. Note that for H a free abelian group, $\operatorname{Ext}(H, \mathcal{G}) = 0$. Since $H_0(X, \mathbb{Z})$ is always free, the above for $q = 1$ yields $H^1(X, \mathcal{G}) \cong \operatorname{Hom}(H_1(X, \mathbb{Z}), \mathcal{G}) = 0$. □

3.3 Stable and Marginally Stable Matrices

Recall that all flexible cliques are marginal cliques, and by Theorem 1.2 the marginal cliques correspond to marginally stable principal submatrices. Therefore, to make the connection to flexible cliques in Sect. 4, we need to consider what happens to a matrix when its principal submatrices are marginally stable, which is not quite the same as having vanishing determinant.

Recall that a matrix is marginally stable if no eigenvalue has strictly positive real part and at least one eigenvalue is purely imaginary. In the case of symmetric matrices, marginal stability implies the existence of a zero eigenvalue, and hence vanishing determinant. This is not in general true for nonsymmetric matrices. However, 2×2 and 3×3 marginally stable matrices with negative diagonal entries do have the following characterization.

Lemma 3.12

(i) *Let A be a 2×2 real matrix with strictly negative diagonal entries. Assume that A is marginally stable. Then $\det(A) = 0$, and the sign matrix $\mathcal{E}_A$ is a symmetric matrix with all entries nonzero.*

(ii) *Let A be a 3×3 real matrix with strictly negative diagonal entries. Assume that A is marginally stable. Then either $\det(A) = 0$, or $\det(A) \neq 0$ and A has a 2×2 stable principal submatrix.*

Proof (i) The matrix A must have a purely imaginary eigenvalue. Since $\mathrm{tr}(A) < 0$, this eigenvalue must be 0, and thus $\det(A) = 0$. As a consequence, the off-diagonal entries of the matrix A must have the same sign and are both nonzero.

(ii) Let λ_1 be a purely imaginary eigenvalue of A. We have two possibilities: either $\lambda_1 = 0$ and thus $\det(A) = 0$, or $\lambda_1 \neq 0$. In the second case, the conjugate $\bar{\lambda}_1 = -\lambda_1$ is also an eigenvalue, and since $\mathrm{tr}(A) < 0$, the third eigenvalue λ_3 must be negative and so $\det(A) \neq 0$. Consider now the characteristic polynomial of A, $P_A(X) = -X^3 + \mathrm{tr}(A)X^2 - M_2(A)X + \det(A)$, where $M_2(A)$ denotes the sum of the principal 2×2 minors of A. Using the usual expression for the coefficients of $P_A(X)$ in terms of eigenvalues of A, we find that $M_2(A) = \lambda_1(-\lambda_1) + \lambda_1\lambda_3 + (-\lambda_1)\lambda_3 = |\lambda_1|^2 > 0$. There thus exists a 2×2 principal submatrix of A with positive determinant and negative trace. This submatrix is stable. □

We also have relationships between the stability of a matrix and its principal submatrices. In the case of symmetric matrices, it follows from Cauchy's interlacing theorem that all principal submatrices of a stable matrix are stable.

Theorem 3.13 (Cauchy's interlacing theorem) *Let A be a symmetric $n \times n$ matrix, and let B be an $m \times m$ principal submatrix of A, where $m \leq n$. If the eigenvalues of A are $\alpha_1 \leq \cdots \leq \alpha_j \leq \cdots \leq \alpha_n$, and those of B are $\beta_1 \leq \cdots \leq \beta_j \leq \cdots \leq \beta_m$, then for all j we have $\alpha_j \leq \beta_j \leq \alpha_{n-m+j}$.*

Corollary 3.14 *Any principal submatrix of a stable symmetric matrix is stable. Any symmetric matrix containing an unstable principal submatrix is unstable.*

Even in the case of nonsymmetric matrices, there are still some constraints of this type. For example, we have the following observation.

