Coding Properties of Firing Rate Models with Low-Rank Synaptic Weight Matrices

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Steven Collazos

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Duane Nykamp

August, 2019
I would like to thank the numerous people that supported me throughout my time in graduate school. I would also like to thank my advisor for his guidance in the process of developing this dissertation.
Abstract

Hebbian theory proposes that ensembles of neurons, that is, groups of co-active neurons, form a basis for neural processing. We model the collection of all possible ensembles of neurons—known as permitted sets, $\mathcal{P}_\Phi(W)$—as a collection of binary strings that indicate which neurons are deemed active. In this model, $\Phi$ is a function that prescribes how neurons respond to inputs, and $W$ is a matrix that captures the strengths of the connections among neurons in the network. We construct $\mathcal{P}_\Phi(W)$ by imposing a threshold on the responsiveness of the neuron to input at the steady state. We investigate how synaptic strengths shape $\mathcal{P}_\Phi(W)$.

When the synaptic weight matrix is almost rank one, we prove two main results about $\mathcal{P}_\Phi(W)$. First, $\mathcal{P}_\Phi(W)$ is a convex code, which is a combinatorial neural code that arises from a pattern of intersections of convex sets. Second, $\mathcal{P}_\Phi(W)$ exhibits nesting, meaning that a permitted set with $k$ co-active neurons contains a permitted subset of $k-1$ co-active neurons. Our results are applicable to neuronal networks whose activation function is $C^1$ with finitely many discontinuities.
Contents

Acknowledgments i
Abstract ii
List of Figures v

1 Introduction 1

2 Background 5
  2.1 Neurobiology ................................................. 5
  2.2 Linear Algebra ................................................ 7
    2.2.1 Spectrum of Perturbed Matrices ....................... 10
  2.3 Dynamical Systems and Stability Theory .................... 12
  2.4 Computational Neuroscience ................................. 13
    2.4.1 Firing Rate Models of Neuronal Networks ............... 13
    2.4.2 Neural codes ........................................... 15

3 Overview of Previous Work 20
  3.1 Permitted Sets .............................................. 20
  3.2 Convex Codes ................................................ 26

4 Convex Coding in Firing Rate Models 28
  4.1 Threshold-Linear Case ..................................... 29
  4.2 Permitted Sets Beyond the Threshold-Linear Regime ....... 34
  4.3 $C^1$ Activation Function Case ............................ 43
List of Figures

1.1 Cell assemblies. ......................................................... 3
1.2 Example of receptive fields and combinatorial neural codes. .... 3
2.1 A schematic of a neuron. ............................................. 6
2.2 Example of a receptive field code. .................................. 17
4.1 A $C^4$ activation function. ....................................... 36
Chapter 1

Introduction

One central unsolved question in neuroscience is how neural activity and network connectivity influence each other. A threshold-linear network, a particular instance of a firing rate model, is a model representing a neuronal network’s activity that has been studied for such a purpose. One aspect of this model that has received much attention by some theoreticians is its asymptotically stable fixed points and their dependence on the network’s synaptic weight matrix, which represents the strength of the connections among neurons in the network. In the computational neuroscience literature, those points are thought to be a model for memories stored in the synaptic connections among the network’s neurons.

The main goal of this dissertation is the study of asymptotically stable fixed points of threshold-linear networks and to generalize beyond threshold-linear models by introducing customizable thresholds. The activity thresholds allow us to identify steady states with sets of active and inactive neurons and interpret our stability results in terms of allowable patterns of coactive neurons. Equipped with such activity thresholds, we can prove results for both the threshold-linear networks and more general firing rate models regarding possible patterns of coactive neurons that can be supported by networks with low-rank synaptic weight matrices. We believe our work paves the way for further analysis of activity patterns for firing rate models.

Our starting point will be the mathematical results presented by Richard Hahnloser et al in the early 2000s in a high-profile article [1]. In their study, they analyzed a neuronal network where the neurons’ response to overall synaptic input is given by a
rectifier, i.e., $\Phi(x) = \max\{x, 0\}$. The input is a stimulus and the output is a collection of units that can become stably coactive, called a permitted set of the network. Similarly, there are groups of neurons that cannot be stably coactive regardless of the choice of stimulus; such a cluster is called a forbidden set. Thus, the authors’ framework combines the interplay between digital and analog coding in neurons. The digital aspect is given by an active group of neurons given a stimulus to the network, whereas the analog feature is modeled by the (possibly) stable firing rates of the active neurons.

Groups of coactive neurons are neural phenomena of interest in neuroscience because they have been observed in a variety of contexts \[2\]. For instance, one connection between co-active neurons and learning is known as Hebb’s rule \[3\], which proposes that synapses between neurons are strengthened if the neurons fire synchronously. In freely-moving rats, subgroups of certain cells in the hippocampus have been shown to become active when the rat enters at specific locations in its environment \[4\]. The modern formulation of the hypothesis that coactive neurons play a role in neural processing is known as the cell assembly hypothesis, an idea that is due to Donald Hebb \[3, 5\]. According to the cell assembly hypothesis, the unit that determines how certain psychological processes evolve in a neural network is the cell assembly, which is a distributed collection of neurons in which excitatory connections can be strengthened via mutual excitation. A sequence of assemblies, a phase sequence, ensues after an assembly becomes active \[3\]. We illustrate the idea of cell assemblies and phase sequences in Figure \[1.1\]. Given a specific neuron $\nu$, what conveys information in the cell assembly hypothesis is the group of neurons that fire along with $\nu$, and not just the fact that $\nu$ is firing. Each assembly occurring in a phase sequence can be thought of as a step in a computation performed by the network \[3\].

A natural way of capturing which neurons are firing together is via a digital code-word, i.e., a binary string. In Figure \[1.1\] we could write 100010 to encode that 1 and 5 fired together in our first cell assembly; similarly, 010011 would encode that neurons 2, 5, 6 were co-active in the second assembly neurons. These kinds of digital codes are called combinatorial neural codes. The collection of permitted sets associated to a firing rate model is a key example of a combinatorial neural code and is the focus of our work.

An additional consideration is the kind of events that neurons respond to. One factor in determining the firing patterns of neurons, including sets of coactive neurons, is the
network connectivity structure. The restriction that connectivity places on the range of codewords, i.e., network activity patterns, is not well understood [6]. The link between connectivity and neural coding was also conjectured in the cell assembly hypothesis [3]. For all that, there are neurons that can respond reliably to certain stimuli in the environment. For instance, since at least the work of Haldan Keffer Hartline, it is known that there are neurons that will become maximally active for specific stimuli [7]. This led to discovering receptive fields of neurons, and these receptive fields often turn out to be convex [8]. We illustrate this concept in Figure 1.2. Examples of convex receptive fields include place fields of neurons in the hippocampus and tuning curves of cells in the visual cortex [9].

Figure 1.2: Three convex receptive fields. The box, denoted $X$, is the stimuli space. If a stimulus lies in $U_1$ and $U_2$ and it avoids $U_3$, the neurons associated to $U_1$ and $U_2$ will fire, but the one associated to $U_3$ will not. The corresponding RF codeword would be 110.
The relationship between combinatorial neural codes and receptive fields is given by \textit{receptive field (RF) codes}, which are codes that keep track of groups of neurons that fire simultaneously when a stimulus falls in their receptive fields. In plain English, RF codes arise from the intersection patterns of subsets of space. As Figure 1.2 suggests, these RF codes are combinatorial neural codes, and they are useful in learning about the stimuli space of neurons \cite{9, 8}. One of the main results in this dissertation is that the collection of permitted sets of a firing rate model is always a convex RF code provided the synaptic weight matrix of the network is close to being rank one.

This manuscript is organized into six chapters; this chapter is the introductory section. In the next chapter, we cover the neurobiology and computational neuroscience prerequisites, and the mathematical tools we will be using. In Chapter 3, we give an overview of the theory of permitted sets of threshold-linear networks, the results that motivated the route we pursued in this dissertation, and we present explicit examples of convex and nonconvex codes as well as a computation showing how to compute the collection of all permitted sets for a small threshold-linear network. In Chapter 4 and Chapter 5, we start by showing that threshold-linear networks satisfy the nesting property and that they exhibit convex coding when the synaptic weight matrix is almost rank one. We then proceed to use the proofs to guide how the demonstrate that almost rank-one firing rate models satisfy the nesting and convex coding properties. Finally, in Chapter 6, we give some closing remarks and future outlook.
Chapter 2

Background

2.1 Neurobiology

We start with a brief, caricatured overview of the neuron, how neurons communicate with each other when they form networks, and receptive fields of neurons. For general neurobiology information, we follow [10, 11].

Neurons, of which there are a wide variety, are one of the constituents that make up nervous tissue. See the schematic in Figure 2.1. They are cells, so they have the same organelles as regular eukaryote cells and have membranes made of phospholipids that form phospholipid bilayers. Neurons differ from regular cells in that they consist of a soma, dendrites, and an axon. A typical soma, or cell body, is about 20 µm in diameter. Dendrites are processes, i.e., outgrowths, extending from the soma, and the longest dendrites tend to be much longer than the diameter of the soma they protrude from. Axons are also projections from a neuron’s soma. The length of axons can range from less than a millimeter to over a meter. Axons branch as they extend away from the soma and they end at axon terminals.

In order to enable neurons to signal to each other, neurons’ membranes have selective ion channels that give the interior of the neuron a negative resting potential. When a neuron’s voltage is large enough, a complex chain of events results in a fast surge in voltage and is followed by quick fall in voltage. This whole electrical impulse which is sent along the neuron’s axon, is called an action potential, or spike. Action potentials are accompanied by a refractory period in which they are unlikely to produce any new
action potentials. After the neuron sends a spike, the membrane potential returns slowly to its resting potential.

Everything we have discussed thus far is applicable to single neurons; putting two or more neurons together introduces new phenomena. In the human brain, which consists of the cerebrum, cerebellum, and brain stem, there are about 85 billion neurons. Here is the stereotypical way in which two neurons, let us call them $\nu_1$ and $\nu_2$, “communicate” with each other: An action potential travels along the axon of $\nu_1$ until it reaches a pre-synaptic terminal in an axon terminal. Then neurotransmitters are released from the pre-synaptic terminal and diffuse onto a point of contact on a dendrite, i.e., a post-synaptic terminal, of $\nu_2$. These points of contact between neurons are called (chemical) synapses. The neurotransmitters that bind onto receptors of $\nu_2$’s dendrites are inputs that induce a change in voltage in $\nu_2$’s soma. Then $\nu_2$ will send an action potential down its axon when it is triggered, and the cycle repeats.

To summarize: Neuronal networks consist of neurons connected to each other via synapses and they instigate each other by firing action potentials.

Next we introduce the neurobiology that motivates the notion of receptive field codes (whose mathematical formulation we will spell out in Section 2.4.2). Researchers have
found a variety of neurons that exhibit receptive fields (RF) [9], which were introduced by Sir Charles Scott Sherrington in 1906 for referring to regions in the body that would trigger a reflex when stimulated. The notion was then extended in 1938 by Haldan Keffer Hartline to include regions of visual space that would make retinal ganglion cells respond [7]. In 1959, David Hubel and Torsten Wiesel further studied RFs of neurons in the primary visual cortex [12]. One particularly striking example—because for the first time there was a neural correlate for a very basic animal behavior, namely navigation in space—of receptive field was discovered in the hippocampus in 1971 by John O’Keefe and Jonathan Dostrovsky [13]. In summary, depending on the modality, a receptive field is the skin area or region of (stimulus) space in which stimuli elicit a response from a neuron [14].

2.2 Linear Algebra

In this section, we introduce some linear algebra notation and ideas we will be using. For the sake of keeping the discussion self-contained, we also prove some lemmas we will need in future chapters.

Let \([N]\) be defined as the set \(\{1, 2, \ldots, N\}\). Given a subset \(\tau = \{i_1, i_2, \ldots, i_k\}\) of \([N]\), where \(i_1 < i_2 < \cdots < i_k\), let \(u_\tau\) denote \((u_{i_1}, u_{i_2}, \ldots, u_{i_k}) \in \mathbb{R}^{|\tau|}\).

Let \(\text{rk}(A)\) denote the rank of \(A\). Among the several ways of characterizing the rank of a matrix, the most useful one for us will be that \(\text{rk}(A) = r\) is the smallest number for which we can express \(A\) as

\[
A = \sum_{j=1}^{r} x_j y_j^T.
\]

where \(x_k, y_k \in \mathbb{C}^N\) for all \(k \in [N]\). In particular, when \(\text{rk}(A) = 1\),

\[
A = u v^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_N \end{pmatrix} = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_N \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_N \\ \vdots & \vdots & \vdots & \vdots \\ u_N v_1 & u_N v_2 & \cdots & u_N v_N \end{pmatrix},
\]

for some \(u = (u_1, u_2, \ldots, u_N)^T, v = (v_1, v_2, \ldots, v_N)^T \in \mathbb{R}^N \setminus \{0\}\). Any two nonzero rows
of $A$ are scalar multiples of each other. A similar equality holds for columns of $A$.

We write $\text{Spec}(A)$ for the multiset of eigenvalues of $A \in \mathbb{R}^{N \times N}$, otherwise known as the spectrum of $A$. For a given $A \in \mathbb{C}^{N \times N}$, the spectrum $\text{Spec}(A)$ is a $N$-tuple $(\lambda_1, \lambda_2, \ldots, \lambda_N) \in \mathbb{C}^N$. If we need to emphasize that a given complex number is the $i$th eigenvalue of a matrix $M \in \mathbb{C}^{N \times N}$, we will write $\lambda_i(M)$. If the eigenvalues of a matrix $A$ have strictly negative real part, it is standard in stability theory to call $A$ a stable matrix. (The motivation for this term is that if $A$ is the Jacobian matrix of system of differential equations evaluated at a steady state, then the steady state will be stable when the eigenvalues of $A$ has negative real part.)

Later we will make use of the following two basic observations:

**Lemma 1.** Let $A \in \mathbb{R}^{N \times N}$ and $D \in \mathbb{R}^{N \times N}$ be a scalar diagonal matrix, i.e., $D = kI$ for some $k \in \mathbb{R}$, where $I \in \mathbb{R}^{N \times N}$ denotes the identity matrix. Then $\text{Spec}(A + D) = \{\lambda + k : \lambda \in \text{Spec}(A)\}$.

A principal submatrix [15] of $A = (a_{ij})$ determined by $\sigma$, denoted by $A_\sigma$, is the submatrix of $A$ whose entries are indexed by $\sigma$—in other words, $A_\sigma = (\tilde{a}_{u,v})$, where $\tilde{a}_{u,v} = a_{u,v}$ where $u, v \in \sigma$. A leading principal submatrix is a principal submatrix determined by $\sigma = [k]$, for some $1 \leq k \leq N$. A leading minor is the determinant of a leading principal submatrix.

Finally, the characteristic polynomial of $A \in \mathbb{C}^{N \times N}$ is defined as

$$\chi_A(t) = \prod_{j=1}^{N} (t - \lambda_j) = t^N + \sum_{j=1}^{N} (-1)^j s_j(\lambda_1, \lambda_2, \ldots, \lambda_N) t^{N-j},$$

where $\lambda_j \in \text{Spec}(A)$ and $s_j(\lambda_1, \ldots, \lambda_N)$ denotes the $j$th elementary symmetric polynomial in $\lambda_1, \ldots, \lambda_N$, i.e.,

$$s_j(\lambda_1, \ldots, \lambda_N) = \sum_{\sigma \in \binom{[N]}{j}} \prod_{k \in \sigma} \lambda_k,$$

where $\binom{[N]}{j}$ denotes the collection of all subsets of $[N]$ of cardinality $j$. It is well known [15] that

$$\chi_A(t) = t^N + \sum_{j=1}^{N} (-1)^j E_j(A) t^{N-j},$$
where

\[ E_j(A) = \sum_{\sigma \subseteq [N] \atop |\sigma| = j} \det(A_\sigma). \]

(Here \(|\sigma|\) denotes the cardinality of \(\sigma\).) Note that \(E_1(A) = \text{tr}(A) = \sum_{i=1}^N a_{ii}\), the trace of \(A\), and \(E_N(A) = \det(A)\), the determinant of \(A\).

We will be using the following two lemmas frequently. Lemma 3 is a generalization of Lemma 2:

**Lemma 2.** Let \(A = uv^T\), where \(u, v \in \mathbb{R}^N \setminus \{0\}\). Then \(\text{Spec}(A) = \{u^T v, 0\}\).

**Proof.** For any \(\sigma \subseteq [N]\) such that \(|\sigma| \geq 2\), we have that \(\det(A_\sigma) = 0\), so

\[ \chi_A(t) = t^N - E_1(A)t^{N-1} = t^{N-1} (t - \text{tr}(A)) = t^{N-1} (t - u^T v). \]

Hence, the eigenvalues of \(A\) are 0 and \(\text{tr}(A)\). \(\square\)

**Lemma 3.** Let \(A = uv^T\), where \(u, v \in \mathbb{R}^N \setminus \{0\}\). Let \(\omega \subseteq [N] = \{1, 2, \ldots, N\}\). Then \(\text{Spec}((A - I)_{\omega}) = \{u^T v_{\omega} - 1, 0\}\), where

\[ u^T v_{\omega} = \sum_{i \in \omega} u_i v_i. \]

**Proof.** For any \(\omega \subseteq [N]\) with \(|\omega| \geq 2\), we have \(A_{\omega} = u_{\omega} v_{\omega}^T\), so \(\text{rk}(A_{\omega}) \leq 1\). Hence, for any \(\sigma \subseteq \omega\), where \(|\omega| \geq 2\), we have \(\det((A_{\omega})_\sigma) = 0\). Thus,

\[ \chi_{A_{\omega}}(t) = t^{|\omega|} - E_1(A_{\omega})t^{|\omega| - 1} = t^{|\omega| - 1} (t - \text{tr}(A_{\omega})) = t^{|\omega| - 1} (t - u^T_{\omega} v_{\omega}), \]

so \(\text{Spec}(A_{\omega}) = \{u^T_{\omega} v_{\omega}, 0\}\). \(\square\)

Let \(\|\cdot\|\) be a norm on \(\mathbb{C}^N\). Then the **matrix norm induced by \(\|\cdot\|\)** is defined \([15]\) as

\[ \|\|A\|\| = \max_{\|x\|=1} \|Ax\|, \]

where \(A \in \mathbb{C}^{N \times N}\). It turns out \([15]\) that \(\rho(A) \leq \|\|A\|\|\), where \(\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|\}\) is the spectral radius of \(A\), and for any \(A, B \in \mathbb{C}^{N \times N}\), we have \(\|\|AB\|\| \leq \|\|A\|\| \|\|B\|\|\).
When $\| \cdot \|$ is the matrix 2-norm, it is known that for all $A \in \mathbb{C}^{N \times N}$,

$$\|A\|_2 = \sqrt{\lambda_{\text{max}}(A^*A)},$$

where $\lambda_{\text{max}}(A^*A)$ denotes the largest eigenvalue of $A^*A$.

Lastly, for the sake of self-containment, we state the following lemma. When $x, y \in \mathbb{C}^N$, we denote the dot product of $x$ and $y$ as $x \cdot y$.

**Lemma 4.** Let $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix. Then $\|D\| = \max_{1 \leq i \leq N} |d_i| = \|D\|_{\text{max}}$, where $d_i$ denotes the $i$th entry along the diagonal of $D$.

We show a proof of this lemma in Appendix A.1.

### 2.2.1 Spectrum of Perturbed Matrices

Our goal here is to present and prove Lemma 5 which we will use in chapters 4 and 5.