Lemma 3.15 *Let A be an $n \times n$ matrix with strictly negative diagonal entries and $n \geq 2$. If A is stable, then there exists a 2×2 principal submatrix of A that is also stable.*

Proof We use the formula for the characteristic polynomial in terms of sums of principal minors:

$$P_A(x) = (-1)^n x^n + (-1)^{n-1} M_1(A) x^{n-1} + (-1)^{n-2} M_2(A) x^{n-2} + \cdots + M_n(A),$$

where $M_k(A)$ is the sum of the $k \times k$ principal minors of A. (Note that $M_1(A) = \mathrm{tr}(A)$ and $M_n(A) = \det(A)$.) The characteristic polynomial also has the well-known formula with coefficients that are symmetric polynomials in the eigenvalues; assuming A is stable, this yields $M_2(A) = \sum_{i<j} \lambda_i \lambda_j > 0$. This implies that at least one 2×2 principal minor is positive. Since the corresponding 2×2 principal submatrix has negative trace, it must be stable. □

In order to prove our main results in Sect. 4, we will also use the following well-known consequences of Cauchy's interlacing theorem. Here, A_k refers to the principal submatrix obtained by taking the upper left $k \times k$ entries of A.

Lemma 3.16 (Stable symmetric matrices) *Let A be a real symmetric $n \times n$ matrix. Then A is stable iff $(-1)^k \det(A_k) > 0$ for all $1 \leq k \leq n$.*

Corollary 3.17 *Let A be a real symmetric $n \times n$ matrix. Then A is stable iff $(-1)^{|\sigma|} \det(A_\sigma) > 0$ for every principal submatrix A_σ.*

4 Maximally Flexible Networks

In this section, we use the matrix results from Sect. 3 in order to prove our main results, Theorems 1.5, 1.7, and 1.8, characterizing maximally flexible networks.

4.1 Flexible vs. Marginal Cliques

Recall that all flexible cliques are marginal cliques, because they can be made both stable and unstable via arbitrarily small perturbations of the network's connection strengths. The converse is not true. The following lemma gives simple, but incomplete, conditions for determining whether or not a marginal clique is flexible in threshold-linear networks.

Lemma 4.1 *Let σ be a marginal clique of a threshold-linear network $(J, D)_G$.*

1. *If there exists $\tau \in X(G)$ such that either* (i) *$\tau \subsetneq \sigma$ and τ unstable, or* (ii) *$\tau \supsetneq \sigma$ and τ stable, then σ is not a flexible clique.*
2. *If, on the other hand,* (i) *for all $\tau \subsetneq \sigma$, τ is a stable clique, and* (ii) *for all $\tau \in X(G)$ such that $\tau \supsetneq \sigma$, τ is an unstable clique, then σ is a flexible clique.*

The proof follows from observing that any marginal clique can be perturbed to become stable or unstable by adding a multiple of the identity matrix to the corresponding principal submatrix, and one can always find a small enough perturbation so that the stability of all stable and unstable principal submatrices in the original matrix is preserved. It is thus straightforward to check the flexibility of marginal cliques if certain patterns of stable/unstable cliques are also present. This is illustrated in the following example.

Example 4.2 Consider the following matrices $-D+J$ for (unconstrained) threshold-linear networks $(D, J) \in \mathcal{N}(3)$:

$$M_1 = \begin{pmatrix} -1 & 0 & -2 \\ -2 & -1 & 0 \\ 0 & -2 & -1 \end{pmatrix}, \qquad M_2 = \begin{pmatrix} -1 & -1 & 1 \\ -1 & -1 & 0 \\ 0 & 1 & -1 \end{pmatrix}, \quad \text{and} \quad M_3 = \begin{pmatrix} -1 & 2 & 1 \\ 1 & -1 & 0 \\ 0 & -1 & -1 \end{pmatrix}.$$

M_1: $\{1,2,3\}$ is a flexible clique since it is marginal and all contained cliques are stable.
M_2: $\{1,2\}$ is a marginal clique but it is not flexible, since $\{1,2,3\}$ is stable.
M_3: $\{1,2,3\}$ is a marginal clique but it is not flexible, since $\{1,2\}$ is unstable.