First we explain what it means for two the spectra of two matrices to be close given that the corresponding matrices are close. In order to do this, we will spell out the notion that eigenvalues of a matrix are continuous functions of the matrix’s entries. The space $\mathbb{C}^{N \times N}$ is naturally topologized by the maximum norm [15], or “max-norm” for short, on $\mathbb{C}^{N^2}$, that is, for any $A = (a_{ij}) \in \mathbb{C}^{N \times N},$

$$\|A\|_{\text{max}} = \max_{i,j} |a_{ij}|.$$

Although there is no natural ordering of $\mathbb{C}^N$, when considering the spectrum $\lambda_1, \lambda_2, \ldots, \lambda_N$ of a matrix $A \in \mathbb{C}^{N \times N}$, we expect that rearranging the list $\lambda_1, \lambda_2, \ldots, \lambda_N$ will still stand for the same spectrum of $A$. In other words, permuting the coordinates of $\text{Spec}(A) \in \mathbb{C}^N$ should be thought of as resulting in an equivalent point of $\mathbb{C}^N$. The collection of all permutations on $N$ objects forms an algebraic structure known as the symmetric group, which we denote as $\mathcal{S}_N$ and has $N! = N(N-1) \cdots 2 \cdot 1$ elements. A mathematical way of saying that $z, w \in \mathbb{C}^N$ are equivalent up to a permutation is by viewing $z$ and $w$ as an elements in $\mathcal{A}_N = \mathbb{C}^N/\mathcal{S}_N$. Therefore, if $z, w \in \mathcal{A}_N$, we have $z \sim w$ if and only if there exists $\pi \in \mathcal{S}_N$ such that

$$(z_1, z_2, \ldots, z_N) = (w_{\pi(1)}, w_{\pi(2)}, \ldots, w_{\pi(N)}).$$
For simplicity, we denote \((w_{\pi(1)}, w_{\pi(2)}, \ldots, w_{\pi(N)})\) by \(\pi \cdot w\). With such a notion of equivalence for \(z, w \in A_N\), a natural notion of distance between \(z\) and \(w\) would be to minimize the maximum norm of \(z\) and the different possible ways of permuting the coordinates of \(w\). In other words, the quotient space \(A_N\) can be topologized via the metric

\[
d(w, z) = \min_{\pi \in S_N} \max_{1 \leq j \leq N} |w_j - z_{\pi(j)}| \]

\[
= \min \left( \max_{1 \leq j \leq N} |w_j - z_{\pi_1(j)}|, \max_{1 \leq j \leq N} |w_j - z_{\pi_2(j)}|, \ldots, \max_{1 \leq j \leq N} |w_j - z_{\pi_N(j)}| \right) \]

\[
= \min \left( ||w - \pi_1 \cdot z||_{\infty}, ||w - \pi_2 \cdot z||_{\infty}, \ldots, ||w - \pi_N \cdot z||_{\infty} \right).
\]

We rewrote the last expression above by using a standard notation for the maximum norm, i.e., for \(x = (x_1, \ldots, x_N) \in \mathbb{R}^N\), we have \(||x||_{\infty} = \max(|x_1|, |x_2|, \ldots, |x_N|)\). This metric turns \(A_N\) into a complete metric space and \(\text{Spec} : \mathbb{C}^{N \times N} \to A_N\) defined by \(M \mapsto \text{Spec}(M)\) is continuous.

Now we can prove the lemma we said at the beginning of this subsection:

**Lemma 5.** Let \(A \in \mathbb{C}^{N \times N}\) and \(\epsilon > 0\) be given. Suppose \(d > 0\) is such that for any \(X \in \mathbb{C}^{N \times N}\) satisfying \(||A - X|| < d\), we have

\[
\min_{\pi \in S_N} \max_{1 \leq j \leq N} |\lambda_j(A) - \lambda_{\pi(j)}(X)| < \epsilon
\]

(where \(\lambda_n(A)\) and \(\lambda_m(X)\) denote an \(n\)th eigenvalue of \(A\) and an \(m\)th eigenvalue of \(X\), respectively). If \(X_0 \in \mathbb{C}^{N \times N}\) is such that \(||A - X_0||_{\text{max}} < d\), then

\[
\text{Spec}(X_0) = \{ \lambda_j + \Delta_j : j \in [N] ; \lambda_j \in \text{Spec}(A) ; |\Delta_j| < \epsilon \}.
\]

**Proof.** By continuity of \(\text{Spec} : \mathbb{C}^{N \times N} \to A_N\) at \(A\), there is \(d > 0\) such that for any \(X \in \mathbb{C}^{N \times N}\) satisfying \(||A - X||_{\text{max}} < d\),

\[
\min_{\pi \in S_N} \max_{1 \leq j \leq N} |\lambda_j(A) - \lambda_{\pi(j)}(X)| < \epsilon.
\]

Let \(X_0 \in \mathbb{C}^{N \times N}\) be such that \(||A - X_0||_{\text{max}} < d\).

We will show next that \(\text{Spec}(X_0)\) has the form in the conclusion of the lemma. Let
π₀ ∈ Θₙ be such that

\[
\min_{\pi \in \Theta_n} \max_{1 \leq j \leq n} |\lambda_j(A) - \lambda_{\pi(j)}(X₀)| = \max_{1 \leq j \leq n} |\lambda_j(A) - \lambda_{\pi₀(j)}(X₀)| .
\]

Define \( \Delta_j = \lambda_{\pi₀(j)}(X₀) - \lambda_j \), where \( \lambda_j \in \text{Spec}(A) \), for all \( j \in [N] \), so we have

\[
\lambda_{\pi₀(j)}(X₀) = \lambda_j + \Delta_j .
\]

Note that by definition \( |\Delta_j| < \epsilon \) for all \( j \in [N] \). Thus,

\[
\text{Spec}(X₀) = \{ \lambda_j + \Delta_j : j \in [N] ; \lambda_j \in \text{Spec}(A) ; |\Delta_j| < \epsilon \} .
\]

\[\square\]

2.3 Dynamical Systems and Stability Theory

Stability of a fixed point is a local property of a dynamical system, and a basic technique for assessing stability is to examine the eigenvalues of the associated linearized system. We review the necessary notions [17].

Consider the system of differential equations \( \dot{x} = F(x) \), where \( \dot{x} = (\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_N) \), with \( \dot{x}_k \) for all \( k \in [N] = \{1, 2, \ldots, N\} \) denoting the derivative of \( x_k \) with respect to \( t \), and \( F : \mathbb{R}^N \rightarrow \mathbb{R}^N \). Systems of this kind are called autonomous systems. A fixed point of \( \dot{x} = F(x) \) is a point \( x^* \in \mathbb{R}^N \) such that \( F(x^*) = 0 \).

Let \( x(t) \) denote a solution to \( \dot{x} = F(x) \). We call \( x(t) \) Liapunov stable (or just “stable” for brevity) if for any \( \epsilon > 0 \), there is \( \delta > 0 \) such that for any solution \( y(t) \) of \( \dot{x} = F(x) \) that satisfies \( |x(t₀) - y(t₀)| < \delta \) for some \( t₀ \in \mathbb{R} \), then \( |x(t) - y(t)| < \epsilon \) for \( t > t₀ \). If, in addition to being Liapunov stable, \( x(t) \) possesses the property that there is a constant \( b > 0 \) such that \( |x(t) - y(t)| \rightarrow 0 \) as \( t \rightarrow 0 \) whenever \( |x(t₀) - y(t₀)| < b \), then we call \( x(t) \) asymptotically stable. Letting \( DF \) denote the matrix of derivatives of \( F \)—that is, \( DF = (∂F_i/∂x_j)_{i,j} \), where \( i, j \in [N] \)—we can check stability of a fixed point \( x^* \) as follows: If all the eigenvalues of \( DF(x^*) \) have negative real part, then \( x^* \) is an asymptotically stable fixed point of \( \dot{x} = F(x) \).

Given the linearized system associated to a system of differential equations, it is
desirable to have criteria for determining whether the eigenvalues have negative real part; we present one such criterion next. The Routh-Hurwitz matrix $\Omega(A) \in \mathbb{R}^{N \times N}$ associated with $A \in \mathbb{R}^{N \times N}$ is

$$
\Omega(A) = \begin{pmatrix}
E_1(A) & E_3(A) & E_5(A) & \cdots & \cdots \\
1 & E_2(A) & E_4(A) & \cdots & \cdots \\
0 & E_1(A) & E_3(A) & \cdots & \cdots \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & E_n(A)
\end{pmatrix}.
$$

Then the Routh-Hurwitz Stability Criterion states [18]:

**Theorem 1.** A matrix $A \in \mathbb{R}^{N \times N}$ is stable if and only if the leading principal minors of $\Omega(A)$ are negative.

For example, letting $P(x) = x^3 + a_1 x^2 + a_2 x + a_3$ be a monic polynomial, the Routh-Hurwitz conditions imply that $P$ will have stable roots if the following three inequalities are satisfied: (1) $a_1 > 0$, (2) $a_3 > 0$, and (3) $a_1 a_2 > a_3$.

### 2.4 Computational Neuroscience

This section consists of two subsections. In the first one, we define what a firing rate model of a recurrent neuronal network is. Along the way, we present some notational conventions we will be using throughout the manuscript. In the second section, we motivate the notion of a neural code and we define combinatorial neural codes. We also introduce receptive field (RF) codes and relevant terminology.

#### 2.4.1 Firing Rate Models of Neuronal Networks

As explained earlier, neurons in networks can have synapses. Synapses can *excitatory* or *inhibitory*; in the former case, the neuron on the post-synaptic terminal will be more likely to fire, and in the latter case the opposite in the case. A network of neurons can be thought of as a directed graph—in the graph-theoretic sense—and the strengths as well as the kinds of synapses can be summarized in a matrix called a *synaptic weight matrix* [19]. If $W = (w_{ij}) \in \mathbb{R}^{N \times N}$ is a synaptic weight matrix, then $w_{ij}$ denotes the
effective connection strength of the synapse of neuron \( j \) onto neuron \( i \). If \( w_{ij} > 0 \), then the synapse is effectively excitatory, and it is effectively inhibitory when \( w_{ij} < 0 \). A synaptic weight matrix can convey some biological features of the associated neuronal network. For instance, a feedforward network is a neuronal network in which cycles cannot occur. If one chooses carefully the labeling of the neurons in the network, the synaptic weight matrix will be upper triangular. Away from that extreme, recurrent networks admit cycles, so the resulting synaptic weight matrix cannot be made upper triangular.

The rate of action potentials of a recurrent neural network can be described by firing-rate models, where the network consists of neuron-like units whose outputs are firing rates. This is in contrast to spiking models, in which the output of neuron-like units is a spike. Two advantages of firing-rate models is that they avoid the short-time-scale dynamics required to simulate action potentials, and they allow us perform analytic calculations of some aspects of network dynamics [19]. However, firing-rate models have drawbacks. For instance, these models cannot account for aspects of spike timing and correlations, and they are restricted to situations in which the firing of neurons in a given network are uncorrelated, and with little synchrony [19].

A threshold-linear network [20] is a recurrent neuronal network model where the dynamics of each neuron is given by

\[
\tau_i \dot{x}_i + x_i = \sum_{j=1}^{N} w_{ij} x_j + b_i ,
\]

where:

- \([\cdot]_+\) is a rectification nonlinearity, \([x]_+ = \max\{x, 0\}\), where \(x \in \mathbb{R}\);
- \(x_i(t)\) is the firing rate of neuron \( i \) at time \( t \);
- \(b_i\) is the input current to the \( i \)th neuron;
- \(\tau_i > 0\) is the \( i \)th neuron’s time scale; and
- \(w_{ij}\) is the effective strength of the synapse of \( j \)th neuron onto the \( i \)th neuron.
The above system of \( N \) differential equations can be written more compactly as

\[
D \dot{x} + x = [W x + b]_+ ,
\]

where \( D = (d_{ij}) \in \mathbb{R}^{N \times N} \) is the diagonal matrix with \( d_{ii} = \tau_i, \dot{x} = (\dot{x}_1, \dot{x}_2, \ldots, \dot{x}_N)^T, x = (x_1, x_2, \ldots, x_N)^T, b = (b_1, b_2, \ldots, b_N)^T \in \mathbb{R}^N, \) and \( W = (w_{ij}) \in \mathbb{R}^{N \times N} \) is the network’s synaptic weight matrix. When \( v = (v_1, v_2, \ldots, v_N) \in \mathbb{R}^N, \) we define \( [v]_+ = ([v_1]_+, [v_2]_+, \ldots, [v_N]_+) \).

In general, a firing rate model of \( N \) neurons has the form

\[
D \dot{x} + x = \Phi(W x + b),
\]

where \( W \in \mathbb{R}^{N \times N} \) is a synaptic weight matrix; \( D \in \mathbb{R}^{N \times N} \) is a diagonal matrix of time constants; \( b \in \mathbb{R}^N \) is a vector that is interpreted as an external stimulus to the network, which could be constant or changing in time; for us, inputs will be constant in time; and \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) is a \( C^1 \) function. When \( v = (v_1, v_2, \ldots, v_N) \in \mathbb{R}^N, \) we define \( \Phi(v) = (\Phi(v_1), \Phi(v_2), \ldots, \Phi(v_N)) \). In the computational neuroscience and machine learning communities, \( \Phi \) is known as an activation function.

2.4.2 Neural codes

In this subsection, we have multiple objectives. First, we explain what is meant by neural coding and a neural code. We introduce combinatorial neural codes and codes of an open cover. The latter notion will enable us to define receptive field (RF) codes. We also introduce convex RF codes, which will be important in this manuscript. We end the section by defining simplicial complexes, maximal intersection-complete codes, and a known theorem that says that maximal intersection-complete codes are convex RF codes. We also define the notion of a rank-\( r \) network.

It is believed in neuroscience, as a result of experiments that can be traced back to the 1700s, that our perception of the world is constructed from sequences of spikes sent from our sensory nerves to the brain [21]. The measurement and characterization of how stimulus attributes are represented by these patterns of spikes is the study of neural coding. There are two “directions” in neural coding: (1) Neural encoding is the map from stimulus to neural response; and (2) neural decoding is the map from neural response to stimulus [19].

There are a variety of ways to analyze the resulting neural code [19]. For example,
rate coding refers to the principle that the number of spikes in a fixed time window following the onset of a static stimulus represents the intensity of that stimulus. This idea emerged from E. D. Adrian’s observations that the rate of neurons’ spiking increases in response to a static stimulus, e.g., a continuous load on a stretch receptor [21]. There is also the notion of temporal coding (from which a temporal code is derived), where the issue is how precisely spike times must be measured in order to extract most of the information from a neural response [12]. Hence, depending on the neural phenomena being considered by a researcher, the term “neural code” can mean different ideas; however, their common thread is that there is a response in a neuron, or group of neurons, that the researcher can associate to some variable.

Given that there is a zoo of neural codes, we will focus on one kind of them in this dissertation. A combinatorial neural code $\mathcal{C} \subseteq 2^{|N|}$, where $|N| = \{1, 2, \ldots, N\}$, referred to hereinafter as “neural code,” can be thought of as a collection of binary string of 0s and 1s, where 1 means that the neuron is active and 0 means that the neuron is silent. Since there is an obvious bijection between subsets of $|N|$ and binary strings of length $N$, we will go back and forth between the terms “codewords” and “binary strings. A codeword is any element of a neural code.

Previous work on combinatorial codes include [22], where the authors construct a combinatorial code by making windows of 8 msecs and determining which neurons are firing in that window and which ones are not. The collection of codewords obtained in such a fashion would be a combinatorial code (and an example of a temporal code, since the 8 ms window was chosen so as to extract as much information as possible from the spiking neurons). Combinatorial neural codes can also be naturally associated to receptive-field overlaps arising from place fields and tuning curves [9]. When these overlaps occur, the stimulus that falls in that overlap makes the corresponding neurons fire at a high frequency.

Let $X$ be a topological space. From a mathematical point of view, define a receptive field as a map $f_i : X \to \mathbb{R}_{\geq 0}$ from a stimuli space $X$ to the average firing rate of a neuron $i$ in response to each stimulus [9]. By abuse of language, we will often refer to the support of $f_i$ as the receptive field of $i$. Let $\mathcal{U} = \{U_1, \ldots, U_N\}$ be a collection of
open subsets of $X$. For a given subset $\sigma$ of $[N]$, let $U_\sigma$ denote $\cap_{i \in \sigma} U_i$. The code of $\mathcal{U}$ is

$$
\mathcal{C}(\mathcal{U}) = \left\{ \sigma \subseteq [N] : U_\sigma \setminus \bigcup_{i \in [N]\setminus \sigma} U_i \neq \emptyset \right\}.
$$

The idea is that $\mathcal{C}(\mathcal{U})$ is the collection of all possible codewords that can be elicited by presenting a stimulus and looking at the combinations of receptive fields that contain the stimulus. Let us consider Figure 2.2 as an example illustrating this idea: We have three neurons $\nu_1, \nu_2,$ and $\nu_3$ with receptive fields $U_1, U_2, U_3$, respectively, that are subsets of a stimulus space $X$. If a stimulus is chosen from $X$ so that it is not contained in any of the receptive fields (which in this case is possible because $U_1, U_2, U_3$ do not cover the entire stimulus space), then none of the neurons will be activated. This would correspond to $\emptyset \in \mathcal{C}(\mathcal{U})$ (or, as a binary string, $000 \in \mathcal{C}(\mathcal{U})$. On the other hand, if a stimulus is chosen so that it lies inside $U_2$, then $\nu_1$ and $\nu_2$ would become active. Hence, $\{1, 2\} \in \mathcal{C}(\mathcal{U})$ or, equivalently, $110 \in \mathcal{C}(\mathcal{U})$. Observe that $\{2\} \notin \mathcal{C}(\mathcal{U})$ because $\nu_2$ can never be made active without triggering $\nu_1$. It is easy to see that $\mathcal{C}(\mathcal{U}) = \{\emptyset, \{1\}, \{3\}, \{1, 2\}, \{1, 3\}\}$. Throughout the manuscript, we will often abbreviate the notation for subsets by omitting the braces and commas. So, for instance, $\mathcal{C}(\mathcal{U})$ would be equivalently expressed as $\mathcal{C}(\mathcal{U}) = \{0, 1, 3, 12\}$.

![Figure 2.2: Three convex receptive fields, $U_1, U_2$ and $U_3$ associated to neurons $\nu_1, \nu_2$, and $\nu_3$. The box, denoted $X$, is the stimuli space.](image)

We say that $\mathcal{C}(\mathcal{U})$ is **convex** if $\mathcal{U}$ consists of convex sets. Additionally, $\mathcal{C}(\mathcal{U})$ is
open convex if \( U \) consists entirely of open convex sets, and **closed convex** if every member of \( U \) is a closed convex set. One reason to insist that RF codes be convex, in addition to its biological plausibility, is that any combinatorial code can be realized as an RF code \([9]\). However, not every RF code can be realized by convex, open sets; for example, \( C = \{\emptyset, 1, 2, 3, 4, 23, 24, 123, 124\} \) is a nonconvex code (due to Vladimir Itskov, personal communication). We prove that \( C \) is nonconvex in Appendix A.2.

Next we state a result that will enable us to conclude that, under suitable conditions, the combinatorial neural codes we introduce in chapter 4 are convex codes. An abstract **simplicial complex** \([23]\), or simplicial complex for brevity, \( \Delta \) on \( [N] = \{1, 2, \ldots, N\} \) is a collection of subsets of \( [N] \) such that for any \( \sigma \in \Delta \), if \( \tau \subseteq \sigma \), then \( \tau \in \Delta \). In other words, a simplicial complex is finite family of sets that is closed under taking subsets. For example, \( C(U) = \{\emptyset, \{1\}, \{1, 2\}, \{3\}, \{1, 3\}\} \) is not a simplicial complex because \( \{2\} \notin C(U) \). Note that \( \{2\} \cup C(U) \) is a simplicial complex. Given a codeword \( c \) in a neural code \( C \subseteq \{0, 1\}^N \), the support \([9]\) of \( c \) is \( \text{supp}(c) = \{i \in [N] : c_i = 1\} \). The **simplicial complex of \( C \)** is \( \Delta(C) = \{\sigma \subseteq [N] : \text{for some } c \in C, \sigma \subseteq \text{supp}(c)\} \).

We call a combinatorial neural code \( C \) **intersection complete** if and only if for every \( \sigma, \omega \in C \), we have that \( \sigma \cap \omega \in C \). A **maximal codeword** is a codeword that is not a subpattern of any other codeword. In terms of subsets, a maximal codeword is simply a maximal subset. So, for example, in \( C(U) = \{\emptyset, \{1\}, \{1, 2\}, \{3\}, \{1, 3\}\} \), the maximal codewords are 12 and 13 (where we are again using the convention for abbreviating subsets). Letting \( \Delta = \Delta(C) \) be the simplicial complex of \( C \), define the set of **max intersections** of \( \Delta \) as

\[
\mathcal{F}_\cap(\Delta) = \left\{ \bigcap_{i=1}^k F_i : F_1, \ldots, F_k \text{ are maximal codewords of } C \right\} \cup \{\emptyset\}.
\]

Say that \( C \) is **maximal intersection-complete** if and only if \( C \supseteq \mathcal{F}_\cap(\Delta) \). In our running example, \( C(U) = \{\emptyset, 1, 3, 12, 13\} \), we see that

\[
\Delta(C(U)) = \{\emptyset, 1, 2, 3, 12, 13\},
\]

\[
\mathcal{F}_\cap(\Delta) = \{\emptyset, 1, 12, 13\},
\]

so \( C(U) \) is maximal intersection-complete. It is also intersection complete.
Lastly, although we will tie together permitted sets and convex coding in the next chapter, we introduce some more terminology. Let $W \in \mathbb{R}^{N \times N}$ be a synaptic weight matrix. If $\text{rk}(W) = r$, then we call $D\dot{x} + x = \Phi(Wx + b)$ a rank-$r$ network. A combinatorial code $C$ associated with a rank-$r$ network will be referred to as a rank-$r$ code. Later we will devote our attention to rank-1 codes.
Chapter 3

Overview of Previous Work

This Chapter has two sections. The first defines what permitted sets are and states known results about the combinatorial constraints permitted sets satisfy; these results led us to develop the nesting property of neural codes, which we discuss in Chapter 5. We also go through two examples of finding the collection of all possible permitted sets of a threshold-linear network with three neurons. The section that follows is brief discussion about permitted sets and convex codes. The main point there is that there are collections of permitted sets that are not convex RF codes.

3.1 Permitted Sets

Permitted sets were first formulated [1, 20] in the context of threshold-linear networks, $D\dot{x} + x = [b + Wx]_+$. A permitted set of a threshold-linear network is a subset $\sigma \subseteq [N] = \{1, 2, \ldots, N\}$ such that there is a stimulus $b \in \mathbb{R}^N$ for which there is an asymptotically stable fixed point $x^* \in \mathbb{R}^N_{\geq 0}$ of with $\operatorname{supp}(x^*) = \sigma$. (Here $\operatorname{supp}(x^*) = \{i \in [N] : x^*_i > 0\}$.) We will denote the collection of all permitted sets by $\mathcal{P}_{[1]+}(W)$. Note that $\mathcal{P}_{[1]+}(W)$ is a combinatorial neural code.

A first question to ask is, given $D\dot{x} + x = [b + Wx]_+$, how one can determine what $\mathcal{P}_{[1]+}(W)$ is. The following result shows the correspondence between stable principal submatrices and permitted sets. Recall that we use the notation $A_\sigma$, where $A \in \mathbb{R}^{N \times N}$ and $\sigma \subseteq [N]$, to denote the principal submatrix of $A$ determined by $\sigma$:  

20
Theorem 2 \([24]\). Consider the threshold-linear network \(D\dot{x} + x = [Wx + b]_+\). A subset of neurons \(\sigma\) is a permitted set of \(D\dot{x} + x = [Wx + b]_+\) if and only if the principal submatrix \((W - I)_\sigma\) is stable.

The above theorem says how we can go from the activity patterns described by the dynamics of \(D\dot{x} + x = [Wx + b]_+\) (as modeled by asymptotically stable fixed for given fixed choice of \(b \in \mathbb{R}^N\)) to \(\mathcal{P}_{[+]_\sigma}(W)\). A natural question, then, is whether given a combinatorial neural code \(\mathcal{C} \subseteq 2^{[N]}\), it is possible whether \(\mathcal{C}\) arises as the collection of activity patterns of a threshold-linear network. In other words, given \(\mathcal{C}\), how can we tell whether there is \(W \in \mathbb{R}^{N \times N}\) such that the permitted sets of \(D\dot{x} + x = [Wx + b]_+\) satisfy \(\mathcal{P}_{[+]_\sigma}(W) = \mathcal{C}\). The following result shows a combinatorial constraint that \(\mathcal{P}_{[+]_\sigma}(W)\) satisfies when \(W\) is a symmetric matrix:

**Proposition 1** \([20]\). Let \(W \in \mathbb{R}^{N \times N}\) be a symmetric matrix. Then the collection of all permitted sets of \(D\dot{x} + x = [Wx + b]_+\) is a simplicial complex.

Drawing again from our running example from the end of the last chapter, \(\mathcal{C} = \{\emptyset, 1, 3, 12, 13\}\) (where we write “12” instead of “\{1, 2\},” etc), we see that \(\mathcal{C}\) cannot possibly be a collection of permitted sets of a threshold-linear network whose \(W\) is symmetric because \(\mathcal{C}\) is not closed under taking subsets, i.e., \(\{2\} \notin \mathcal{C}\), so \(\mathcal{C}\) is not a simplicial complex.

Next we find all the possible permitted sets of a threshold-linear network with three neurons when \(W\) is symmetric:

**Example 1.** We will determine \(\mathcal{P}_{[+]_\sigma}(W)\) when \(W\) is assumed to be a symmetric matrix. In other words, we will consider every possible principal submatrix of \(W - I\), where

\[
W - I = \begin{pmatrix}
-1 & w_{12} & w_{13} \\
w_{12} & -1 & w_{23} \\
w_{13} & w_{23} & -1
\end{pmatrix},
\]

and determine which ones are stable, i.e., which principal submatrices have spectra in which the real parts of eigenvalues are negative.

In order to do this, we go through every possible subset \(\sigma\) of \(\{1, 2, 3\}\), starting with the singletons, and determine for what values of \(w_{iij}\) the corresponding principal
submatrix is stable. In what follows, we will be using the Routh-Hurwitz Criterion for stability.

- Suppose $\sigma$ is a singleton, that is, $\sigma = \{1\}, \{2\}$ or $\{3\}$. For each one of these subsets, $(W - I)_\sigma = [-1]$, which has $-1$ as an eigenvalue, so these submatrices are stable, which implies that $1, 2, 3 \in P_{[1]}^+(W)$. Notice, in particular, that since this computation does not depend on the synaptic weights, the singletons $1, 2, 3$ are elements of any $P_{[1]}^+(W)$.

- Suppose $\sigma$ is a 2-subset, i.e., $\sigma = \{1, 2\}, \{1, 3\}$, or $\{2, 3\}$. Let us start with $\{1, 2\}$, in which case

$$(W - I)_{\{1, 2\}} = \begin{pmatrix} -1 & w_{12} \\ w_{12} & -1 \end{pmatrix},$$

which has eigenvalues $-1 \pm w_{12}$. Therefore, $(W - I)_{\{1, 2\}}$ is stable if and only if $w_{12} \in (-1, 1)$. A similar analysis can be performed for $\{1, 3\}, \{2, 3\}$.

- Finally, consider $\sigma = \{1, 2, 3\}$. Here $(W - I)_\sigma = W - I$. In order for $(W - I) = 0$ to be satisfied,

$$\lambda^3 + 3\lambda^2 + (3 - w_{12}^2 - w_{13}^2 - w_{23}^2)\lambda + (1 - w_{12}^2 - w_{13}^2 - w_{23}^2 - 2w_{12}w_{13}w_{23}) = 0.$$

According to the Routh-Hurwitz Criterion, the constraints that must be satisfied by $w_{12}, w_{13}, w_{23}$ in order for $W - I$ to be stable are

$$1 > w_{12}^2 + w_{13}^2 + w_{23}^2 + 2w_{12}w_{13}w_{23} \quad (3.1)$$

and

$$4 > w_{12}^2 + w_{13}^2 + w_{23}^2 - w_{12}w_{13}w_{23}. \quad (3.2)$$

Now we verify that there are choices of synaptic weights such that actually satisfy these constraints, as well as choices that satisfy one but not the other. First, when $w_{12} = w_{13} = w_{23} = 0$ or $w_{12} = w_{13} = 1/2$ and $w_{23} = -1/2$ are synaptic weights
that satisfy both (3.1) and (3.2). However, it is also possible to pick weights that violate one of the inequalities. For example, if \( w_{12} = w_{13} = w_{23} = \frac{8}{9} \), then

\[
\frac{w_{12}^2 + w_{13}^2 + w_{23}^2 + 2w_{12}w_{13}w_{23}}{27} > 1,
\]

so (3.1) is not satisfied, in which case \( W - I \) is not stable.

This is the first step in our analysis because all we have done so far is determine conditions on the synaptic weights such that a given subset \( \sigma \subseteq [N] \) is permitted. Now we want to know what are all possible collections of permitted sets. In order to deduce this from our computations, we choose synaptic weights so that certain subsets are in \( P_{\{\cdot\}}(W) \) and others are not. (Below we follow the convention of writing \( \{1, 12\} \) instead of \( \{\{1\}, \{1, 2\}\} \), etc.) Here are two examples:

- The choice of \( W \) such that \( P_{\{\cdot\}}(W) = \{1, 2, 3\} \) is when its entries are \( w_{12}, w_{13}, w_{23} \in (-1, 1) \) because \( 1, 2, 3 \) are always permitted sets and \( \{i, j\} \notin P_{\{\cdot\}}(W) \) if and only if \( w_{ij} \notin (-1, 1) \). (Observe that the choices of \( w_{12}, w_{13}, w_{23} \) ensure that \( 123 \) is not a permitted set.)

- The choices of synaptic weights such that \( P_{\{\cdot\}}(W) = \{1, 2, 3, 12\} \) are \( w_{12} \in (-1, 1) \) and \( w_{13}, w_{23} \notin (-1, 1) \) because \( 12 \) is permitted if and only if \( w_{12} \in (-1, 1) \), whereas \( 13, 23 \) are not permitted if and only if \( w_{13}, w_{23} \notin (-1, 1) \). (Just as in the first examples, the values of \( w_{13}, w_{23} \) prevent \( 123 \) from being a permitted set.)

An identical analysis can be performed in order to list every possible \( P_{\{\cdot\}}(W) \) given a suitable \( W \). Notice, in particular, that \( P(W) \) is in each case a simplicial complex—that is, a family of subsets that is closed under taking subsets.

Now that we are done with the example, let us consider what happens when \( W \) is not symmetric. The next result gives us a criterion for deciding whether a given neural code \( C \subseteq 2^{[N]} \) could be realized as a collection of permitted sets of a threshold-linear network:

**Lemma 6** (2-by-2 Minor Lemma, [25]). Let \( A \in \mathbb{R}^{N \times N} \) with strictly negative diagonal and \( N \geq 2 \). If \( A \) is stable, then there exists a stable \( 2 \times 2 \) principal submatrix of \( A \).
While the proof is elementary, for the sake of completeness, we present a proof.

**Proof.** Recall that in Section 2.2 we that the characteristic polynomial of $A$ is

$$\chi_A(t) = \sum_{i=0}^{N} (-1)^{N-i} E_i(A)t^{N-i}.$$ 

Since $A$ is stable, $(-1)^k E_k(A) > 0$; in particular, $E_2(A) > 0$. Hence, there is at least one $2 \times 2$ principal minor that is positive. The corresponding matrix has negative diagonal, so it is stable by the Routh-Hurwitz Criterion.

As a consequence of Lemma 6, given a neural code $C \subseteq 2^{[N]}$, where $[N] = \{1, 2, \ldots, N\}$, if there is a codeword containing no subpattern consisting of two coactive neurons, then $C$ is not realizable as $P_{\lceil \cdot \rceil_+}(W)$ for any threshold-linear network $W$. For instance, $C = \{(1), (1, 2, 3)\}$ cannot be realized as $P_{\lceil \cdot \rceil_+}(W)$ for any $W$ at all because $\{1, 2\}, \{1, 3\}, \{2, 3\} \notin C$.

Just as in the symmetric regime, we explicitly list all the possible permitted sets of a threshold-linear network with three neurons when $W$ is non-symmetric:

**Example 2** (Collection of all permitted sets when $W$ is non-symmetric.). We will determine the possible $P_{\lceil \cdot \rceil_+}(W)$ when

$$W - I = \begin{pmatrix} -1 & w_{12} & w_{13} \\ w_{21} & -1 & w_{23} \\ w_{31} & w_{32} & -1 \end{pmatrix}.$$ 

Just as in the first example, we will determine what conditions need to be imposed on the synaptic weights in order for the principal submatrices of $W - I$ to be stable.

- Suppose $\sigma = \{1\}, \{2\}, \text{ or } \{3\}$. Then $(W - I)_\sigma = [-1]$ is stable regardless of the values of the synaptic weights.

- Suppose $\sigma = \{1, 2\}, \{1, 3\}$ or $\{2, 3\}$. If $\sigma = \{1, 2\}$, then

$$\det((W - I)_\sigma - \lambda I) = \lambda^2 + 2\lambda + 1 - w_{ij}w_{ji} = 0,$$
so \( \lambda = -1 \pm \sqrt{w_{ij}w_{ji}} \). If \( w_{ij} \) and \( w_{ji} \) have different signs, or if one of them is zero, then the eigenvalues of \((W - I)_{\sigma}\) have negative real part. On the other hand, if \( w_{ij}, w_{ji} \) have the same sign, then \( \lambda < 0 \) whenever \( w_{ij}w_{ji} < 1 \). Therefore, \((W - I)_{\sigma}\) will be stable when \( w_{ij}w_{ji} < 1 \).

- Finally, suppose \( \sigma = \{1, 2, 3\} \). The eigenvalues of \((W - I)_{\sigma} = W - I\) are the roots of the polynomial (in \( \lambda \))

\[
\lambda^3 + 3\lambda^2 + (3 - w_{12}w_{21} - w_{13}w_{31} - w_{23}w_{32})\lambda \\
+ (1 - w_{12}w_{21} - w_{13}w_{31} - w_{23}w_{32} - w_{12}w_{23}w_{31} - w_{13}w_{21}w_{32}).
\]

The Routh-Hurwitz Criterion says that \( W - I \) will be stable when

\[
1 > w_{12}w_{21} + w_{13}w_{31} + w_{23}w_{32} + w_{12}w_{23}w_{31} + w_{13}w_{21}w_{32} \tag{3.3}
\]

and

\[
8 > 2(w_{12}w_{21} + w_{13}w_{31} + w_{23}w_{32}) - (w_{12}w_{23}w_{31} + w_{13}w_{21}w_{32}). \tag{3.4}
\]

Just as in our previous example when \( W \) is symmetric, the above constraints can be put together in order to rule out which collections subsets of \( \{1, 2, 3\} \) can be realized as \( P[\cdot, +](W) \). First off, we observe that unlike the case when \( W \) is symmetric, it is possible to come up with weights such that \( W - I \) is stable, yet there is a subset \( \sigma \subseteq \{1, 2, 3\} \) with \((W - I)_{\sigma}\) not stable. For instance, consider \( w_{12} = w_{21} = \sqrt{2} \), so that \((W - I)_{\sigma}\) (with \( \sigma = \{1, 2\} \)) is not stable. Let

\[
w_{13} = w_{23} = -1 \quad \text{and} \quad w_{31} = w_{32} = 1,
\]

so the submatrices associated to the subsets \( \{1, 3\} \) and \( \{2, 3\} \) are stable. This choice of weights satisfy (3.3) and (3.4). Hence, the permitted sets in this case are \( P[\cdot, +](W) = \{1, 2, 3, 13, 23, 123\} \). In particular, \( P[\cdot, +](W) \) is not a simplicial complex in general when \( W \) is not symmetric.

Finally, it is also possible to come up with weights so that \( P[\cdot, +](W) = \{1, 2, 3, 23, 123\} \).
For instance, 

\[ w_{12} = w_{21} = w_{13} = w_{31} = \sqrt{2}, \quad w_{23} = 1, \quad w_{32} = -2, \]

so \((W - I)\sigma\) is not stable when \(\sigma = \{1, 2\}\) and \(\sigma = \{1, 3\}\), but \((W - I)_{\{2,3\}}\) is stable. To summarize, the possible \(P[\cdot]+(W)\) consists of the same list we showed when \(W\) is symmetric, as well as the following six collections of subsets:

\[
\{1, 2, 3, 13, 23, 123\}, \{1, 2, 3, 12, 13, 123\}, \{1, 2, 3, 12, 23, 123\}, \\
\{1, 2, 3, 12, 123\}, \{1, 2, 3, 13, 123\}, \{1, 2, 3, 23, 123\}.
\]

### 3.2 Convex Codes

Recall that in Section 2.4.2 we brought up the code \(C = \emptyset, 1, 2, 3, 4, 23, 24, 123, 124\) as an instance of a nonconvex code. Note that Lemma 6 will not help us figure out whether there exists \(W \in \mathbb{R}^{4 \times 4}\) such that \(P(W) = C\). The reason why is because \(23 \subset 123\) as well as \(24 \subset 124\), and \(23, 24 \in C\). It turns out that \(C\) can be realized as a collection of permitted sets of a threshold-linear network: Let

\[
W = \begin{pmatrix}
0 & 0.5 & 2 & 1 \\
2 & 0 & -0.1 & -0.9 \\
0.5 & -0.7 & 0 & 0.5 \\
1 & -0.7 & 2 & 0
\end{pmatrix}.
\]

Then one can verify that \(P[\cdot]+(W) = C\). Therefore, \(C\) is an example of a collection of permitted sets that is a non-convex RF code.

The strength of the recurrent connection from neuron \(j\) to neuron \(i\) is denoted by \(w_{ij}\). Dale’s law says [19] that neurons are either purely excitatory or inhibitory, i.e., for a fixed \(j\), for all \(i \in [N] \setminus \{j\}\), we have that either \(w_{ij} > 0\) or \(w_{ij} < 0\). So if a synaptic weight matrix \(W\) satisfies Dale’s Law, then entries on the same column—possibly leaving the diagonal out, which is typically 0 by assumption—have the same sign. A natural question at this juncture is if we require \(W\) to satisfy Dale’s law, can we rule out nonconvex permitted sets? The following synaptic weight matrix satisfying
Dale’s law shows there are collections of permitted sets that do not form convex codes:

\[ W = \begin{pmatrix} 0 & -0.15 & 2.16 & -0.42 \\ 1.86 & 0 & 2.58 & -2.52 \\ 1.77 & -1.35 & 0 & -0.42 \\ 1.98 & -0.72 & 2.19 & 0 \end{pmatrix} \]

Then \( \mathcal{P}_{[\cdot]}(W) = \{\emptyset, 1, 2, 3, 4, 12, 14, 23, 34, 123, 134\} \). We prove in Appendix A.2 that \( \mathcal{P}_{[\cdot]}(W) \) is nonconvex.

The next result says that maximal intersection-complete codes are convex:

**Theorem 3** (Theorem 4.4, [26]). *Let \( \mathcal{C} \) be a neural code on \( N \) neurons. If \( \mathcal{C} \) is maximal intersection complete, then \( \mathcal{C} \) is open and closed convex.*
Chapter 4

Convex Coding in Firing Rate Models

This is the first chapter in which we present our new contribution to the study of permitted sets. Our main concern will be showing that a combinatorial neural code (the permitted sets) can be associated to a firing rate model and that such a code exhibits convex coding, i.e., the neural code arises, roughly speaking, from a pattern of intersections of receptive fields of the neurons in the network. We first present results when the activation function $\Phi$ is a rectifier, $\Phi(x) = \max(x, 0)$. Then we proceed to prove analogous results for $C^1$ activation functions with finitely many singularities.

One take-away from this chapter is that although it is not plausible that the synaptic weight matrix of neuronal networks is rank one, it is a regime in which neurobiologically relevant properties can be theoretically addressed, as is the case with convex coding. We remark that rank-one synaptic weight matrices have been considered in other theoretical analyses of neuronal networks. For example, for the threshold-linear networks, rank-one synaptic weight matrices are special in that there is a learning algorithm for which the network can learn and unlearn combinatorial neural codes with maximal flexibility [24]. Rank-one synaptic weight matrices also play a role for the theoretical analysis in the encoding of combinatorial neural codes for threshold-linear networks [25]. Therefore, it is not unreasonable to start analyzing the relationship between combinatorial neural codes and synaptic weight matrices by assuming that said matrices are rank one (and
then work with perturbations in order to explore how robust the results are).

## 4.1 Threshold-Linear Case

First we prove that given a threshold-linear network $\dot{x} + x = [Wx + b]_+$, where $W$ is a rank-one matrix, has $\mathcal{P}_{[\cdot]}(W)$ that is a convex code. Recall (Section 2.4.2) that a maximal intersection-complete code is a combinatorial neural code $\mathcal{C} \subseteq 2^{|N|}$ such that intersections of any two maximal subsets is also a codeword.

**Proposition 2.** Let $W = uv^T$, where $u, v \in \mathbb{R}^N \setminus \{0\}$, and $D \in \mathbb{R}^{N \times N}$ a diagonal matrix of time constants. Let $\mathcal{P}_{[\cdot]}(W)$ be the collection of permitted sets of $\dot{x} + x = [Wx + b]_+$. If $\mu$ and $\sigma$ are two codewords in $\mathcal{P}_{[\cdot]}(W)$, i.e., $\mu, \sigma \in \mathcal{P}_{[\cdot]}(W)$, where $\mu$ is a maximal codeword, then $\mu \cap \sigma$ is also a codeword in $\mathcal{P}_{[\cdot]}(W)$. In particular, $\mathcal{P}_{[\cdot]}(W)$ is maximal intersection-complete, so $\mathcal{P}_{[\cdot]}(W)$ is convex by Theorem 3.

**Proof.** Since $\mu$ is maximal, for any $i \notin \mu$, we have that $\mu \cup \{i\}$ is not a codeword, so $\mu \cup \{i\}$ is not a permitted set. By Theorem 2 we know that $\mu \cup \{i\}$ is not permitted if and only if $(W - I)_{\mu \cup \{i\}}$ is not a stable matrix, that is, there is an eigenvalue of $(W - I)_{\mu \cup \{i\}}$ whose real part is nonnegative. By Lemma 3,

$$\text{Spec}((W - I)_{\mu \cup \{i\}}) = \left\{ u^T_{\mu \cup \{i\}}v_{\mu \cup \{i\}} - 1, -1 \right\}.$$

The only possibility, then, for $(W - I)_{\mu \cup \{i\}}$ to not be stable is for $u^T_{\mu \cup \{i\}}v_{\mu \cup \{i\}} - 1 \geq 0$ to hold. This implies that $u_i v_i > 0$ for any $i \notin \mu$.