Note that Lemma 4.1 says nothing about the situation where the cliques contained by or containing a given marginal clique are themselves also marginal. It is much more difficult to check for flexible cliques in a network with many marginal cliques. We investigate precisely this case, as we look for properties of networks with the maximal number of flexible cliques.

4.2 Proof of Theorem 1.5

We begin with an example of a network in which all the cliques with at least two neurons are flexible. Such a network is maximally flexible, and provides a reference point in proving that all rank 1 networks are maximally flexible in $\mathcal{N}(n)$. The proof relies on the following determinant formulas.

Lemma 4.3 *Let $W_n(\varepsilon, \alpha)$, for $n \geq 2$, be the symmetric $n \times n$ matrix with entries*

$$W_n(\varepsilon,\alpha)_{ij} = \begin{cases} -1, & \text{if } i = j, \\ -1+\alpha\varepsilon & \text{if } \{i,j\} = \{1,2\}, \\ -1+\varepsilon & \text{if } \{i,j\} \neq \{1,2\}. \end{cases} \tag{12}$$

Then

$$\det W_n(\varepsilon,\alpha) = (-1)^n \alpha\varepsilon^{n-1}\big(2n-2-(2n-4)\varepsilon - \big(n-2-(n-3)\varepsilon\big)\alpha\big). \tag{13}$$

In particular,

$$\det W_n(\varepsilon, 1) = (-1)^n \varepsilon^{n-1}\big(n-(n-1)\varepsilon\big). \tag{14}$$

Proof This is a straightforward determinant computation. □

We now show that the matrix with all entries -1 corresponds to a network on n neurons that has the maximal number $2^n - n - 1$ of flexible cliques, and is thus maximally flexible.

Proposition 4.4 *Let $(J, D) \in \mathcal{N}(n)$ be the network with the matrix $-D + J = -\mathbf{1}$, where $-\mathbf{1}$ is the $n \times n$ matrix having all entries -1. Then any subset $\sigma \subset \{1,\ldots,n\}$ with at least two neurons is a flexible clique. In particular,* $\text{flex}(J, D) = 2^n - n - 1$.

Proof Let σ be any subset with $|\sigma| = k \geq 2$ neurons. To show that σ is flexible, it suffices to show that there exists an $\varepsilon_0 > 0$ so that for every $0 < \varepsilon < \varepsilon_0$, there exist ε-perturbations A_s and A_u of (J, D) under which σ becomes a maximally stable clique and a minimally unstable clique, respectively. We show this via explicit construction of A_s and A_u.

Let A_s be the symmetric matrix with 0 entries on the diagonal, entries $(A_s)_{ij} = \varepsilon$ for distinct $i, j \in \sigma$, and $(A_s)_{ij} = -\varepsilon$ if either $i \notin \sigma$ or $j \notin \sigma$. Clearly, A_s is an ε-perturbation. We need to show that σ is a maximally stable clique for $(J + A_s, D)$; i.e., σ is a stable clique of the perturbed network, and any clique τ which properly contains σ is unstable. By Theorem 1.2, it is enough to show that the corresponding principal submatrices of $-\mathbf{1} + A_s$ are stable and unstable, respectively.

Recall (12), and note that the principal submatrix $(-\mathbf{1} + A_s)_\sigma = W_k(\varepsilon, 1)$, where $k = |\sigma|$. Using (14), we obtain

$$\det(-\mathbf{1} + A_s)_\sigma = (-1)^k \varepsilon^{k-1}\big(k - (k-1)\varepsilon\big). \tag{15}$$

Note that the same expression holds for any $\sigma' \subset \sigma$, with $k = |\sigma'|$. By Corollary 3.17, it follows that $(-\mathbf{1} + A_s)_\sigma$ is stable for all $0 < \varepsilon \leq 1$. To show that σ is maximally stable, observe that any clique τ properly containing σ must also contain an order 2 clique whose corresponding principal submatrix is

$$\begin{pmatrix} -1 & -1-\varepsilon \\ -1-\varepsilon & -1 \end{pmatrix},$$

which is unstable for all $\varepsilon > 0$. Since the matrix $(-\mathbf{1} + A_s)_\tau$ is symmetric, it follows from Corollary 3.14 that $(-\mathbf{1} + A_s)_\tau$, for any $\tau \supsetneq \sigma$, is unstable for all $0 < \varepsilon \leq 1$.

To generate a perturbation A_u for which the clique σ is minimally unstable, we proceed as follows. Let $0 < \varepsilon \leq 1$, and choose two neurons $i_1, i_2 \in \sigma$ such that $i_1 = \min(\sigma)$ and $i_2 = \min(\sigma - \{i_1\})$. Let A_u be the symmetric matrix with entries $(A_u)_{ij} = \varepsilon$ for distinct $i, j \in \sigma$ unless $\{i, j\} = \{i_1, i_2\}$. We let the entries $(A_u)_{i_1 i_2} = (A_u)_{i_2 i_1} = \alpha\varepsilon$, with α to be determined later. All other entries of A_u are set to 0. To show that σ is minimally unstable, we need to choose α so that $(-\mathbf{1} + A_u)_\sigma$ is unstable while all its proper principal submatrices are stable. Since $-\mathbf{1} + A_u$ is symmetric, Corollary 3.17 tells us that this is accomplished if the determinant of $(-\mathbf{1} + A_u)_\sigma$ has the "wrong" sign $(-1)^{k+1}$, where $k = |\sigma|$, and all $j \times j$ principal minors of $(-\mathbf{1} + A_u)_\sigma$, with $j < k$, have the "right" sign $(-1)^j$.

Observing that $(-\mathbf{1} + A_u)_\sigma = W_k(\varepsilon, \alpha)$, we have from (13) that

$$\det(-\mathbf{1} + A_u)_\sigma = (-1)^k \alpha \varepsilon^{k-1}\Big(2k - 2 - (2k-4)\varepsilon - \big(k - 2 - (k-3)\varepsilon\big)\alpha\Big). \tag{16}$$

There are two types of proper principal submatrices. The first are those that correspond to the cliques $\tau \subsetneq \sigma$ that contain both i_1 and i_2, with $j = |\tau|$, and are equal to the matrices $W_j(\varepsilon, \alpha)$. From (13), these have determinants

$$\det(-\mathbf{1} + A_u)_\tau = (-1)^j \alpha \varepsilon^{j-1}\Big(2j - 2 - (2j-4)\varepsilon - \big(j - 2 - (j-3)\varepsilon\big)\alpha\Big). \tag{17}$$

The second type of principal submatrices correspond to cliques $\nu \subsetneq \sigma$ that do not contain both i_1 and i_2. Letting $j = |\nu|$, these are equal to the matrices $W_j(\varepsilon, 1)$, and by (14) have determinants

$$\det(-\mathbf{1}+A_u)_\nu = (-1)^j \varepsilon^{j-1}\bigl(j-(j-1)\varepsilon\bigr). \tag{18}$$

Using Corollary 3.17, we see that the cliques of type ν are all stable for $0<\varepsilon\le 1$. It remains to choose α so that (16) has sign $(-1)^{k+1}$ and (17) has sign $(-1)^j$ for all $j=2,\dots,k-1$.

For $k>3$, we choose α so that

$$\frac{2k-2-(2k-4)\varepsilon}{k-2-(k-3)\varepsilon} < \alpha < \min\left\{\frac{2j-2-(2j-4)\varepsilon}{j-2-(j-3)\varepsilon}\right\}_{j=2,\dots,k-1}.$$

This is always possible, since for $0<\varepsilon\le 1$ the sequence on the right is decreasing; the minimum is attained for $j=k-1$ and is greater than the term on the left, corresponding to $j=k$. Since for $k>3$ we also have $\alpha\varepsilon < 2\frac{k-2}{k-3}\varepsilon$ the matrix A_u is a 4ε-perturbation. When $k=3$, one can choose $0<\varepsilon<\frac{1}{2}$ and $\alpha=4$, while in the case $k=2$ one needs simply to choose $\alpha<0$ so that σ is a minimally unstable clique. □

We now show that all the symmetric rank 1 networks are maximally flexible.