Let $\tau = \mu \cap \sigma$. Suppose that $\tau$ is a proper subset of $\sigma$. Observe that

$$u^T_{\tau}v_{\tau} - 1 = u^T_{\sigma}v_{\sigma} - 1 - \sum_{k \in \sigma \setminus \tau} u_kv_k$$

$$< u^T_{\sigma}v_{\sigma} - 1$$

$$< 0,$$

where the first inequality holds because $k \in \sigma \setminus \tau$ implies $k \notin \mu$. Hence, $\tau \in \mathcal{P}_{[\cdot]}(W)$. \qed

Earlier, in section 2.4.2 we introduced intersection complete codes. The identifying property is that the intersection of any two codewords will also be a codeword. In
particular, every intersection complete code is maximal intersection-complete. Next we exhibit an example of a rank-one synaptic weight matrix such that $\mathcal{P}_{\subseteq}^+(W)$ is maximal intersection-complete and not intersection complete.

**Example 3.** Let $W = uv^T$, where $u = (2, 1, 1, -0.2)^T$ and $v = (-0.2, 0.9, 0.2, 2)^T$, so that

$$
W = \begin{pmatrix}
-0.4 & 1.8 & 0.4 & 4 \\
-0.2 & 0.9 & 0.2 & 2 \\
-0.2 & 0.9 & 0.2 & 2 \\
0.04 & -0.18 & -0.04 & -0.4
\end{pmatrix}.
$$

Then we have that $123, 124 \in \mathcal{P}_{\subseteq}^+(W)$ because

$$
u_{123}^T v_{123} = -0.4 + 0.9 + 0.2 < 1, \text{ and } u_{234}^T v_{234} = 0.9 + 0.2 - 0.4 < 1,
$$

whereas $123 \cap 234 = 23 \notin \mathcal{P}_{\subseteq}^+(W)$ because

$$
u_{23}^T v_{23} = 0.9 + 0.2 > 1.
$$

Note that this does not contradict Proposition[2] because $1234 \in \mathcal{P}_{\subseteq}^+(W)$.

Another natural question that one can ask is whether $\mathcal{P}_{\subseteq}^+(W)$ is necessarily convex when $W$ is rank 2. The following example shows that this is not the case.

**Example 4.** Let $W = xy^T + uv^T$, where $x, y, u, v \in \mathbb{R}^4$ such that

$$
x = \begin{pmatrix} 8.3, 2, -8, 3.1 \end{pmatrix},
$$

$$
y = \begin{pmatrix} -3.6, 1, 8.1, -6 \end{pmatrix},
$$

$$
u = \begin{pmatrix} -9, -3.6, -6.1, -7 \end{pmatrix},
$$

$$
v = \begin{pmatrix} 0, 4, -8.6, 6.6 \end{pmatrix},
$$
so we have that
\[
W = \begin{pmatrix}
-29.88 & -27.7 & 144.63 & -109.2 \\
-7.2 & -12.4 & 47.16 & -35.76 \\
28.8 & -32.4 & -12.34 & 7.74 \\
-11.16 & -24.9 & 85.31 & -64.8
\end{pmatrix}.
\]

Then \( \mathcal{P}[\cdot]_+(W) = \{\emptyset, 1, 2, 3, 4, 12, 14, 23, 34, 124, 234\} \). Note that \( \mathcal{P}[\cdot]_+(W) \) is not maximal-intersection complete. An elementary argument shows that \( U_2 \cap U_4 \) is disconnected, so either \( U_2 \) or \( U_4 \) is not convex.

A reasonable question to ask is whether \( \mathcal{P}[\cdot]_+(W) \) is convex when, in some sense, \( W \) is close to a rank-one matrix. This is indeed the case, so next we set to prove a perturbed version of Proposition 2. Before doing so, we need the following lemma. It says that maximal codewords must contain the negative diagonal entries of \( W \). Roughly speaking, neurons whose self-coupling strength is effectively inhibitory participate in the formation of maximal neural ensembles.

**Lemma 7.** Let \( W = u v^T \), where \( u, v \in \mathbb{R}^N \setminus \{0\} \). Define \( [N]_{<0} \) to be the collection of all \( i \in [N] = \{1, 2, \ldots, N\} \) such that \( u_i v_i < 0 \) and suppose that \( [N]_{<0} \neq \emptyset \). Then there is a neighborhood \( \mathcal{U} \) of \( 0 \in \mathbb{R}^{N \times N} \) such that if \( \delta \in \mathcal{U} \), then \( [N]_{<0} \subseteq \mu \) for any \( \mu \in \mathcal{P}[\cdot]_+(W + \delta) \) that is a maximal permitted set.

**Proof.** Let
\[
0 < \epsilon < \min \left( 1, \min_{i \in [N]_{<0}} \frac{|u_i v_i|}{2} \right).
\]
For any \( \sigma \in 2^[[N]] \setminus \emptyset \), by continuity of \( \text{Spec} : \mathbb{C}^{[\sigma] \times [\sigma]} \rightarrow \mathcal{A}_{[\sigma]} \) at \( W_{\sigma} - I_{\sigma} \), there is \( d_{\sigma} > 0 \) such that for any \( X \in \mathbb{C}^{[\sigma] \times [\sigma]} \) with \( ||W_{\sigma} - I_{\sigma} - X||_{\text{max}} < d_{\sigma} \), we have
\[
\min_{\pi \in \mathcal{G}_{[\sigma]}} \max_{1 \leq j \leq [\sigma]} \left| \lambda_j(W_{\sigma} - I_{\sigma}) - \lambda_{\pi(j)}(X) \right| < \epsilon.
\]
Let \( d = \min_{\sigma} d_{\sigma} \) and define \( \mathcal{U} = \{X \in \mathbb{C}^{N \times N} : ||X||_{\text{max}} < d\} \). Let \( \delta \in \mathcal{U} \) and suppose \( \mu \in \mathcal{P}[\cdot]_+(W + \delta) \) is a maximal codeword. Then
\[
||W_{\mu} - I_{\mu} - (W_{\mu} + \delta_{\mu} - I_{\mu})||_{\text{max}} = ||\delta_{\mu}||_{\text{max}}.
\]
Observe that if $A \in \mathbb{C}^{N \times N}$, then $||A_{\sigma}||_{\text{max}} \leq ||A||_{\text{max}}$ for any $\sigma \subseteq [N]$. Hence,

$$||\delta_{\mu}||_{\text{max}} \leq ||\delta||_{\text{max}} < d \leq d_{\mu},$$

so by continuity at $W_{\mu} - I_{\mu}$ and Lemma 5,

$$\text{Spec}(W_{\mu} + \delta_{\mu} - I_{\mu}) = \{u_{\mu}^Tv_{\mu} - 1 + \Delta_1, \Delta_2 - 1, \ldots, \Delta_{|\mu|} - 1 \},$$

where $|\mu|$ denotes the cardinality of $\mu$ and $|\Delta_k| < \epsilon$ for all $k \in \{1, 2, \ldots, |\mu|\}$. Observe that

$$u_{\mu}^Tv_{\mu} - 1 + \text{Re}(\Delta_1) < 0$$

by our assumption that $\mu \in \mathcal{P}_{[\cdot]}(W + \delta)$.

Now we prove $[N]_{< 0} \subseteq \mu$ by way of contradiction: Let $s \in [N]_{< 0}$ be such that $s \notin \mu$ and define $\tilde{\mu} = \mu \cup \{s\}$. Observe that

$$||W_{\tilde{\mu}} - I_{\tilde{\mu}} - (W_{\tilde{\mu}} + \delta_{\tilde{\mu}} - I_{\tilde{\mu}})|| = ||\delta_{\tilde{\mu}}||_{\text{max}} \leq ||\delta||_{\text{max}} < d \leq d_{\tilde{\mu}}.$$

By continuity at $W_{\tilde{\mu}} - I_{\tilde{\mu}}$ and Lemma 5,

$$\text{Spec}(W_{\tilde{\mu}} + \delta_{\tilde{\mu}} - I_{\tilde{\mu}}) = \{u_{\mu}^Tv_{\mu} - 1 + \Delta_1, \Delta_2 - 1, \ldots, \Delta_{|\tilde{\mu}|} - 1 \},$$

where $|\Delta_k| < \epsilon$ for every $k \in \{1, 2, \ldots, |\tilde{\mu}|\}$. Define $\eta = \Delta_1 - \Delta$. Then

$$u_{\mu}^Tv_{\mu} - 1 + \text{Re}(\Delta_1) = u_{\mu}^Tv_{\mu} - 1 + \text{Re}(\eta) + \text{Re}(\Delta_1) + u_s v_s,$$

where we noted that $u_{\mu}^Tv_{\mu} = u_{\mu}^Tv_{\mu} + u_s v_s$. Since $\mu \in \mathcal{P}_{[\cdot]}(W + \delta)$, we have $u_{\mu}^Tv_{\mu} - 1 + \text{Re}(\Delta_1) < 0$. Further, let $s' = \text{argmin}_{s \in [N]_{< 0}} - u_s v_s / 2$, so that $-u_{s'} v_{s'} \leq -u_s v_s$ for every $s \in [N]_{< 0}$. Therefore, $\text{Re}(\eta) < 2\epsilon \leq -u_{s'} v_{s'}$, so

$$\text{Re}(\eta) + u_s v_s < -u_{s'} v_{s'} + u_s v_s \leq 0,$$

so it follows that $u_{\mu}^Tv_{\mu} - 1 + \text{Re}(\Delta_1) < 0$, which implies that $\tilde{\mu} \in \mathcal{P}_{[\cdot]}(W + \delta)$. However, $\mu \subset \tilde{\mu}$ by construction, contradicting maximality of $\mu$. We conclude that $[N]_{< 0} \subseteq \mu$. □
As a consequence of Lemma 7, if \( k \notin \mu \), then \( u_k v_k \geq 0 \). Next we present a result that shows that Proposition 2 is robust: If a synaptic weight matrix \( W \) is a small perturbation of a rank-one matrix, say \( W + \delta \), then the resulting \( \mathcal{P}_{[\cdot]+}(W + \delta) \) will be a convex code just as in the rank-one regime.

**Proposition 3.** Let \( W = uv^T \), where \( u, v \in \mathbb{R}^N \setminus \{0\} \). Suppose that \( u_i v_i \neq 0 \) for all \( i \in [N] = \{1, 2, \ldots, N\} \). Then there is a neighborhood \( U \) of \( 0 \in \mathbb{R}^{N \times N} \) such that if \( \delta \in U \), then \( \mu \cap \sigma \in \mathcal{P}_{[\cdot]+}(W + \delta) \) for any \( \mu, \sigma \in \mathcal{P}_{[\cdot]+}(W + \delta) \), where \( \mu \) is a maximal codeword. In particular, \( \mathcal{P}_{[\cdot]+}(W + \delta) \) is maximal intersection-complete, so \( \mathcal{P}_{[\cdot]+}(W + \delta) \) convex by Theorem 3.

**Proof.** Define \( [N]_{<0} = \{i \in [N] : u_i v_i < 0\} \), \( [N]_{>0} = \{i \in [N] : u_i v_i > 0\} \), and

\[
0 < \epsilon < \min \left(1, \min_{i \in [N]_{<0}} \frac{u_i v_i}{2}, \min_{i \in [N]_{>0}} \frac{u_i v_i}{2}\right).
\]

(Define \( \min(\emptyset) = +\infty \).) For every \( \omega \in 2^{[N]} \setminus \{\emptyset\} \), there is \( d_\omega > 0 \) such that for all \( X \in \mathbb{C}^{\omega \times \omega} \) satisfying \( ||(W - I)_\omega - X||_{\max} < d_\omega \),

\[
\min_{\pi \in \mathcal{S}_|\omega|} \max_{1 \leq j \leq |\omega|} \lambda_j(W_\omega - I_\omega) - \lambda_{\pi(j)}(X) < \epsilon.
\]

Let \( d = \min_{\omega \in 2^{[N]} \setminus \{\emptyset\}} d_\omega \). Define \( U = \{X \in \mathbb{C}^{N \times N} : ||X||_{\max} < d\} \).

If \( \delta \in U \), then

\[
||(W - I)_\omega - (W - I + \delta)_\omega||_{\max} = ||\delta_\omega||_{\max} < d \leq d_\omega,
\]

so we get

\[
\text{Spec}(W_\omega - I_\omega + \delta_\omega) = \{u_{\omega}^T v_\omega - 1 + \Delta_1^\omega, \Delta_2^\omega - 1, \ldots, \Delta_{|\omega|}^\omega - 1\},
\]

where \( |\Delta_k^\omega| < \epsilon \) for all \( k \in \{1, 2, \ldots, |\omega|\} \), by continuity at \( (W - I)_\omega \) and Lemma 5.

Let \( \mu, \sigma \in \mathcal{P}_{[\cdot]+}(W + \delta) \), where \( \mu \) is a maximal codeword. Define \( \tau = \mu \cap \sigma \). We want to show that \( \tau \in \mathcal{P}_{[\cdot]+}(W + \delta) \). First, observe that

\[
u^T_\tau v_\tau - 1 + \text{Re}(\Delta^\tau_{\mu}) = u^T_\sigma v_\sigma - 1 + \text{Re}(\Delta^\sigma_{\mu}) + \text{Re}(\Delta^\tau_{\mu} - \Delta^\sigma_{\mu}) - u^T_{\sigma \backslash \tau} v_{\sigma \backslash \tau}.
\]
We know that $u^T v_\sigma - 1 + \text{Re}(\Delta^q_1) < 0$ because $\sigma \in \mathcal{P}_{[\cdot]}(W + \delta)$ by assumption. As for $\text{Re}(\Delta^q_1 - \Delta^q_1) - u^T v_{\sigma \setminus \tau}$, we know that $u_i v_i \neq 0$ for all $i \in \{1, 2, \ldots, N\}$. Therefore, since $k \notin \mu$ for every $k \in \sigma \setminus \tau$, it follows that $-u^T v_{\sigma \setminus \tau} < 0$ because $k \notin [N]_{<0}$ by Lemma 7. Furthermore,

$$\text{Re}(\Delta^q_1 - \Delta^q_1) < 2\epsilon < u^T v_{\tau},$$

where $p' = \arg\min_{p \in [N]_{>0}} u_p v_p$ so that $u_{p'} v_{p'} \leq u_p v_p$ for all $p \in [N]_{>0}$. Thus,

$$\text{Re}(\Delta^q_1 - \Delta^q_1) - u^T v_{\sigma \setminus \tau} < u_{p'} v_{p'} - u^T v_{\sigma \setminus \tau} \leq 0,$$

so $u^T v_{\tau} - 1 + \text{Re}(\Delta^q_1) < 0$, so we conclude that $\tau \in \mathcal{P}_{[\cdot]}(W + \delta)$. Note that if $\sigma$ is a maximal codeword, then we get that $\mu \cap \sigma$ is a codeword, which implies that $\mathcal{P}_{[\cdot]}(W + \delta)$ is maximal intersection-complete. 

We remark that it is possible to state a version of Proposition 3 in which less can be assumed about $\delta$; we refer the reader to the Appendix. We also remark that our assumption in Proposition 3 that $u_i v_i \neq 0$ for all $i \in \{1, 2, \ldots, N\}$ amounts to ruling out neurons that either have no synapses onto any neurons in the network or have no synapses from any other neurons in the network onto them.

### 4.2 Permitted Sets Beyond the Threshold-Linear Regime

In section 4.1, we showed that the collection of permitted sets of $D \dot{x} + x = [W x + b]_+$, where $\text{rk}(W) = 1$, is a convex code. We then extended that result to $D \dot{x} + x = [(W + \delta)x + b]_+$, where $\text{rk}(W) = 1$ and $\delta$ is a small enough perturbation. Since the relationship between input and firing rates of neuron is not in reality threshold-linear, a better approximation would be to replace $[\cdot]_+$ by a well-behaved function $\Phi$. Therefore, next we consider neuronal networks whose dynamics are modeled by a rate model $D \dot{x} + x = \Phi(W x + b)$, where $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ is a $C^1$ activation function with finitely many discontinuities. We refer the reader to section 2.4.1 for quick overview about rate models. This section has three goals: (1) Defining a notion of permitted set for $D \dot{x} + x = \Phi(W x + b)$. The collection of permitted sets in this regime will be denoted by $\mathcal{P}_{\Phi}(W)$; (2) showing that $\mathcal{P}_{\Phi}(W)$ is a convex code when $\text{rk}(W) = 1$; and (3) showing that $\mathcal{P}_{\Phi}(W + \delta)$ is a convex code when $\text{rk}(W) = 1$ and $\delta$ is a suitable perturbation.
Let us address the first item, i.e., defining permitted sets for $D\dot{x} + x = \Phi(Wx + b)$.

The next lemma says that any point in $\Phi(\mathbb{R})^N = \Phi(\mathbb{R}) \times \cdots \times \Phi(\mathbb{R})$ can be made into a fixed point of $D\dot{x} + x = \Phi(Wx + b)$ by applying a suitable external stimulus $b$ to the network.

**Lemma 8.** Let $W \in \mathbb{R}^{N \times N}$ be a synaptic weight matrix; $D \in \mathbb{R}^{N \times N}$ a diagonal matrix of time constants; and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be an activation function. Let $x^* = (\Phi(s_1), \ldots, \Phi(s_N)) \in \Phi(\mathbb{R})^N$. Then there is $b \in \mathbb{R}^N$ such that $x^*$ is a fixed point of $D\dot{x} + x = \Phi(Wx + b)$.

**Proof.** Let $b = f^* - Wx^*$, where $f^* \in \Phi^{-1}(x^*_1) \times \Phi^{-1}(x^*_2) \times \cdots \times \Phi^{-1}(x^*_N)$. Then

$$\Phi(Wx^* + b) = \Phi(Wx^* + f^* - Wx^*) = (\Phi(f^*_1), \Phi(f^*_2), \ldots, \Phi(f^*_N)) = (x^*_1, x^*_2, \ldots, x^*_N) = x^*.$$ 

\[\square\]

Notice that $b = s - Wx^*$ is one possible choice of $b$, where $s = (s_1, \ldots, s_N)$; therefore, for this choice of $b$, the overall input into the $i$th neuron is $s_i$. In deciding whether $x^* \in \Phi(\mathbb{R})^N$ is a stable fixed point of $D\dot{x} + x = \Phi(Wx + b)$ given some $b \in \mathbb{R}^N$, we will be determining whether the Jacobian $JF$ of $F(x) = \Phi(Wx + b) - x$ is stable at $x^*$. (Note that $D^{-1}$ will not change the sign of the real part of the eigenvalues in the spectrum of $JF(x)$, so we do not have to deal with the Jacobian of $D^{-1}(\Phi(Wx + b) - x)$.)

**Lemma 9.** Let $W \in \mathbb{R}^{N \times N}$ be a synaptic weight matrix, $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix of time constants, and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be an activation function that is $C^1$ except possibly at finitely many points. Let $F(x) = \Phi(Wx + b) - x$, and define

$$\Lambda(x) = \begin{pmatrix}
\Phi'(W^{(1)} \cdot x + b_1) & 0 & \cdots & 0 \\
0 & \Phi'(W^{(2)} \cdot x + b_2) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \Phi'(W^{(N)} \cdot x + b_N)
\end{pmatrix}.$$
Then the Jacobian of $F$ is $JF(x) = \Lambda(x)W - I$.

**Proof.** We have

$$JF(x)_{ij} = \begin{cases} W_{ii} \Phi'(W^{(i)} \cdot x + b_i) - 1 & \text{if } i = j, \\ W_{ij} \Phi'(W^{(i)} \cdot x + b_i) & \text{if } i \neq j. \end{cases}$$

In other words, $JF(x) = \Lambda(x)W - I$. 

Let $\Lambda(x)_{kk}$ denote the $k$th diagonal entry of $\Lambda(x)$, i.e., $\Lambda(x)_{kk} = \Phi'(W^{(k)} \cdot x + b_k)$ for any $k \in [N]$. Observe that since $b = f^* - Wx^*$, where $f^* \in \Phi^{-1}(x^*)$,

$$\Lambda(x^*)_{kk} = \Phi'(W^{(k)} \cdot x^* + b_k) = \Phi'(W^{(k)} \cdot x^* + f^*_k - W^{(k)} \cdot x^*) = \Phi'(f^*_k)$$

for any $k \in [N]$. Therefore, assessing whether $x^*$ is asymptotically stable will require computing the spectrum of $\Lambda(x^*)W - I$ for potentially several $f^* \in \Phi^{-1}(x^*)$. Specifically, if there is no $f^* \in \Phi^{-1}(x^*)$ such that $\Lambda(x^*)W - I$ is stable, then $x^*$ will not be asymptotically stable. Activation functions such that $|\Phi^{-1}(x^*)| > 1$ do have biological significance. For example, the F-I curve of neurons generally exhibit depolarization block, so the firing rate of neurons increases up to a certain point and from there it declines (and $\Phi(J) \to 0$ as $J \to \infty$), in which case $|\Phi^{-1}(x^*)| \geq 2$. See figure 4.1 for a schematic outlining situation we are describing.

![Figure 4.1](image.png)

**Figure 4.1:** Here is a situation that might arise when assessing the stability of $\Lambda(x^*)W - I$: Note that even though $\Phi(f^*_1) = \Phi(f^*_2) = x^*_1$, we have that $\Phi'(f^*_2) < 0 < \Phi'(f^*_1)$. Here $\Phi^{-1}(x^*_1) = \{f^*_1, f^*_2\}$, so $|\Phi^{-1}(x^*_1)| = 2$. 
To our knowledge, there is no definition of permitted set for firing rate models outside of threshold-linear networks. Our goal for the remainder of this section is to address this gap in the literature by proposing a definition for active/inactive neurons in a network, and then proving that \( P(W) \) satisfies the nesting property when \( \text{rk}(W) = 1 \). The following result that suggests a flexible notion of when one might say that a neuron is unresponsive to a stimulus, an essential concept for sensibly introducing permitted sets for firing rate models. First recall (Section 3.1) that for a threshold-linear network

\[
D \dot{x} + x = [Wx + b]_+ ,
\]

a subset \( \sigma \subseteq [N] \) is a permitted set if and only if there is \( b \in \mathbb{R}^N \) such that there is an asymptotically stable fixed point \( x^* \in \mathbb{R}^N \) satisfying \( x^*_i > 0 \) for every \( i \in \sigma \) and \( x^*_j = 0 \) otherwise. Suppose, then, we naively decided to call \( \sigma \) a permitted set of \( D \dot{x} + x = \Phi(Wx + b) \) if and only there is \( b \in \mathbb{R}^N \) such that there is an asymptotically stable fixed point \( x^* = (\Phi(I_1), \Phi(I_2), \ldots, \Phi(I_N)) \in \Phi(\mathbb{R})^N \) satisfying \( \Phi(I_j) > r_1 \) for \( i \in \sigma \) and \( \Phi(I_j) < r_2 \) otherwise, for some \( 0 < r_1 < r_2 \). Then Proposition 4 says that if the threshold for deciding whether a neuron is active is chosen too low, then any subset of neurons would be permitted:

**Proposition 4.** Let \( W \in \mathbb{R}^{N \times N} \) be a synaptic weight matrix, \( D \in \mathbb{R}^{N \times N} \) a diagonal matrix of time constants, and \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) be an activation function that is \( C^1 \) except possibly at finitely many points such that \( \Phi'(s) \to 0 \) as \( s \to -\infty \). Let \( r_1, r_2 \in \mathbb{R} \) be such that

\[
0 < r_1 < r_2 < 1/|||W|||.
\]

Suppose that there is \( J \in \mathbb{R} \) such that \( |\Phi'(J)| > r_2 \). For any \( \sigma \subseteq [N] = \{1, 2, \ldots, N\} \), there is \( b \in \mathbb{R}^N \) so that \( D \dot{x} + x = \Phi(Wx + b) \) has an asymptotically stable fixed point \( x^* = (\Phi(I_1), \Phi(I_2), \ldots, \Phi(I_N)) \in \Phi(\mathbb{R}) \) such that \( |\Phi'(I_i)| \geq r_2 \) for all \( i \in \sigma \) and \( |\Phi'(I_j)| \leq r_1 \) for all \( j \notin \sigma \).

**Proof.** Let \( J \in \mathbb{R} \) be such that \( |\Phi'(J)| > r_2 \). Continuity of \( \Phi' \) and the fact that \( \Phi'(s) \to 0 \) as \( s \to -\infty \) imply that there is an open set \( \mathcal{N}_1 \) such that for all \( x \in \mathcal{N}_1 \), we have \( r_2 < |\Phi'(x)| < 1/|||W||| \). Continuity of \( \Phi' \) and the fact that \( \Phi'(s) \to 0 \) as \( s \to -\infty \) also imply that there is an open set \( \mathcal{N}_2 \) such that for all \( x \in \mathcal{N}_2 \), we have \( |\Phi'(x)| < r_1 \). Let \( x^* \) be such that for all \( i \in \sigma \), we have \( x^*_i = \Phi(I_i) \), where \( I_i \in \mathcal{N}_1 \); and for all \( j \notin \sigma \), we have \( x^*_j = \Phi(I_j) \), where \( I_j \in \mathcal{N}_2 \). By Lemma 8 if \( b = f^* - Wx^* \), where \( f^*_i \in \Phi^{-1}(x^*_i) \cap \mathcal{N}_1 \subseteq \Phi^{-1}(x^*_i) \) for all \( i \in \sigma \) and \( f^*_j \in \Phi^{-1}(x^*_j) \cap \mathcal{N}_2 \subseteq \Phi^{-1}(x^*_j) \) for all \( j \notin \sigma \),
then $x^*$ is a fixed point. (The reason why we consider $\Phi^{-1}(x_i^*) \cap N_1$, $\Phi^{-1}(x_j^*) \cap N_2$ instead of just $\Phi^{-1}(x_i^*)$ is that we want $f_i^*$ such that $\Phi(f_i^*) = x_i^*$ as well as $|\Phi'(f_i^*)| < 1/||W||$.)

To prove that $x^*$ is asymptotically stable, we prove that $JF(x^*) = \Lambda(x^*)W - I$ has eigenvalues whose real part is negative. Since $\Lambda(x^*)$ is a diagonal matrix such that every entry is strictly bounded by $1/||W||$, it follows that $|||\Lambda(x^*)||| < 1/||W||$ by Lemma 4. Then we observe

$$|||\Lambda(x^*)W||| \leq |||\Lambda(x^*)||||W||| \quad \text{(by submultiplicativity of } ||| \cdot |||)$$

$$< \frac{1}{|||W||||} |||W|||$$

$$= 1.$$ 

Since the spectral radius of $\Lambda(x^*)W$ satisfies the inequality $\rho(\Lambda(x^*)W) \leq |||\Lambda(x^*)W|||$, we have that $\rho(\Lambda(x^*)W) < 1$. Letting $\lambda_1, \lambda_2, \ldots, \lambda_N \in \mathbb{C}$ denote the eigenvalues of $\Lambda(x^*)W$, we know by definition of spectral radius that

$$\rho(\Lambda(x^*)W) = \max(|\lambda_1|, |\lambda_2|, \ldots, |\lambda_N|).$$

In particular, $|\text{Re}(\lambda_k)| \leq |\lambda_k| < 1$ for all $k \in [N]$, so $\text{Re}(\lambda_k) - 1 < 0$ for all $k \in [N]$. Therefore, since the eigenvalues of $\Lambda(x^*)W - I$ are of the form $\lambda - 1$ by Lemma 1 where $\lambda$ is an eigenvalue of $\Lambda(x^*)W$, we conclude that $x^*$ is asymptotically stable.

As we said earlier in the introductory remarks for this section, a $C^1$ activation function $\Phi$ (with finitely many singularities) is a better approximation than the threshold nonlinearity $[\cdot]_+$. Proposition 4 illustrates the challenges to dichotomize the activity into active and inactive groups of neurons. We present in Definition 1 our approach to generalize what is considered an active or inactive neuron:

**Definition 1.** Let $W \in \mathbb{R}^{N \times N}$ be a synaptic weight matrix, $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix of time constants, and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be an activation function that is $C^1$ except at finitely many points, and $\Phi'(I) \to 0$ as $I \to -\infty$. Let $r_1, r_2 > 0$ be such that $r_1 < r_2$. Define

$$U = \{ I \in \mathbb{R} : |\Phi'(I)| < r_1 \} , \text{ and } A = \{ I \in \mathbb{R} : |\Phi'(I)| \geq r_2 \} .$$
Let \( x^* \in \Phi(\mathbb{R})^N \) be a vector of firing rates of a network with \( N \) units whose dynamics are described by \( D\dot{x} + x = \Phi(Wx + b) \). The firing rates for which \( x^*_i \in \Phi(\mathcal{A}) \) are the rates where we say that neuron \( i \) is unresponsive, and \( x^*_i \in \Phi(\mathcal{U}) \) are the rates where neuron \( i \) is active. We say that a subset \( \emptyset \subset \sigma \subseteq [N] \) of neurons is coactive if for all \( i \in \sigma \), we have \( x^*_i \in \Phi(\mathcal{A}) \), and \( x^*_j \in \Phi(\mathcal{U}) \) for \( j \notin \sigma \).

Next we define permitted and forbidden sets in the context of firing rate models:

**Definition 2.** Let \( W \in \mathbb{R}^{N \times N} \) be a synaptic weight matrix, and \( D \in \mathbb{R}^{N \times N} \) be a diagonal matrix of time constants. Let \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) be an activation function that is \( C^1 \) except at finitely many points, and \( \Phi'(I) \to 0 \) as \( I \to -\infty \). Let \( r_1, r_2 > 0 \) be such that \( r_1 < r_2 \). Let \( \mathcal{U} \) and \( \mathcal{A} \) be the set of inputs that make a neuron unresponsive and active, respectively. For \( \emptyset \subset \sigma \subseteq [N] \), then:

1. We call \( \sigma \) a permitted set of the network dynamics defined by \( D\dot{x} + x = \Phi(Wx + b) \) if there exists \( b \in \mathbb{R}^N \) so that there is an asymptotically stable fixed point \( x^* = (\Phi(I_1), \ldots, \Phi(I_N)) \) satisfying
   - for all \( i \in \sigma \), we have \( |\Phi'(I_i)| \geq r_2 \); and
   - for all \( j \notin \sigma \), we have \( |\Phi'(I_j)| < r_1 \).

2. We say that \( \sigma \) is a marginally forbidden set of the network dynamics defined by \( D\dot{x} + x = \Phi(Wx + b) \) if \( \sigma \) is not permitted and there exists \( b \in \mathbb{R}^N \) so that there is an asymptotically stable fixed point \( x^* = (\Phi(I_1), \ldots, \Phi(I_N)) \) satisfying
   - for all \( i \in \sigma \), we have \( r_1 \leq |\Phi'(I_i)| < r_2 \); and
   - for all \( j \notin \sigma \), we have \( |\Phi'(I_j)| < r_1 \).

3. If \( \sigma \) is neither permitted nor marginally forbidden, then we call \( \sigma \) a forbidden set.

We denote the collection of all permitted sets by \( \mathcal{P}_\Phi(W) \), unless context makes it clear what the neuronal network’s activation function is. We use the expression “\( x^* \) supports \( \sigma \)” when \( x^* = (\Phi(I_1), \ldots, \Phi(I_N)) \in \Phi(\mathbb{R})^N \) is a fixed point of the dynamics such that \( |\Phi'(I_i)| \geq r_2 \) for \( i \in \sigma \) and \( |\Phi'(I_j)| < r_1 \) for any \( j \notin \sigma \).
We remark that the constants $r_1$ and $r_2$ can be chosen arbitrarily; Proposition 4 makes a suggestion for what a reasonable choice of those constants should be. For instance, we recover the concept of permitted sets in the threshold-linear regime by selecting $0 < r_2 < 1$.

Before presenting examples, the next lemma gives a formula for the one eigenvalue we will usually be concerned with when dealing with rank-one $W \in \mathbb{R}^{N \times N}$:

**Lemma 10.** Let $W = uv^T$, where $u, v \in \mathbb{R}^N \setminus \{0\}$, be a synaptic weight matrix, $D \in \mathbb{R}^{N \times N}$ a diagonal matrix of time constants, and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be an activation function that is $C^1$ except possibly at finitely many points. Let $x^* \in \Phi(\mathbb{R})^N$ and $b \in \mathbb{R}^N$ such that $x^*$ is a fixed point of $D\dot{x} + x = \Phi(Wx + b)$. Then $x^*$ is an asymptotically stable fixed point if and only if

$$\text{tr}(\Lambda(x^*)W) = \sum_{i=1}^N u_i v_i \Phi'(f^*_i) < 1,$$

where $f^* \in \Phi^{-1}(x^*_1) \times \Phi^{-1}(x^*_2) \times \cdots \times \Phi^{-1}(x^*_N)$.

**Proof.** First note that $x^*$ is an asymptotically stable fixed point of $D\dot{x} + x = \Phi(Wx + b)$ if and only if $JF(x^*) = \Lambda(x^*)W - I$ is stable. The eigenvalues of $JF(x^*)$ are of the form $\lambda - 1$, by Lemma 1, where $\lambda \in \text{Spec}(\Lambda(x^*)W)$. Since $\Lambda(x^*)W$ is rank-one, it follows by Lemma 2 that $\text{Spec}(\Lambda(x^*)W) = \{0, \text{tr}(\Lambda(x^*)W)\}$, so $\text{Spec}(\Lambda(x^*)W - I) = \{-1, \text{tr}(\Lambda(x^*)W) - 1\}$. Asymptotic stability of $x^*$ implies $\text{tr}(\Lambda(x^*)W) - 1 < 0$.

Since $x^*$ is a fixed point of $D\dot{x} + x = \Phi(Wx + b)$, we have that $b = f^* - Wx^*$, where $f^* \in \Phi^{-1}(x^*_1) \times \Phi^{-1}(x^*_2) \times \cdots \times \Phi^{-1}(x^*_N)$. In particular, for all $i \in [N]$,

$$\Lambda(x^*)_{ii} = \Phi'(W^{(i)} \cdot x^* + b_i)$$

$$= \Phi'(W^{(i)} \cdot x^* + f^*_i - W^{(i)} \cdot x^*)$$

$$= \Phi'(f^*_i).$$

Therefore, we can rewrite $\text{tr}(\Lambda(x^*)W) - 1 < 0$ as

$$\text{tr}(\Lambda(x^*)W) = \sum_{i=1}^N u_i v_i \Phi'(f^*_i) < 1.$$
Now we present two examples. In the first example, we explicitly show that a certain subset is forbidden, whereas in the second example we show that a certain subset is permitted.

**Example 5.** Let $W \in \mathbb{R}^{2 \times 2}$ be such that $W_{ij} = 1/2$ for all $i, j \in \{1, 2\}$. In other words,

$$W = \begin{pmatrix}
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\
\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}
\end{pmatrix}.$$

Let $D \in \mathbb{R}^{2 \times 2}$ be a diagonal matrix of time constants and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be defined by $\Phi(x) = e^x$. Consider the rate model $D \dot{x} + x = \exp(Wx + b)$.

Let $r_2 = 1/\|W\|_2$ and $0 < r_1 < r_2$. We will show that $\{1, 2\} \notin \mathcal{P}_\Phi(W)$, i.e., $\{1, 2\}$ is not a permitted set. We know by Lemma 8 that any point $x^* = (\Phi(J_1), \Phi(J_2)) \in \Phi(\mathbb{R})^2$ can be made into a fixed point of $D \dot{x} + x = \Phi(Wx + b)$ for a suitable choice of $b \in \mathbb{R}^2$. Therefore, to show that $\{1, 2\}$ is not permitted, we will prove that there is no choice of $J_1, J_2 \in \mathbb{R}$ such that (1) $|\Phi'(J_i)| \geq r_2 = 1$ and (2) $x^* = (\Phi(J_1), \Phi(J_2))$ is asymptotically stable.

When $W$ is rank one, note that $\|W\|_2 = \|u\| \|v\| = 1$. Observe that $|\Phi'(J)| = e^J \geq r_2 = 1$ if and only if $J \geq 0$. For any $J_1, J_2 \geq 0$, we have $x_i^* = \Phi(J_i) = e^{J_i} \geq 1$, where $i = 1, 2$. For such a choice of $x^* = (x^*_1, x^*_2)$, by Lemma [8] $b = \log(x^*) - Wx^*$ makes $x^*$ a fixed point of $D \dot{x} + x = \exp(Wx + b)$. Next we want to check whether $x^*$ is stable. It turns out that $x^*$ is stable if

$$\sum_{k=1}^{n} u_kv_kx_k^* < 1$$

by Lemma 10. However,

$$\sum_{k=1}^{2} u_kv_kx_k^* = \sum_{k=1}^{2} \frac{1}{2} e^{J_k} \geq \sum_{k=1}^{2} \frac{1}{2} = 1,$$
so $x^*$ is not an asymptotically stable fixed point of $D\dot{x} + x = \Phi(Wx + b)$. Since we have shown that every $x^* = (\Phi(J_1), \Phi(J_2)) = (e^{J_1}, e^{J_2}) \in \Phi(\mathbb{R})^2$ such that $|\Phi'(J_i)| \geq r_2 = 1$ (which was equivalent to insisting that $J_1, J_2 \geq 0$) is not an asymptotically stable fixed point, we conclude that $\{1, 2\}$ is not a permitted set.

**Example 6.** Let $W = uv^T \in \mathbb{R}^{2\times 2}$ be such that $W = \begin{pmatrix} \frac{1}{\sqrt{6}} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{\sqrt{6}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} \end{pmatrix}$, and let $D \in \mathbb{R}^{2\times 2}$ be a diagonal matrix of time constants. Let $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be defined by $\Phi(x) = e^x$. We will again consider the rate model $D\dot{x} + x = \Phi(Wx + b)$.

Let $r_2 = 6/5$ and $0 < r_1 < r_2$. We will show that $\{1, 2\}$ is a permitted set of $D\dot{x} + x = \Phi(Wx + b)$. Observe that $\|W\|_2 = \|u\| \|v\| = 5/6$. For every $J > \log(6/5) \approx 0.1823$,

$$|\Phi'(J)| = e^J > \frac{1}{\|W\|_2} = \frac{6}{5} = r_2.$$

Let $x^*_i = \Phi(J_i) = e^{J_i}$, where $\log(\sqrt{6}/2) > J_i > \log(6/5)$ and $i = 1, 2$. By Lemma 8, setting $b = \log(x^*) - Wx^*$, we have that $x^*$ is a fixed point of $D\dot{x} + x = \exp(Wx + b)$. (Note that $\log(\sqrt{6}/2) \approx 0.2027$.) Then

$$\sum_{k=1}^{2} u_kv_kx_k^* = \sum_{k=1}^{2} \frac{1}{\sqrt{6}}e^{J_k}$$

$$< \frac{1}{\sqrt{6}} \sum_{k=1}^{2} \sqrt{6} = 1.$$

Hence, $x^*$ is an asymptotically stable fixed point. Since we have found $J_1, J_2 \in \mathbb{R}$ such that $|\Phi'(J_i)| \geq r_2$ and $x^* = (\Phi(J_1), \Phi(J_2))$ is an asymptotically stable fixed point, we have found an asymptotically stable fixed point where both neurons are considered active. In other words, we have shown that $\{1, 2\}$ is a permitted set.

We remark that although we chose $r_2 = 1/\|W\|_2$, we were free to choose $r_2$ larger than $1/\|W\|_2$. 


4.3 $C^1$ Activation Function Case

Next we set to address the second goal of this chapter, namely proving that $\mathcal{P}_\Phi(W)$ is a convex code when $\text{rk}(W) = 1$. In the proof of Proposition 5, the point is to use the asymptotically stable fixed points associated to the codewords $\sigma$ and $\mu$, say $x^*$ and $y^*$, and construct a fixed point $z^*$ associated to $\sigma \cap \mu$. From there, we will be able to find a suitable $b \in \mathbb{R}^N$ in order to prove that $z^*$ is in fact asymptotically stable.

**Proposition 5.** Let $W = uv^T$, where $u, v \in \mathbb{R}^N \setminus \{0\}$, be a synaptic weight matrix; $D \in \mathbb{R}^{N \times N}$ a diagonal matrix of time constants; and $\Phi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ be a $C^1$ activation function (with finitely many discontinuities) such that $\Phi'(J) \rightarrow 0$ as $J \rightarrow -\infty$. Fix activation thresholds $0 < r_1 < r_2$. Then $\sigma \cap \mu \in \mathcal{P}_\Phi(W)$, where $\sigma, \mu \in \mathcal{P}_\Phi(W)$ and $\mu$ is maximal. In particular, $\mathcal{P}_\Phi(W)$ is maximal intersection-complete, so $\mathcal{P}_\Phi(W)$ is convex by Theorem 3.

**Proof.** Since $\mu$ is a maximal codeword, it follows that for any $k \notin \mu$, then $\tilde{\mu} = \mu \cup \{k\}$ is not permitted—in other words, if $x^* = (\Phi(I_1), \ldots, \Phi(I_N))$ is a fixed point of $D\dot{x} + x = \Phi(Wx + b)$ such that $|\Phi'(I_i)| \geq r_2$ for all $i \in \tilde{\mu}$ and $|\Phi'(I_j)| < r_1$ for all $j \notin \tilde{\mu}$, then $\text{tr}(\Lambda(x^*)W) \geq 1$.

Let $x^* = (\Phi(I_1), \ldots, \Phi(I_N))$ denote an asymptotically stable fixed point associated with $\mu$, which implies that $\text{tr}(\Lambda_{\mu}W) < 1$, where $\Lambda_{\mu} = \Lambda(x^*)$. Let

$$y^* = (\Phi(L_1), \ldots, \Phi(L_N))$$

be an asymptotically stable fixed point associated with $\sigma$, where $\sigma \in \mathcal{P}_\Phi(W)$, which implies that $\text{tr}(\Lambda_{\sigma}W) < 1$, where $\Lambda_{\sigma} = \Lambda(y^*)$.