Proposition 4.5 *Let $(J,D)\in\mathcal{N}(n)$ be a symmetric rank* 1 *network. Then (J,D) is maximally flexible in $\mathcal{N}(n)$ and has flexibility 2^n-n-1.*

Proof Recall that by definition $(J,D)\in\mathcal{N}(n)$ is a symmetric rank 1 network if the matrix $-D+J$ is a symmetric rank 1 matrix. Since the matrix $-D+J$ has negative entries on the diagonal, there exists a vector $x\in\mathbb{R}^n$ so that $-D+J=-xx^T$. Let $\operatorname{diag}(x)$ be the $n\times n$ diagonal matrix associated to the vector $x\in\mathbb{R}^n$. Then

$$-D+J=-xx^T=\operatorname{diag}(x)(-\mathbf{1})\operatorname{diag}(x),$$

where $-\mathbf{1}$ is the $n\times n$ matrix with all entries -1 that we encountered in Proposition 4.4.

Since multiplication of a matrix on the left and right by the same diagonal matrix does not alter the sign of any principal minor, we have for any perturbation A

$$\det\bigl(-xx^T+\operatorname{diag}(x)A\operatorname{diag}(x)\bigr)_\sigma=\det(-\mathbf{1}+A)_\sigma,$$

for any $\sigma\subset\{1,\dots,n\}$. Moreover, if A is a symmetric perturbation, then so is $\operatorname{diag}(x)A\operatorname{diag}(x)$, and the stability of any principal submatrix of $-\mathbf{1}+A$ or $-xx^T+\operatorname{diag}(x)A\operatorname{diag}(x)$ is determined entirely by the signs of the principal minors (Corollary 3.17). We can thus obtain stable and unstable perturbations $\tilde{A}_s$ and $\tilde{A}_u$ of $-xx^T$ for any subset σ consisting of $|\sigma|\ge 2$ neurons by modifying the perturbations A_s and A_u in Proposition 4.4 accordingly: $\tilde{A}_s=\operatorname{diag}(x)A_s\operatorname{diag}(x)$ and $\tilde{A}_u=\operatorname{diag}(x)A_u\operatorname{diag}(x)$. We conclude that the network (J,D) is maximally flexible in $\mathcal{N}(n)$, with flexibility 2^n-n-1. □

Before proving Theorem 1.5, which extends the above results to G-constrained nonsymmetric networks in $\mathcal{N}(G)$, we need to define the notion of "pruning" of a network. We say that a graph $\tilde{G}$ is a *pruning* of the graph G if the two graphs have the same vertices and the edges of $\tilde{G}$ form a subset of the edges of G. We say that a network $(\tilde{J}, \tilde{D})_{\tilde{G}}$ is a *pruning* of $(J, D)_G$ if $\tilde{G}$ is a pruning of G, $\tilde{D} = D$, and $\tilde{J}_{ij} = J_{ij}$ for all edges $(ij) \in \tilde{G}$. The following lemma shows that flexible cliques are "inherited" by pruning.

Lemma 4.6 *Let $(\tilde{J}, D)_{\tilde{G}}$ be a pruning of $(J, D)_G$. Consider a clique $\sigma \in X(\tilde{G}) \subset X(G)$. If σ is a flexible clique of $(J, D)_G$, then σ is also a flexible clique of $(\tilde{J}, D)_{\tilde{G}}$.*

Proof This follows from the Definition 1.3 of flexible cliques. If $\sigma \in X(G)$ is a flexible clique, then there exist perturbations A_s and A_u consistent with G so that σ is maximally stable for $(J + A_s, D)_G$ and minimally unstable for $(J + A_u, D)_G$. If we also have $\sigma \in X(\tilde{G})$, define the perturbations $\tilde{A}_s$ and $\tilde{A}_u$, consistent with $\tilde{G}$, by setting all the entries in A_s and A_u corresponding to the pruned edges to 0. Since $X(\tilde{G}) \subset X(G)$, the perturbations $\tilde{A}_s$ and $\tilde{A}_u$ realize σ as a flexible clique of $(\tilde{J}, D)_{\tilde{G}}$. $\square$

We now have all the ingredients necessary for proving Theorem 1.5.

Proof of Theorem 1.5 We prove first that all rank 1 networks $(J, D) \in \mathcal{N}(n)$ are maximally flexible. Since the matrix $-D + J$ has rank 1 and negative entries on the diagonal, there exists two vectors $x, y \in \mathbb{R}^n$, with $x_i y_i > 0$ for all $i = 1, \ldots, n$, so that $-D + J = -xy^T$. Using these two vectors, we construct the diagonal matrix $d = \text{diag}(\sqrt{\frac{y_i}{x_i}})$. Let P be the matrix obtained from $-D + J$ by conjugation with the matrix d, i.e., $P = d(-D + J)d^{-1}$. It has entries

$$P_{ij} = \sqrt{\frac{y_i}{x_i}}(-x_i y_j)\sqrt{\frac{x_j}{y_j}} = -\sqrt{x_i y_i}\sqrt{x_j y_j},$$

and is therefore a rank 1 symmetric matrix. By Proposition 4.5, the network (dJd^{-1}, dDd^{-1}) is a maximally flexible network in $\mathcal{N}(n)$. Since P and $-D + J$ are similar matrices, related via conjugation by a diagonal matrix, it follows that all corresponding principal submatrices P_σ and $(-D + J)_\sigma$ are also similar. Hence, a perturbation A of the network (J, D) has exactly the same stable and unstable cliques as a perturbation dAd^{-1} of the network (dJd^{-1}, dDd^{-1}). Since (dJd^{-1}, dDd^{-1}) is maximally flexible, it follows that (J, D) is also a maximally flexible network in $\mathcal{N}(n)$.

Now let $(J, D)_G \in \mathcal{N}(G)$ be a G-constrained network with a rank 1 completion. We can think of the graph G as a pruning of the complete graph K_n on n vertices. Let $(\bar{J}, D) \in \mathcal{N}(n)$ be a rank 1 completion of the network $(J, D)_G$. By the previous arguments, the network $(\bar{J}, D)$ is maximally flexible in $\mathcal{N}(n)$, and has flexibility $2^n - n - 1$. In particular, any clique $\sigma \in X(G)$ with $|\sigma| \geq 2$ is a flexible clique of the network $(\bar{J}, D)$. By Lemma 4.6, σ is also a flexible clique of $(J, D)_G$. Since $X(G)$ has $|X(G)| - n - 1$ cliques with more than two neurons, it follows that the flexibility of (J, D) is $|X(G)| - n - 1$, which is maximal. $\square$

4.3 Proof of Theorems 1.7 and 1.8

First, we prove our main result, Theorem 1.7.

Proof of Theorem 1.7 Let $(J, D)_G$ be a maximally flexible threshold-linear network in $\mathcal{N}(G)$. This means that all the cliques $\sigma \in X(G)$ with at least two neurons are flexible. Since all flexible cliques are marginal cliques, Theorem 1.2 gives that the corresponding principal submatrices of $(-D + J)_\sigma$ are all marginally stable. In particular, all 2×2 and 3×3 principal submatrices are marginally stable, and thus by Lemma 3.12

$$\det(-D + J)_\sigma = 0, \quad \text{for all } \sigma \in X(G) \text{ with } |\sigma| = 2 \text{ or } 3.$$