Set $\tilde{\mu} = \mu \cup \{k\}$, where $k \notin \mu$. Further define $z^* = (\Phi(J_1), \ldots, \Phi(J_N))$ such that $z_i^* = x_i^*$ for $i \in [N] \setminus \{k\}$, and $z_k^* = \Phi(J_k)$ with $J_k \in \mathbb{R}$ satisfying $|\Phi'(J_k)| \geq r_2$. Note that $z^*$ represents the firing rates for $\tilde{\mu}$. By the hypothesis that $\mu$ is maximal, $\text{tr}(\Lambda_{\tilde{\mu}}W) \geq 1$, where $\Lambda_{\tilde{\mu}} = \Lambda(z^*)$. By putting together the inequalities

$$0 > -1 + \text{tr}(\Lambda_{\mu}W) = -1 + u_kv_k\Phi'(I_k) + \sum_{i \in [N] \setminus \{k\}} u_iv_i\Phi'(I_i), \quad \text{and} \quad -1 + \text{tr}(\Lambda_{\tilde{\mu}}W) = -1 + u_kv_k\Phi'(J_k) + \sum_{i \in [N] \setminus \{k\}} u_iv_i\Phi'(I_i) \geq 0,$$
we see that \( u_k v_k (\Phi'(J_k) - \Phi'(I_k)) > 0 \).

Set \( \tau = \mu \cap \sigma \) and assume that \( \tau \subset \sigma \). Define \( \widetilde{x}^* = (\Phi(\widetilde{I}_1), \ldots, \Phi(\widetilde{I}_N)) \) to be so that \( \widetilde{x}^*_i = y_i \) for any \( i \in \tau \) or \([N] \setminus \sigma\), and \( \widetilde{x}^*_j = \Phi(I_j) \) for \( j \in \sigma \setminus \tau \). In other words, \( \widetilde{x}^* \) is a fixed point associated with \( \tau \). Next we want to show that \( \widetilde{x}^* \) is asymptotically stable. In other words, letting \( \Lambda_\tau = \Lambda(z^*) \), we want to show that \( \text{tr}(\Lambda_\tau W) < 1 \). First, we simply rearrange terms:

\[
-1 + \text{tr}(\Lambda_\tau W) = -1 + \sum_{i \in \tau} u_i v_i \Phi'(\widetilde{I}_i) + \sum_{j \notin \tau} u_j v_j \Phi'(\widetilde{I}_j)
\]

\[
= -1 + \sum_{i \in \tau} u_i v_i \Phi'(\widetilde{I}_i) + \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(\widetilde{I}_j) + \sum_{j \in \sigma \setminus \tau} u_j v_j \Phi'(\widetilde{I}_j)
\]

If add and subtract \( \sum_{l \in \sigma \setminus \tau} u_l v_l \Phi'(L_l) \) to the last expression above,

\[
-1 + \text{tr}(\Lambda_\tau W) = \sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(\widetilde{I}_l) - \Phi'(L_l)) + \left( -1 + \sum_{i \in \tau} u_i v_i \Phi'(\widetilde{I}_i) + \sum_{l \in \sigma \setminus \tau} u_l v_l \Phi'(L_l) + \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(\widetilde{I}_j) \right)
\]

Observe that the last expression implies that \( -1 + \text{tr}(\Lambda_\tau W) < 0 \) because

\[
\sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(\widetilde{I}_l) - \Phi'(L_l)) < 0
\]

(as \( l \in \sigma \setminus \tau \) implies \( l \notin \mu \)) and

\[
-1 + \sum_{i \in \tau} u_i v_i \Phi'(\widetilde{I}_i) + \sum_{l \in \sigma \setminus \tau} u_l v_l \Phi'(L_l) + \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(\widetilde{I}_j) = \text{tr}(\Lambda_\sigma W) - 1 < 0
\]

(because \( \widetilde{I}_i = L_i \) for any \( i \in \tau \) or \( i \in [N] \setminus \sigma \)) by the assumption that \( \sigma \) is permitted and \( y^* \) is an asymptotically stable fixed point associated with the permitted set \( \sigma \).

We just showed that \( \widetilde{x}^* = (\Phi(\widetilde{I}_1), \Phi(\widetilde{I}_2), \ldots, \Phi(\widetilde{I}_N)) \in \Phi(\mathbb{R})^N \) is an asymptotically stable fixed point of \( D\dot{x} + x = \Phi(Wx + b) \), where \( b = s - Wx^* \) satisfies \( s_i = L_i \) for \( i \in \tau \) or \([N] \setminus \sigma\) and \( s_j = I_j \) for \( j \in \sigma \setminus \tau \). By our choice of \( b \), we have that \( |\Phi(s_i)| \geq r_2 \) for
\(i \in \tau\) and \(|\Phi'(s_j)| < r_1\) otherwise. Therefore, \(\tau\) is a permitted set.

For the remainder of this section, we tackle the last goal of this chapter, i.e., proving that \(P_{\Phi}(W + \delta)\) is a convex code, where \(\text{rk}(W) = 1\) and \(\delta\) is a suitable perturbation. The outline for establishing that \(P_{\Phi}(W + \delta)\) is convex is analogous to the corresponding result for \(P_{[\cdot]}(W + \delta)\): First we show that, in a suitable sense, neurons whose self-coupling efficacy is inhibitory will always participate in maximal codewords of \(P_{\Phi}(W + \delta)\). Second, we use that observation to establish in Theorem 4 that \(P_{\Phi}(W + \delta)\) is convex.

Lemma 11. Let \(W = uv^T\), where \(u, v \in \mathbb{R}^N \setminus \{0\}\); \(D \in \mathbb{R}^{N \times N}\) a diagonal matrix of time constants; and \(\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}\) a \(C^1\) activation function with finitely many singularities and such that \(\Phi'(J) \to 0\) as \(J \to -\infty\). Fix activation thresholds \(0 < r_1 < r_2\). Define \([N]_{< 0} = \{i \in [N] : \text{there is } \tilde{I}_i \in \mathbb{R} \text{ such that } sgn(u_i v_i \Phi'(\tilde{I}_i)) = -1 \text{ and } |\Phi'(\tilde{I}_i)| \geq r_2\}\).

Suppose that \([N]_{< 0} \neq \emptyset\). Let \(\delta \in \mathbb{R}^{N \times N}\). Then there is \(d = d(\delta) > 0\) depending on \(\delta\) such that if \(||\delta||_{\text{max}} < d\), then \([N]_{< 0} \subseteq \mu\) for any \(\mu \in P_{\Phi}(W + \delta)\) that is a maximal codeword.

Proof. Let \(\delta \in \mathbb{R}^{N \times N}\) be such that \(P_{\Phi}(W + \delta) \neq \emptyset\). For any codeword \(\omega \in P_{\Phi}(W + \delta)\), we know that there is an asymptotically stable fixed point, say

\[
x^*(\omega) = (\Phi(I_1^\omega), \Phi(I_2^\omega), \ldots, \Phi(I_N^\omega)),
\]

that supports \(\omega\); we can therefore take for granted (and fix) a collection of asymptotically stable fixed points \(x^*(\omega)\) where each \(x^*(\omega)\) supports exactly one codeword, namely \(\omega\). Define

\[
M = \max_{\omega \in P_{\Phi}(W + \delta)} ||\Lambda(x^*(\omega))||_{\text{max}} > 0,
\]

where \(x^*(\omega) = (\Phi(I_1^\omega), \Phi(I_2^\omega), \ldots, \Phi(I_N^\omega))\), where \(I_k^\omega \in \mathbb{R}\), is any asymptotically stable fixed point supporting \(\omega\). Define

\[
0 < \epsilon < \min \left(1, \min_{\omega \in P_{\Phi}(W + \delta)} \min_{s \in \omega \cap [N]_{< 0}} -\frac{u_s v_s}{2} \left(\Phi'(\tilde{I}_s) - \Phi'(I_s^\omega)\right)\right),
\]
where $\omega^c$ denotes the complement of $\omega$, $\tilde{I}_s \in \mathbb{R}$ is such that

$$r_2 \leq |\Phi'(\tilde{I}_s)| \leq \max_{\omega \in \mathcal{P}} \max_{i \in \omega} |\Phi'(I_\omega^i)|$$

and $\text{sgn}\left(u_s v_s \Phi'(\tilde{I}_s)\right) = -1$ for some $s \in [N]$. Note that $\omega^c \neq \emptyset$ for at least one maximal $\omega$ such that $[N]_{<0} \not\subseteq \omega$. Further note that for any $s \in \omega^c$, we have $|\Phi'(I_\omega^s)| < r_1$. It turns out that $u_s v_s (\Phi'(\tilde{I}_s) - \Phi'(I_\omega^s)) < 0$ for any $s \in \omega^c \cap [N]_{<0}$.

By way of contradiction, suppose that $[N]_{<0} \not\subseteq \mu$ for at least one codeword in $\mathcal{P}_\Phi^\text{max}(W + \delta)$. Let $\mu \in \mathcal{P}_\Phi(W + \delta)$ be a maximal codeword and

$$x^*(\mu) = (\Phi(I_1^\mu), \Phi(I_2^\mu), \ldots, \Phi(I_N^\mu))$$

be an asymptotically stable fixed point supporting $\mu$. Suppose $s \in [N]_{<0}$ is such that $s \notin \mu$. Set $\tilde{\mu} = \mu \cup \{s\}$; we will demonstrate that $\tilde{\mu} \in \mathcal{P}_\Phi(W + \delta)$ (which will contradict $\mu$ being a maximal codeword).

First we construct a fixed point supporting $\tilde{\mu}$ and determine what $d > 0$ we want for bounding $||\delta||_{\max}$. Let

$$x^*(\tilde{\mu}) = (\Phi(I_1^{\tilde{\mu}}), \Phi(I_2^{\tilde{\mu}}), \ldots, \Phi(I_N^{\tilde{\mu}})) \in \Phi(\mathbb{R})^N$$

be such that $I_i^{\tilde{\mu}} = I_i^\mu$ whenever $i \neq s$, and $I_s^{\tilde{\mu}} = \tilde{I}_s$ (where

$$r_2 \leq |\Phi'(\tilde{I}_s)| \leq \max_{i \in \mu} |\Phi'(I_i^\mu)|.$$}

Now we introduce $d > 0$ as follows: For any $\omega \in \mathcal{P}_\Phi(W + \delta)$, define $d_\omega > 0$ to be such that for any $X \in \mathbb{C}^{N \times N}$ with $||\Lambda(x^*(\omega))W - I - X||_{\max} < d_\omega$, we have

$$\min_{\pi \in \mathcal{S}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*(\omega))W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$}

Further, for each possible $\mu' \cup \{s\}$, where $s \notin \mu'$ and $\mu'$ is a maximal codeword, let $d_{\mu' \cup \{s\}} > 0$ be such that for any $X \in \mathbb{C}^{N \times N}$ with $||\Lambda(x^*(\tilde{\mu}))W - I - X||_{\max} < d_{\mu' \cup \{s\}}$,

$$\min_{\pi \in \mathcal{S}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*(\tilde{\mu}))W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$
Let $\tilde{d} = \min(\min_{\mu' \cup \{s\}} d_{\mu' \cup \{s\}}, \min_{\omega \in \mathcal{P}_b(W+\delta)} d_{\omega})$ and assume that $||\delta||_{\max} < d$, where $d = \frac{\tilde{d}}{M}$.

Next we will show that the spectra of $\Lambda(x^*(\mu))(W+\delta) - I$ and $\Lambda(x^*(\tilde{\mu}))(W+\delta) - I$ are close to the spectra of $\Lambda(x^*(\mu))W - I$ and $\Lambda(x^*(\tilde{\mu}))W - I$, respectively. Observe that

$$||\Lambda(x^*(\mu))W - I - (\Lambda(x^*(\mu))(W+\delta) - I)||_{\max} = ||\Lambda(x^*(\mu))\delta||_{\max} \leq ||\Lambda(x^*(\mu))||_{\max} ||\delta||_{\max} \leq M ||\delta||_{\max} < \tilde{d},$$

so that

$$\text{Spec}(\Lambda(x^*(\mu))(W+\delta) - I) = \{\text{tr}(\Lambda(x^*(\mu))W) - 1 + \Delta^\mu_1, \Delta^\mu_2 - 1, \ldots, \Delta^\mu_N - 1\},$$

where $|\Delta_i| < \epsilon$ for every $i \in [N]$. As for the spectrum of $\Lambda(x^*(\tilde{\mu}))(W + \delta) - I$,

$$||\Lambda(x^*(\tilde{\mu}))W - I - (\Lambda(x^*(\tilde{\mu}))(W + \delta) - I)||_{\max} = ||\Lambda(x^*(\tilde{\mu}))\delta||_{\max} \leq ||\Lambda(x^*(\tilde{\mu}))||_{\max} ||\delta||_{\max} \leq ||\Lambda(x^*(\mu))||_{\max} ||\delta||_{\max} \leq M ||\delta||_{\max} \leq \tilde{d},$$

so we have

$$\text{Spec}(\Lambda(x^*(\tilde{\mu}))(W + \delta) - I) = \{\text{tr}(\Lambda(x^*(\tilde{\mu}))W) - 1 + \Delta^\tilde{\mu}_1, \Delta^\tilde{\mu}_2 - 1, \ldots, \Delta^\tilde{\mu}_N - 1\},$$

where $|\Delta^\tilde{\mu}_i| < \epsilon$ for every $i \in [N]$.

Finally, we prove that $\Lambda(x^*(\tilde{\mu}))(W + \delta) - I$ is stable, which will imply that $\tilde{\mu}$ is a permitted set (and, hence, makes us reach the desired contradiction to maximality of $\mu$). It suffices to check $\text{tr}(\Lambda(x^*(\tilde{\mu}))W) - 1 + \text{Re}(\Delta^\tilde{\mu}_1) < 0$ because $\text{Re}(\Delta^\tilde{\mu}_1) - 1 < 0$ for
every $i \in [N]$ for $|\Delta^\mu_i| < \epsilon < 1$ by construction. First,

\[
\begin{align*}
\text{tr}(\Lambda(x^*(\bar{\mu}))W) - 1 + \text{Re}(\Delta^\mu_1) &= -1 + \sum_{i \in \bar{\mu}} u_iv_i\Phi'(I^\mu_i) + \sum_{j \notin \mu} u_jv_j\Phi'(I^\mu_j) + \text{Re}(\Delta^\mu_1) \\
&= -1 + \sum_{i \in \mu} u_iv_i\Phi'(I^\mu_i) + u_sv_s\Phi'(\bar{I}_s) \\
&\quad + \sum_{j \notin \mu} u_jv_j\Phi'(I^\mu_j) - u_sv_s\Phi'(I^\mu_s) + \text{Re}(\Delta^\mu_1) + \text{Re}(\eta),
\end{align*}
\]

where $\eta = \Delta^\mu_1 - \Delta^\mu_1$ and we used the fact that

\[
\sum_{i \in \bar{\mu}} u_i v_i \Phi'(I^\mu_i) = u_s v_s \Phi'(\bar{I}_s) + \sum_{i \in \mu} u_i v_i \Phi'(I^\mu_i)
\]

(because $I^\mu_i = I^\mu_i$ for $i \neq s$, and $I^\mu_s = \bar{I}_s$) and

\[
\sum_{j \notin \mu} u_j v_j \Phi'(I^\mu_j) = -u_s v_s \Phi'(I^\mu_s) + \sum_{j \notin \mu} u_j v_j \Phi'(I^\mu_j)
\]

(because $I^\mu_j = I^\mu_j$ for every $j \notin \mu$ and $j \neq s$ for all $j \notin \bar{\mu} = \mu \cup \{s\}$). By our setup,

\[
-1 + \sum_{i \in \mu} u_i v_i \Phi'(I^\mu_i) + \sum_{j \notin \mu} u_j v_j \Phi'(I^\mu_j) + \text{Re}(\Delta_1) = \text{tr}(\Lambda(x^*(\mu))W) - 1 + \text{Re}(\Delta^\mu_1) < 0.
\]

Furthermore,

\[
\text{Re}(\eta) + u_s v_s \Phi'(\bar{I}_s) - u_s v_s \Phi'(I^\mu_s) = \text{Re}(\eta) + u_s v_s \left( \Phi'(\bar{I}_s) - \Phi'(I^\mu_s) \right)
\]

\[
< 2\epsilon + u_s v_s \left( \Phi'(\bar{I}_s) - \Phi'(I^\mu_s) \right)
\]

\[
< 0.
\]

Hence, the eigenvalue $\text{tr}(\Lambda(x^*(\bar{\mu}))W) - 1 + \Delta^\mu_1$ has negative real part, so $\bar{\mu}$ is a permitted set, which is a contradiction. We conclude that $[N]_{<0} \subseteq \mu$. \qed

Theorem 4 is the main result of this chapter. It says that the permitted sets of $D\dot{x} + x = \Phi((W + \delta)x + b)$, where $\text{rk}(W) = 1$ and $\delta$ is a perturbation, form a convex code. Since $\Phi$ is assumed to be $C^1$ except at finitely many points, it generalizes Proposition
Once again, we remark that the results hold for arbitrary activation thresholds 0 < r_1 < r_2.

**Theorem 4.** Let \( W = uv^T \), where \( u, v \in \mathbb{R}^N \setminus \{0\} \) and \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) be a \( C^1 \) activation function with finitely many singularities such that \( \Phi'(J) \to 0 \) as \( J \to -\infty \). Suppose that \( u_i v_i \neq 0 \) for all \( i \in \{1, 2, \ldots, N\} \). Fix activation thresholds 0 < r_1 < r_2. Let \( \delta \in \mathbb{R}^{N \times N} \). Then there is \( d = d(\delta) > 0 \) such that if \( ||\delta||_{\text{max}} < d \), then for any \( \mu, \sigma \in \mathcal{P}_\Phi(W + \delta) \), where \( \mu \) is maximal, we have \( \mu \cap \sigma \in \mathcal{P}_\Phi(W + \delta) \). In particular, \( \mathcal{P}_\Phi(W + \delta) \) is maximal intersection-complete, so \( \mathcal{P}_\Phi(W + \delta) \) is convex by Theorem 3.

**Proof.** Suppose \( \delta \in \mathbb{R}^{N \times N} \) is such that \( \mathcal{P}_\Phi(W + \delta) \neq \emptyset \). For any codeword \( \omega \in \mathcal{P}_\Phi(W + \delta) \), we know that there is an asymptotically stable fixed point, say \( x^*(\omega) = (\Phi(I_{\omega}^1), \Phi(I_{\omega}^2), \ldots, \Phi(I_{\omega}^N)) \), that supports \( \omega \); we can therefore take for granted (and fix) a collection of asymptotically stable fixed points \( x^*(\omega) \) where each \( x^*(\omega) \) supports exactly one codeword, namely \( \omega \).

Define

\[
[N]_{<0} = \left\{ i \in [N] : \text{there is } \tilde{I}_i \in \mathbb{R} \text{ such that } \sgn(u_i v_i \Phi'(\tilde{I}_i)) = -1 \text{ and } |\Phi'(\tilde{I}_i)| \geq r_2 \right\},
\]

and

\[
[N]_{>0} = \left\{ i \in [N] : \text{there is } \tilde{I}_i \in \mathbb{R} \text{ such that } \sgn(u_i v_i \Phi'(\tilde{I}_i)) = 1 \text{ and } |\Phi'(\tilde{I}_i)| \geq r_2 \right\}.
\]

Choose

\[
0 < \epsilon < \min (1, M_1, M_2),
\]

where

\[
M_1 = \min_{\omega \in \mathcal{P}_\Phi(W + \delta)} \min_{s \in \omega \cap [N]_{<0}} -\frac{u_s v_s}{2} \left( \Phi'(I^s_s) - \Phi'(I_{\omega}^s) \right),
\]

and

\[
M_2 = \min_{\omega \in \mathcal{P}_\Phi(W + \delta)} \min_{p \in \omega \cap [N]_{>0}} \frac{u_p v_p}{2} \left( \Phi'(I^p_p) - \Phi'(I_{\omega}^p) \right),
\]
with $\tilde{I}_p, \tilde{I}_s \in \mathbb{R}$ such that

$$r_2 \leq |\Phi'(\tilde{I}_s)|, |\Phi'(\tilde{I}_p)| \leq \max_{\omega \in \mathcal{P}_\Phi(W + \delta)} \max_{i \in \omega} |\Phi'(I_{\omega}^i)|,$$

(Should $[N]_{<0} = \emptyset$ or $[N]_{>0} = \emptyset$ occur, define $\min(\emptyset) = +\infty$.)

Define

$$M = \max_{\omega \in \mathcal{P}_\Phi(W + \delta)} ||\Lambda(x^*(\omega))||_{\max} > 0,$$

Next we define the $d > 0$ we want for the bound $||\delta||_{\max} < d$. For every $\omega \in \mathcal{P}_\Phi(W + \delta)$, let $d_{\omega} > 0$ be such that whenever $X \in \mathbb{C}^{N \times N}$ satisfies $||\Lambda(x^*(\omega))(W - I) - X||_{\max} < d_{\omega},$

$$\min_{\pi \in \mathcal{G}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*(\omega))(W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$

Further, let $d_{\tilde{\mu} \cap \tilde{\sigma}} > 0$, where $\tilde{\mu}$ is any maximal codeword and $\tilde{\sigma}$ is a codeword, be such that whenever $X \in \mathbb{C}^{N \times N}$ satisfies $||\Lambda(x^*(\tilde{\mu} \cap \tilde{\sigma}))(W - I) - X||_{\max} < d_{\tilde{\mu} \cap \tilde{\sigma}},$

$$\min_{\pi \in \mathcal{G}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*(\tilde{\mu} \cap \tilde{\sigma}))(W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$

Finally, let $d_{\tilde{\mu} \cup \{s\}} > 0$, where $\tilde{\mu}$ is any maximal codeword and $s \notin \tilde{\mu}$, be such that whenever $X \in \mathbb{C}^{N \times N}$ satisfies $||\Lambda(x^*(\tilde{\mu} \cup \{s\}))(W - I) - X||_{\max} < d_{\tilde{\mu} \cup \{s\}},$

$$\min_{\pi \in \mathcal{G}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*(\tilde{\mu} \cup \{s\}))(W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$

(In other words, $d_{\tilde{\mu} \cup \{s\}}$ is defined in the same way as in Lemma $\square$.) Let

$$\tilde{d} = \min \left( \min_{\omega \in \mathcal{P}_\Phi(W + \delta)} d_{\omega}, \min_{\tilde{\mu} \cap \tilde{\sigma}} d_{\tilde{\mu} \cap \tilde{\sigma}}, \min_{\tilde{\mu} \cup \{s\}} d_{\tilde{\mu} \cup \{s\}} \right).$$

Define $d = \tilde{d}/M$ and assume $||\delta||_{\max} < d$.

The above finishes our setting up what properties $\delta$ satisfies. Now let $\mu, \sigma \in \mathcal{P}_\Phi(W + \delta)$, where $\mu$ is maximal. Let $x^*(\mu) = (\Phi(I_{1}^\mu), \ldots, \Phi(I_{N}^\mu))$ be an asymptotically stable fixed point supporting $\mu$, and $x^*(\sigma) = (\Phi(I_{1}^\sigma), \ldots, \Phi(I_{N}^\sigma))$ an asymptotically stable fixed point supporting $\sigma$. Let $\tau = \mu \cap \sigma$, and suppose that $\tau \subset \sigma$. Define $x^*(\tau) = (\Phi(I_{1}^\tau), \ldots, \Phi(I_{N}^\tau))$ to be such that $I_{i}^\tau = I_{i}^\sigma$ for every $i \in \tau$ or $i \in [N] \setminus \sigma$, and $I_{j}^\tau = I_{j}^\mu$.
for every $j \in \sigma \setminus \tau$. In other words, $x^*(\tau)$ is a fixed point supporting $\tau$. Our task next is to show that $x^*(\tau)$ is an asymptotically stable fixed point supporting $\tau$. Lastly, note that by Lemma 11, we have that $j \in [N]_{>0}$ for any $j \in \sigma \setminus \tau \neq \emptyset$ because $|\Phi'(I^\sigma_j)| \geq r_2$ and $j \notin \mu$ (and the latter implies that $j \notin [N]_{<0}$).

Since

$$||\Lambda(x^*(\sigma))W - I - (\Lambda(x^*(\sigma))(W + \delta) - I)||_{\text{max}} = ||\Lambda(x^*(\sigma))\delta||_{\text{max}} \leq ||\Lambda(x^*(\sigma))||_{\text{max}} ||\delta||_{\text{max}} \leq M||\delta||_{\text{max}} < \tilde{d},$$

so that

$$\text{Spec}(\Lambda(x^*(\sigma))(W + \delta) - I) = \{\text{tr}(\Lambda(x^*(\sigma))W) - 1 + \Delta^\sigma_1, \Delta^\sigma_2, \ldots, \Delta^\sigma_N - 1\},$$

where $|\Delta^\sigma_k| < \epsilon$ for every $k \in [N]$. Similarly,

$$||\Lambda(x^*(\tau))W - I - (\Lambda(x^*(\tau))(W + \delta) - I)||_{\text{max}} = ||\Lambda(x^*(\tau))\delta||_{\text{max}} \leq ||\Lambda(x^*(\tau))||_{\text{max}} ||\delta||_{\text{max}} \leq M||\delta||_{\text{max}} < \tilde{d},$$

implies

$$\text{Spec}(\Lambda(x^*(\tau))(W + \delta) - I) = \{\text{tr}(\Lambda(x^*(\tau))W) - 1 + \Delta^\tau_1, \Delta^\tau_2, \ldots, \Delta^\tau_N - 1\},$$

where $|\Delta^\tau_k| < \epsilon$ for every $k \in [N]$.

Now define $\eta = \Delta^\tau_1 - \Delta^\sigma_1$. We next show that $\text{tr}(\Lambda(x^*(\tau))W) - 1 + \Delta^\tau_1$ has negative
real part:

\[-1 + \text{tr}(\Lambda(x^*(\tau))W) + \text{Re}(\Delta_1^\tau) = -1 + \sum_{i \in \tau} u_i v_i \Phi'(I_i^\tau) + \sum_{j \notin \tau} u_j v_j \Phi'(I_j^\tau) + \text{Re}(\Delta_1^\tau)\]

\[= -1 + \sum_{i \in \tau} u_i v_i \Phi'(I_i^\tau) + \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(I_j^\tau)\]

\[+ \sum_{j \in \sigma \setminus \tau} u_j v_j \Phi'(I_j^\tau) + \text{Re}(\eta) + \text{Re}(\Delta_1^\sigma)\]

\[= \sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(I_l^\tau) - \Phi'(I_l^\sigma)) + \text{Re}(\eta)\]

\[+ \left( -1 + \sum_{i \in \tau} u_i v_i \Phi'(I_i^\tau) + \sum_{l \in \sigma \setminus \tau} u_l v_l \Phi'(I_l^\sigma) \right)\]

\[+ \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(I_j^\sigma) + \text{Re}(\Delta_1) \right)\]

Observe that the last expression implies that 

\[-1 + \text{tr}(\Lambda(x^*(\tau))W) + \text{Re}(\Delta_1^\tau) < 0:\]

First,

\[\sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(I_l^\tau) - \Phi'(I_l^\sigma)) + \text{Re}(\eta) < \sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(I_l^\tau) - \Phi'(I_l^\sigma)) + 2\epsilon\]

\[\leq \sum_{l \in \sigma \setminus \tau} u_l v_l (\Phi'(I_l^\tau) - \Phi'(I_l^\sigma))\]

\[+ u_{p'} v_{p'} (\Phi'(I_{p'}^\tau) - \Phi'(I_{p'}^\sigma))\]

\[< 0,\]

where \(p'\) is the index minimizing

\[
\min_{\omega \in \mathcal{P}_\Phi(W + \delta)} \min_{p \in \omega \cap [N]_{>0}} \frac{u_p v_p}{2} \left( \Phi'(I_p^\tau) - \Phi'(I_p^\sigma) \right),
\]

because \(u_l v_l (\Phi'(I_l^\tau) - \Phi'(I_l^\sigma)) < 0\) (as \(l \in [N]_{>0}\) and \(l \notin \mu\) so that \(|\Phi'(I_l^\tau)| < r_1\)) and
\[ u_l v_l \left( \Phi'(I_l^\tau) - \Phi'(I_l^\sigma) \right) + u_{l'} v_{l'} \left( \Phi'(I_{l'}^\nu) - \Phi'(I_{l'}^\mu) \right) < 0 \quad \text{for every } l \in \sigma \setminus \tau. \] Second,

\[
-1 + \sum_{i \in \tau} u_i v_i \Phi'(I_i^\tau) + \sum_{l \in \sigma \setminus \tau} u_l v_l \Phi'(I_l^\sigma) + \sum_{j \in [N] \setminus \sigma} u_j v_j \Phi'(I_j^\tau) + \text{Re}(\Delta_1^\sigma)
\]

\[
= \text{tr}(\Lambda(x^*(\sigma))W) - 1 + \text{Re}(\Delta_1^\sigma) < 0
\]

by the assumption that \( \Lambda(x^*(\sigma))(W + \delta) - I \) is stable and \( I_i^\tau = I_i^\sigma \) for every \( i \in \tau \) or \( i \in [N] \setminus \sigma \). Therefore, we have exhibited a fixed point \( x^*(\tau) \) that is asymptotically stable and that supports \( \tau \), so it follows that \( \tau \) is a permitted set.

Let us summarize what we have done in this chapter. We showed that rank-one synaptic weight matrices, and their perturbations, are a family of matrices for which \( \mathcal{P}_{[1]+}(W + \delta) \) is a convex combinatorial neural code. From there, we moved on to analyze firing rate models with \( C^1 \) activation functions that have finitely many singularities. One issue we addressed was that, given a state of the neuronal network, it is necessary to dichotomize the activity of neurons as active or inactive; this issue was dealt with in Definition 2. Then we proceeded to prove that \( \mathcal{P}_{\Phi}(W + \delta) \) is a convex code, which as a corollary implies that \( \mathcal{P}_{[1]+}(W + \delta) \) is a convex code.
Chapter 5

Nesting of Codewords in Firing Rate Models

Our goal is to establish relationships between the neural code $\mathcal{P}_\Phi(W)$ and the structure of $W$. One might at first be tempted to conjecture the following generalization of Lemma 6: If $A$ is an $N \times N$ real matrix with negative diagonal that is stable, then there is a stable $(N-1) \times (N-1)$ principal submatrix of $A$. In terms of permitted sets, this says that if $[N]$ is a permitted set, then there is $\sigma \subset [N]$ such that $\sigma$ is permitted and $|\sigma| = N - 1$. This leads us to the following definition:

**Definition 3.** A combinatorial neural code $\mathcal{C}$ satisfies the nesting property if given $\sigma \in \mathcal{C}$ with $|\sigma| = k$, there exists $\tau \in \mathcal{C}$ such that $\tau \subset \sigma$ and $|\tau| = k - 1$.

**Example 7.** Let $W \in \mathbb{R}^{N \times N}$ be a symmetric synaptic weight matrix such that $W_{ii} = 0$. It is known that $\mathcal{P}_{[\cdot]_+}(W)$ is a simplicial complex, so $\mathcal{P}_{[\cdot]_+}(W)$ satisfies the nesting property. (Recall that a simplicial complex is a family of subsets of set that is closed under taking subsets.)

**Example 8.** Let $W \in \mathbb{R}^{3 \times 3}$ be a synaptic weight matrix such that $W - I$ has negative diagonal. Then $\mathcal{P}_{[\cdot]_+}(W)$ satisfies the nesting property.

In general, $\mathcal{P}_{[\cdot]_+}(W)$ does not satisfy the nesting property:
Example 9. Consider

\[
W = \begin{pmatrix}
0 & 2.03 & 4.11 & 3.39 \\
4.18 & 0 & 1.46 & 2.41 \\
2.47 & 2.14 & 0 & 1.14 \\
1.24 & 0.54 & 2.78 & 0
\end{pmatrix}.
\]

Then \( W - I \) is a 4-by-4 stable matrix with no stable 3-by-3 principal submatrices.

One way the nesting property could be interpreted as, in given a neuronal network, is being able to find a stimulus that falls within the receptive of some neurons while missing the receptive fields of the rest. If a neural code satisfies the nesting property, then if a given cluster of neurons activates (and the rest are unresponsive), then it is possible to find a stimulus that activates all the neurons in the cluster except one.

It is easy to exhibit examples, with as few as three neurons, of convex RF codes that do not satisfy the nesting property. (In fact, there are convex RF codes containing the singletons that do not satisfy the nesting property. The collection of permitted sets of a network are typically codes that contain all the singletons because we usually insist that the synaptic weight matrix \( W \) has zeros along the diagonal.) The two main results in this chapter are that

1. first, \( P_{\tau_+}(W) \) always satisfies the nesting property when \( W \) is almost rank one;
   and

2. extending that result to network dynamics described by \( D\dot{x} + x = \Phi(Wx + b) \), where \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) is a function that is \( C^1 \) except possibly at finitely many points and \( \Phi'(J) \to 0 \) as \( J \to -\infty \). It will turn out that \( P_{\Phi}(W + \delta) \) satisfies the nesting when \( \operatorname{rk}(W) = 1 \) and \( \delta \) is a perturbation.

5.1 Threshold-Linear Case

Recall (section 2.2) that given a subset \( \tau = \{i_1, i_2, \ldots, i_k\} \) of \( [N] \), where \( i_1 < i_2 < \cdots < i_k \), let \( u_\tau \) denote \((u_{i_1}, u_{i_2}, \ldots, u_{i_k}) \in \mathbb{R}^{|\tau|}\).
Proposition 6. Let $W = uv^T$, where $u, v \in \mathbb{R}^N \setminus \{0\}$. Let $\sigma \subseteq [N]$ be a permitted set in $\mathcal{P}_{\{+\}}(W)$. Let $\sigma_{\geq 0} = \{i \in \sigma : u_i v_i \geq 0\}$ and $\sigma_{< 0} = \sigma \setminus \sigma_{\geq 0}$. Define

$$
\omega = \{j \in \sigma_{< 0} : -u_{\sigma_{\geq 0}}^T v_{\sigma_{\geq 0}} > u_{\sigma_{< 0} \setminus \{j\}}^T v_{\sigma_{< 0} \setminus \{j\}} - 1\}.
$$

Let $\tau \subset \sigma$, where $|\tau| = |\sigma| - 1$ or, equivalently, $\tau = \sigma \setminus \{p\}$ for some $p \in \sigma$. Then $\tau$ is a permitted set if and only if $p \in \sigma_{\geq 0} \cup \omega$. In particular, since $\sigma_{\geq 0} \cup \omega \neq \emptyset$, a permitted set $\sigma$ with $|\sigma|$ co-active neurons satisfies the nesting property.

Proof. If $\sigma$ is permitted, then $W_\sigma - I_\sigma$ is stable. Therefore, $u_\sigma^T v_\sigma - 1 < 0$ because $\text{Spec}(W_\sigma - I_\sigma) = \{-1, u_\sigma^T v_\sigma - 1\}$. Let $\tau$ be as above.

Necessity: Suppose that $\tau = \sigma \setminus \{p\}$ is permitted. We want to show that $p \in \sigma_{\geq 0} \cup \omega$.

Suppose that $p \notin \omega$ and $p \in \sigma_{< 0}$ such that $u_{\sigma_{\geq 0}}^T v_{\sigma_{\geq 0}} \leq u_{\sigma_{< 0} \setminus \{p\}}^T v_{\sigma_{< 0} \setminus \{p\}} - 1$. Since $\sigma_{\geq 0} \cup \sigma_{< 0} \setminus \{p\} = \sigma \setminus \{p\} = \tau$, it follows that $0 \geq u_\tau^T v_\tau - 1$, contradicting that $\tau$ is a permitted set.

Sufficiency: We prove the contrapositive. If $\tau = \sigma \setminus \{p\}$ with $p \notin \sigma_{\geq 0} \cup \omega$, then $u_p v_p < 0$ and

$$
-u_{\sigma_{\geq 0}}^T v_{\sigma_{\geq 0}} \leq u_{\sigma_{< 0} \setminus \{p\}}^T v_{\sigma_{< 0} \setminus \{p\}} - 1 \quad \implies \quad 0 \leq u_{\sigma_{\geq 0}}^T v_{\sigma_{\geq 0}} + u_{\sigma_{< 0} \setminus \{p\}}^T v_{\sigma_{< 0} \setminus \{p\}} - 1 = u_\tau^T v_\tau - 1,
$$

implying that $\tau$ is not permitted.

Example 10 (A rank-two network such that $\mathcal{P}_{\{+\}}(W)$ does not satisfy the nesting property.). Consider

$$
W = \begin{pmatrix}
0.1342 & 0.4913 & 0.4235 & 0.4590 \\
0.8912 & 0.5488 & 0.8560 & 0.7200 \\
0.4414 & 0.4730 & 0.5690 & 0.5292 \\
0.1328 & 0.1684 & 0.1900 & 0.1816
\end{pmatrix},
$$

which is a rank-two synaptic weight matrix. It turns out that $W - I$ is stable and the spectrum of any 3-by-3 principal submatrix of $W - I$ will include eigenvalues whose real part is positive. In other words, every 3-by-3 principal submatrix of $W - I$ is unstable, so $\mathcal{P}_{\{+\}}(W)$ does not satisfy the nesting property.
A natural follow-up question to ask would be whether the nesting property holds if \( W \) is close to being rank 1; next we show that this is indeed the case. Note that even though the Proposition 7 deals with \( N \times N \) matrices, it is applicable to any principal submatrix of \( W - I + \delta \) associated with a proper subset of \([N]\).

**Proposition 7.** Let \( W = uv^T \), where \( u, v \in \mathbb{R}^{N \times N} \setminus \{0\} \), be a synaptic weight matrix.

Then there is a neighborhood \( U \) of \( 0 \in \mathbb{R}^{N \times N} \) satisfying the following property: If \( \delta \in U \) is such that \( W - I + \delta \) is stable, then \( W - I + \delta \) has an \((N - 1) \times (N - 1)\) stable principal submatrix.

**Proof.** First some set-up:

- If there is \( p \in [N] = \{1, 2, \ldots, N\} \) such that \( u_p v_p > 0 \), then let \( 0 < \epsilon < \min\{1, u_p v_p / 2\} \). In this case, let \( \sigma = [N] \setminus \{p\} \).

- Otherwise, let \( 0 < \epsilon < 1 \) and \( \sigma = [N] \setminus \{m\} \), where \( m \) is any element of \([N]\).

- By continuity of \( \text{Spec} : \mathbb{C}^{N \times N} \to \mathcal{A}_N \) at \( W - I \), there is \( d_2 > 0 \) such that for all \( X \in \mathbb{C}^{N \times N} \) with \( \|W - I - X\|_{\max} < d_1 \), we have

  \[
  \min_{\pi \in S_N} \max_{1 \leq j \leq N} |\lambda_j(W - I) - \lambda_{\pi(j)}(X)| < \epsilon.
  \]

- By continuity of \( \text{Spec} : \mathbb{C}^{(N-1) \times (N-1)} \to \mathcal{A}_{N-1} \) at \( W_\sigma - I_\sigma \), there is \( d_3 > 0 \) such that for all \( X \in \mathbb{C}^{(N-1) \times (N-1)} \) with \( \|W_\sigma - I_\sigma - X\|_{\max} < d_2 \), we have

  \[
  \min_{\pi \in S_{N-1}} \max_{1 \leq j \leq N-1} |\lambda_j(W_\sigma - I_\sigma) - \lambda_{\pi(j)}(X)| < \epsilon.
  \]

Define \( d = \min\{d_1, d_2\} \) and \( U = \{X \in \mathbb{C}^{N \times N} : \|X\|_{\max} < d\} \). We claim that \( U \) is the neighborhood with the desired property.

Suppose \( \delta \in U \) is such that \( W - I + \delta \) is stable. We will show that \( W_\sigma - I_\sigma + \delta_\sigma \) is stable.

First observe

\[
\|W - I - (W - I + \delta)\|_{\max} = \|\delta\|_{\max} < d \leq d_2,
\]
by how we defined $\delta$. By continuity at $W - I$ and Lemma 5,

$$\text{Spec}(W - I + \delta) = \{u^Tv - 1 + \Delta_1, \Delta_2 - 1, \ldots, \Delta_N - 1\},$$

where $|\Delta_k| < \epsilon$ for all $k = 1, 2, \ldots, N$.

As for $W_\sigma - I_\sigma$,

$$||W_\sigma - I_\sigma - (W_\sigma - I_\sigma + \delta_\sigma)||_{\text{max}} = ||\delta_\sigma||_{\text{max}} \leq ||\delta||_{\text{max}} < d \leq d_2,$$

so by continuity at $W_\sigma - I_\sigma$ and Lemma 5,

$$\text{Spec}(W_\sigma - I_\sigma + \delta_\sigma) = \{u^T_\sigma v_\sigma - 1 + \tilde{\Delta}_1, \tilde{\Delta}_2 - 1, \ldots, \tilde{\Delta}_{N-1} - 1\},$$

where $|\tilde{\Delta}_k| < \epsilon$ for all $k = 1, 2, \ldots, N - 1$.

Now we proceed to argue that $W_\sigma - I_\sigma + \delta_\sigma$ is stable. First notice that for all $k = 2, \ldots, N - 1$, we have $\text{Re}(\tilde{\Delta}_k) - 1 < 0$ because $|\tilde{\Delta}_k| < \epsilon < 1$. Therefore, we focus on demonstrating that $u^T_\sigma v_\sigma - 1 + \text{Re}(\tilde{\Delta}_1) < 0$ by considering two separate cases:

1. Suppose that for all $m \in [N]$, we have $u_mv_m \leq 0$; in this case we introduced $0 < \epsilon < 1$ and $\sigma = [N] \setminus \{m\}$, where $m$ is any member of $[N]$. Then

$$u^T_\sigma v_\sigma - 1 + \text{Re}(\tilde{\Delta}_1) < 0$$

follows because $u^T_\sigma v_\sigma \leq 0$ and $\text{Re}(\tilde{\Delta}_1) - 1 < 0$.