Applying Lemma 3.6 to $-D + J$, it follows that for all $(ij) \in G$, the entry $J_{ij} \neq 0$. Thus, the network $(J, D)_G$ has no silent connections. By Lemma 3.11, the homology condition $H_1(X(G); \mathbb{Z}) = 0$ implies that $H^1(X(G); \mathbb{Z}_2) = H^1(X(G); \mathbb{R}) = 0$, and then by Proposition 3.9 it follows that the matrix $-D + J$ has a rank 1 completion. At the level of networks, this translates to $(J, D)_G$ having a rank 1 completion. □

Theorem 1.8 states that for the set $\mathcal{N}(n)$ of unconstrained threshold-linear networks, the maximally flexible networks are exactly the rank 1 networks. The proof is a direct application of Theorems 1.7 and 1.5.

Proof of Theorem 1.8 (⇒) This direction is a direct consequence of Theorem 1.7. Let $(J, D) \in \mathcal{N}(n)$ be a maximally flexible network. Its graph is the complete graph K_n, and thus the corresponding clique complex $X(K_n)$ is contractible and satisfies $H_1(X(K_n), \mathbb{Z}) = 0$. By Theorem 1.7 (J, D) is a rank 1 network. (⇐) This follows from first part of Theorem 1.5. □

We also give a second proof of Theorem 1.8, without appealing to the homological arguments used in the proof of Theorem 1.7.

Proof of Theorem 1.8 Without Homology/Cohomology (⇒) Suppose $(J, D) \in \mathcal{N}(n)$ is a maximally flexible network. This means (J, D) must have flexibility $2^n - n - 1$. In particular, all 2×2 and 3×3 principal submatrices of $-D + J$ must be marginally stable, and so by Lemma 3.12 all 2×2 and 3×3 principal minors must vanish. This, together with the fact that the diagonal entries are strictly negative, implies that $-D + J$ satisfies the hypotheses of Lemma 3.7, whose proof does not rely on cohomology arguments, and is thus rank 1. (⇐) This follows from first part of Theorem 1.5, which does not use homology or cohomology arguments. □

Acknowledgements C.C. was supported by NSF DMS-0920845 and a Courant Instructorship. A.D. was partially supported by NSF DMS-0505767. V.I. was supported by NSF DMS-0967377 and the Swartz Foundation.

References

Abbott, L. F., & Regehr, W. G. (2004). Synaptic computation. *Nature*, *431*(7010), 796–803.

Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience*. Cambridge: MIT Press.

Ermentrout, G. B., & Terman, D. H. (2010). *Mathematical foundations of neuroscience*. Berlin: Springer.

Geffen, M. N., Broome, B. M., Laurent, G., & Meister, M. (2009). Neural encoding of rapidly fluctuating odors. *Neuron*, *61*(4), 570–586.

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

Hatcher, A. (2002). *Algebraic topology*. Cambridge: Cambridge University Press.

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

Itskov, V., Curto, C., Pastalkova, E., & Buzsáki, G. (2011). Cell assembly sequences arising from spike threshold adaptation keep track of time in the hippocampus. *J. Neurosci.*, *31*(8), 2828–2834.

Kahle, M. (2009). Topology of random clique complexes. *Discrete Math.*, *309*(6), 1658–1671.

Kerchner, G. A., & Nicoll, R. A. (2008). Silent synapses and the emergence of a postsynaptic mechanism for LTP. *Nat. Rev., Neurosci.*, *9*(11), 813–825.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M. B. (2006). Path integration and the neural basis of the 'cognitive map'. *Nat. Rev., Neurosci.*, *7*(8), 663–678.

Romani, S., & Tsodyks, M. (2010). Continuous attractors with morphed/correlated maps. PLoS Comput Biol *6*(8).

Rutishauser, U., Mamelak, A. N., & Schuman, E. M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. *Neuron*, *49*(6), 805–813.

Samsonovich, A., & McNaughton, B. L. (1997). Path integration and cognitive mapping in a continuous attractor neural network model. *J. Neurosci.*, *17*(15), 5900–5920.