2. Suppose there is $p \in [N]$ such that $u_pv_p > 0$, so $0 < \epsilon < \min\{1, u_pv_p/2\}$ and $\sigma = [N] \setminus \{p\}$. Let $\eta = \tilde{\Delta}_1 - \Delta_1$. Then

$$u^T_\sigma v_\sigma - 1 + \text{Re}(\tilde{\Delta}_1) = u^T v - 1 + \text{Re}(\tilde{\Delta}_1) - u_mv_m - \text{Re}(\Delta_1) + \text{Re}(\Delta_1)$$

$$= u^T v - 1 + \text{Re}(\eta) + \text{Re}(\Delta_1) - u_mv_m.$$

Observe that $u^T v - 1 + \text{Re}(\Delta_1) < 0$ by stability of $W - I + \delta$. Further,

$$\text{Re}(\eta) < 2\epsilon < u_mv_m \implies \text{Re}(\eta) - u_mv_m < 0.$$
Thus, \( u^T_\sigma v_\sigma - 1 + \text{Re}(\tilde{\Delta}_1) < 0 \).

We conclude that \( W_\sigma - I_\sigma + \delta_\sigma \) is stable. \( \square \)

### 5.2 \( C^1 \) Activation Function Case

Now that we have a working definition of permitted set for networks with \( C^1 \) activation functions, we tackle the nesting property for such neuronal network models.

**Proposition 8.** Let \( W = uv^T \), where \( u, v \in \mathbb{R}^{N \times N} \setminus \{0\} \) be a synaptic weight matrix and \( \Phi : \mathbb{R} \to \mathbb{R}_{\geq 0} \) be a \( C^1 \) activation function with finitely many singularities such that \( \Phi(J) = \Phi'(J) \to 0 \) as \( J \to -\infty \). Let \( 0 < r_1 < r_2 \) be activation thresholds. Suppose \( \sigma \) is a permitted set of \( D\dot{x} + x = \Phi(Wx + b) \). Then there is a permitted \((||\sigma|| - 1)\)-subset \( \tau \subset \sigma \).

**Proof.** Suppose that \( \Phi \) is not monotonically decreasing. Suppose \( \sigma \) is a permitted set so that \( x^* = (\Phi(I_1), \ldots, \Phi(I_N)) \) is an asymptotically stable fixed point of \( D\dot{x} + x = \Phi(Wx + b) \), where \( b = i^* - Wx^* \) and \( i^* = (I_1, \ldots, I_N) \in \Phi^{-1}(x^*) \). (Additionally, recall that, by definition, \( |\Phi'(I_j)| \geq r_2 \) for all \( i \in \sigma \) and \( |\Phi'(I_j)| < r_1 \) for all \( j \notin \sigma \).) There are two cases. We fix the following notation: If \( a \in \mathbb{R} \), let \( \text{sgn}(a) = 1 \) if \( a > 0 \) and \( \text{sgn}(a) = -1 \) if \( a < 0 \).

- Suppose that \( p \in \sigma \) such that \( \text{sgn}(u_pv_p\Phi'(I_p)) = 1 \). Let \( \tau = \sigma \setminus \{p\} \) for any such \( p \in \sigma \), and \( \tilde{I} \in \mathbb{R} \) be such that \( |\Phi'(\tilde{I})| < r_1 \) and \( \text{sgn}(\Phi'(\tilde{I})\Phi'(I_p)) = 1 \), which is possible because \( \Phi' \) is continuous and \( \Phi'(J) \to 0 \) as \( J \to -\infty \). In particular, \( u_pv_p\Phi'(I_p) > u_pv_p\Phi'(\tilde{I}) \). Let \( \tilde{x}^* \) be such that \( \tilde{x}^*_i = x^*_i \) for all \( i \in [N] \setminus \{p\} \), and \( \tilde{x}^*_p = \Phi(\tilde{I}) \), and \( \tilde{b} = \tilde{i}^* - W\tilde{x}^* \), where \( \tilde{I} \in \Phi^{-1}(\tilde{x}^*) \). Then \( \tilde{x}^* \) is a fixed point of \( D\dot{x} + x = \Phi(Wx + \tilde{b}) \) and it is stable because

\[
1 > \text{tr}(\Lambda(x^*)W) = u_pv_p\Phi'(I_p) + \sum_{i \in \tau} u_iv_i\Phi'(I_i) + \sum_{j \notin \tau} u_jv_j\Phi'(I_j) \\
> u_pv_p\Phi'(\tilde{I}) + \sum_{i \in \tau} u_iv_i\Phi'(I_i) + \sum_{j \notin \tau} u_jv_j\Phi'(I_j) \\
= \sum_{i \in \tau} u_iv_i\Phi'(\tilde{I}_i) + \sum_{j \notin \tau} u_jv_j\Phi'(\tilde{I}_j) \\
= \text{tr}(\Lambda(\tilde{x}^*)W).
\]
Thus, $\tau$ is permitted.

- Suppose that for all $m \in \sigma$ we have $\text{sgn}(u_m v_m \Phi'(I_m)) = -1$. Let $\tau = \sigma \setminus \{m\}$. Since $\Phi'(s) \to 0$ as $s \to -\infty$, it follows that there is $\bar{I} \in \mathbb{R}$ satisfying $|\Phi'(\bar{I})| < r_1$ and $\text{sgn}(\Phi'(\bar{I})\Phi'(I_m)) = 1$, and there is $J \in \mathbb{R}$ with

$$u_m v_m \Phi'(\bar{I}) + \sum_{j \notin \sigma} u_j v_j \Phi'(J) < 1.$$ 

Define $\bar{x}^*$ so that $\bar{x}^*_i = x^*_i = \Phi(I^*_i)$ for $i \in \tau$, $\bar{x}^*_m = \Phi(\bar{I})$, and $\bar{x}^*_j = \Phi(J)$ for all $j \notin \sigma$; and define $\bar{b} = \bar{I} - W \bar{x}^*$, where $\bar{I} \in \Phi^{-1}(\bar{x}^*)$. Then $\bar{x}^*$ is a fixed point of $D\bar{x} + x = \Phi(Wx + \bar{b})$. Furthermore,

$$\text{tr}(\Lambda(\bar{x}^*)W) = u_m v_m \Phi'(\bar{I}) + \sum_{i \in \tau} u_i v_i \Phi'(I_i) + \sum_{j \notin \sigma} u_j v_j \Phi'(J)$$

$$< 1 + \sum_{i \in \tau} u_i v_i \Phi'(I_i)$$

$$< 1,$$

so $\tau$ is permitted.

We conclude that there is a permitted set $\tau$ of cardinality $|\sigma| - 1$ contained in $\sigma$, so the system $D\bar{x} + x = \Phi(Wx + \bar{b})$ satisfies the nesting property of neural codes.

Next we present the main result of this chapter, namely that $\mathcal{P}(W + \delta)$ satisfies the nesting property, where $\text{rk}(W) = 1$ and $W + \delta$ is in a sense close to being a rank-one matrix.

**Theorem 5.** Let $W = uv^T$, where $u, v \in \mathbb{R}^{N \times N} \setminus \{0\}$, be a synaptic weight matrix and $\Phi : \mathbb{R} \to \mathbb{R}_{\geq 0}$ be a $C^1$ activation function with finitely many discontinuities such that $\Phi(J) = \Phi'(J) \to 0$ as $J \to -\infty$. Let $0 < r_1 < r_2$ be activation thresholds. If there is $\delta \in \mathbb{R}^{N \times N}$ such that

- $\sigma \in \mathcal{P}(W + \delta)$ is a permitted set with associated fixed point $x^*$; and
- there is $d = d(\delta) > 0$ depending on $\delta$ such that $\|\Lambda(x^*)\|_{\text{max}} \|\delta\|_{\text{max}} < d$,

then there is $\tau$, where $\tau \subset \sigma$ has $|\sigma| - 1$ neurons, such that $\tau \in \mathcal{P}(W + \delta)$.  

Proof. Let $\delta \in \mathbb{R}^{N \times N}$, and $0 < r_1 < r_2$. Suppose $\sigma \in \mathcal{P}_\Phi(W + \delta)$. Let $x^* = (\Phi(I_1), \Phi(I_2), \ldots, \Phi(I_N))$ be an asymptotically stable fixed point such that $|\Phi'(I_i)| < r_1$ for all $i \notin \sigma$ and $|\Phi'(I_j)| \geq r_2$ for all $j \in \sigma$.

Define $\epsilon$, $\bar{x}^*$, and $\Lambda_2(\bar{x}^*)$ as follows:

- Suppose there is $p \in \sigma$ such that $\text{sgn}(u_p v_p \Phi'(I_p)) = 1$. Then let $\bar{I} \in \mathbb{R}$ be such that $|\Phi'(\bar{I})| < r_1$ and $\text{sgn}(\Phi'(\bar{I})\Phi'(I_p)) = 1$, which is possible because $\Phi'$ is continuous and $\Phi'(J) \to 0$ as $J \to -\infty$. Let $\bar{x}^*$ be such that $\bar{x}_{i}^* = x_{i}^*$ for all $i \in [N] \setminus \{p\}$, and $\bar{x}_{p}^* = \Phi(\bar{I})$. Let $\bar{b} = \bar{I} - W\bar{x}^*$, where $\bar{I}^* = \Phi^{-1}(\bar{x}^*)$, and $\Lambda_2(\bar{x}^*) \in \mathbb{R}^{N \times N}$ be a diagonal matrix such that $\Lambda_2(\bar{x}^*)_{ii} = \Phi'(W^{(i)} \cdot \bar{x}^* + \bar{b}_i)$. Then define

$$0 < \epsilon < \min \left\{ 1, \frac{u_p v_p}{2} \left( \Phi'(I_p) - \Phi'(\bar{I}_p) \right) \right\}.$$

- otherwise, if $\text{sgn}(u_m v_m \Phi'(I_m)) = -1$ for all $m \in \sigma$, then let $\epsilon < 1$. Since $\Phi'(I) \to 0$ as $I \to -\infty$, it follows that there is $\bar{I} \in \mathbb{R}$ satisfying $|\Phi'(\bar{I})| < r_1$ and $\text{sgn}(\Phi'(\bar{I})\Phi'(I_m)) = 1$, and there is $J \in \mathbb{R}$ with

$$u_m v_m \Phi'(\bar{I}) + \sum_{j \notin \sigma} u_j v_j \Phi'(J) < 0.$$

Define $\bar{x}^*$ so that $\bar{x}_{i}^* = x_{i}^* = \Phi(I_i)$ for $i \in \tau$, $\bar{x}_{m}^* = \Phi(\bar{I})$, and $\bar{x}_{j}^* = \Phi(J)$ for all $j \notin \sigma$. Let $\bar{b} = \bar{I} - W\bar{x}^*$, where $\bar{I}^* = \Phi^{-1}(\bar{x}^*)$, and $\Lambda_2(\bar{x}^*) \in \mathbb{R}^{N \times N}$ be a diagonal matrix such that $\Lambda_2(\bar{x}^*)_{ii} = \Phi'(W^{(i)} \cdot \bar{x}^* + \bar{b}_i)$.

Let $\tau = \sigma \setminus \{q\}$, where $q$ is an element of $\sigma$ falling into either of the above two cases. Next, in order to set up the proof:

- By continuity at $\Lambda(x^*)W - I$, there is $d_1 > 0$ so that for all $X \in \mathbb{C}^{N \times N}$ with $||\Lambda(x^*)W - I - X||_{\max} < d_1$, it follows that

$$\min_{\pi \in \mathfrak{S}_N} \max_{1 \leq j \leq N} |\lambda_j(\Lambda(x^*)W - I) - \lambda_{\pi(j)}(X)| < \epsilon.$$

- By continuity at $\Lambda_2(\bar{x}^*)W - I$, there is $d_2 > 0$ such that for all $X \in \mathbb{C}^{N \times N}$
satisfying \( \|A_2(\bar{x}^*)W - I - X\|_{\text{max}} < d_2 \), then

\[
\min_{\pi \in \mathcal{S}_{N-1}} \max_{1 \leq j \leq N-1} |\lambda_j (A_2(\bar{x}^*)W - I) - \lambda_{\pi(j)}(X)| < \epsilon.
\]

Let \( d = \min\{d_1, d_2\} \). This \( d > 0 \) is the constant for which \( \|A(x^*)\|_{\text{max}} \|\delta\|_{\text{max}} < d_\delta \) is satisfied. The Jacobian for the perturbed system is \( A(x^*)(W + \delta) - I \). Observe

\[
\|A(x^*)W - I - (A(x^*)(W + \delta) - I)\|_{\text{max}} = \|A(x^*)\delta\|_{\text{max}} \leq \|A(x^*)\|_{\text{max}} \|\delta\|_{\text{max}} < d.
\]

Thus, by continuity at \( A(x^*)W - I \) and Lemma 5

\[
\text{Spec}(A(x^*)(W + \delta) - I) = \{\text{tr}(A(x^*)W) - 1 + \Delta_1, \Delta_2 - 1, \ldots, \Delta_n - 1\},
\]

where \( |\Delta_k| < \epsilon \) for all \( k \in [N] \). Next, note that

\[
\|A_2(\bar{x}^*)W - I - (A_2(\bar{x}^*)(W + \delta) - I)\|_{\text{max}} = \|A_2(\bar{x}^*)\delta\|_{\text{max}} \\
\leq \|A_2(\bar{x}^*)\|_{\text{max}} \|\delta\|_{\text{max}} \\
\leq \|A(\bar{x}^*)\|_{\text{max}} \|\delta\|_{\text{max}} \\
< d
\]

where we have observed that \( \|A_2(\bar{x}^*)\|_{\text{max}} \leq \|A(\bar{x}^*)\|_{\text{max}} \). Hence, by continuity of the spectrum at \( A_2(\bar{x}^*)W - I \) and Lemma 5

\[
\text{Spec}(A_2(\bar{x}^*)(W + \delta) - I) = \left\{\text{tr}(A_2(\bar{x}^*)W) - 1 + \tilde{\Delta}_1, \tilde{\Delta}_2 - 1, \ldots, \tilde{\Delta}_{N-1} - 1\right\},
\]

with \( |\tilde{\Delta}_k| < \epsilon \) for all \( k \in [N - 1] \).

The final step of the proof is showing that \( \text{Spec}(A_2(\bar{x}^*)(W + \delta) - I) \) is indeed stable.

First observe that \( \text{Re}(\tilde{\Delta}_k) - 1 < 0 \) for all \( k \in \{2, 3, \ldots, N - 1\} \) as \( \epsilon < 1 \) and \( |\tilde{\Delta}_k| < \epsilon \), so there is only one eigenvalue in the spectrum of \( A_2(\bar{x}^*)(W + \delta) - I \) left to analyze:

- First we assume \( \text{sgn}(u_mv_m \Psi'(I_m)) = -1 \) for all \( m \in \sigma \). In this case,

\[
\text{tr}(A_2(\bar{x}^*)W) - 1 + \text{Re}(\tilde{\Delta}_1) < 0
\]
since \( \text{Re}(\tilde{\Delta}_1) - 1 < 0 \) and \( \text{tr}(\Lambda_2(\tilde{x}^*)W) < 0 \).

- Assume there is \( p \in \sigma \) such that \( \text{sgn}(u_p v_p \Phi'(I_p)) = 1 \). (Recall that in this case \( 0 < \epsilon < \min\{1, \frac{u_p v_p}{2} \Phi'(I_p)\} \)). Let \( \eta = \Delta_1 - \Delta_1 \). Then

\[
\text{tr}(\Lambda_2(\tilde{x}^*)W) - 1 + \text{Re}(\Delta_1) = \gamma_1 + \text{Re}(\eta) - u_p v_p \Phi'(I_p) + u_p v_p \Phi'(\tilde{I}_p) \\
= \gamma_1 + \text{Re}(\eta) - u_p v_p \left( \Phi'(I_p) - \Phi'(\tilde{I}_p) \right),
\]

where

\[ \gamma_1 = \text{tr}(\Lambda(x^*)W) - 1 + \text{Re}(\Delta_1). \]

Since \( \text{Spec}(\Lambda(x^*)(W + P) - I) \) is stable by assumption, \( \gamma_1 < 0 \). Furthermore,

\[
\text{Re}(\eta) = \text{Re}(\Delta_1) - \text{Re}(\Delta_1) \\
< 2\epsilon \\
< u_p v_p \left( \Phi'(I_p) - \Phi'(\tilde{I}_p) \right),
\]

so that

\[
\text{Re}(\eta) - u_p v_p \left( \Phi'(I_p) - \Phi'(\tilde{I}_p) \right) < 0,
\]

which shows that

\[
\text{tr}(\Lambda_2(\tilde{x}^*)W) - 1 + \text{Re}(\Delta_1) < 0.
\]

We conclude that \( \Lambda_2(\tilde{x}^*)(W + \delta) - I \) is stable, so it follows that \( \tau \) is a permitted set.

Next we present examples of firing rate models and their permitted sets:

**Example 11** (A network that does not satisfy the nesting property.). We consider an activation that we think comes closest to the rectified-linear unit activation function \( \Phi(x) = \max\{x, 0\} \), namely the soft-plus activation function \( \Phi(x) = \ln(1 + e^x) \). Consider the network whose synaptic weights are described by

\[
W = \begin{pmatrix}
0 & -11.69 & -7.6 & -3.26 \\
8.2 & 0 & 24.7 & -18.34 \\
17.63 & -10.2 & 0 & 5.38 \\
19.91 & 14.87 & 0.64 & 0
\end{pmatrix}.
\]
Using MATLAB, $1/\|W\|_2 \approx 0.0297$. Let $r_1 = 0.01$ and $r_2 = 0.9$, so note that

$$r_1 < 1/\|W\|_2 < r_2.$$ 

In order to find the permitted sets of $D\dot{x} + x = \Phi(Wx + b)$, we fix $\sigma \subseteq [4] = \{1, 2, 3, 4\}$, and choose uniformly at random $I_k \in \mathbb{R}$ such that

$$\frac{1}{1 + e^{-I_i}} \geq r_2 \quad \text{for } i \in \sigma,$$

$$\frac{1}{1 + e^{-I_j}} < r_1 \quad \text{for } j \notin \sigma;$$

in other words,

$$I_i \geq \ln \left( \frac{r_2}{1 - r_2} \right) \quad \text{for } i \in \sigma,$$

$$I_j < \ln \left( \frac{r_1}{1 - r_1} \right) \quad \text{for } i \in \sigma.$$

Written in interval notation, $I_i \in \left( \ln \left( \frac{r_2}{1 - r_2} \right), \infty \right)$ for $i \in \sigma$, and $I_j \in \left( -\infty, \ln \left( \frac{r_1}{1 - r_1} \right) \right)$ for $j \notin \sigma$. Since $\Phi'(10) \approx 1$ and $\Phi'(-20) \approx 0$, we choose $I_k$ uniform randomly from

$$I_i \in \left( \ln \left( \frac{r_2}{1 - r_2} \right), 10 \right) \quad \text{for } i \in \sigma,$$

$$I_j \in \left( -20, \ln \left( \frac{r_1}{1 - r_1} \right) \right) \quad \text{for } j \notin \sigma.$$

For each subset $\sigma$, we generate up to 100,000 different $I_k$'s and check whether $\Lambda(x^*)W - I$ is stable, where $x^* = (\Phi(I_1), \ldots, \Phi(I_4))$. After running the code, we find that

$$\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34, 1234\}$$

so $\mathcal{P}_\Phi(W)$ does not satisfy the nesting property.

**Example 12** (Network with exponential activation function. Combinatorial neural
code does not satisfy the nesting property.). Let

\[
W = \begin{pmatrix}
0 & -0.24 & 0.8 & 0.48 \\
0.42 & 0 & -0.4 & 0.59 \\
0.52 & -0.78 & 0 & 1.11 \\
0.58 & 1.49 & 0.08 & 0 \\
\end{pmatrix},
\]

so that \(1/\|W\|_2 \approx 0.53553\). Suppose \(\Phi(x) = e^x\). Let \(r_1 = 3.9\) and \(r_2 = 7.8\). Then \(\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 1234\}\). Next we show the resulting neural code for a variety of choices of \(r_1 < r_2\):

- \(r_1 = 3.9\) and \(r_2 = 5\): \(\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 123, 1234\}\).
- \(r_1 = 3.9\) and \(r_2 = 4\): \(\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 123, 124, 1234\}\).
- \(r_1 = 1\) and \(r_2 = 3\): \(\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 34, 1234\}\).
- \(r_1 = 0.3\) and \(r_2 = 1\): \(\mathcal{P}_\Phi(W) = 2^{[4]} \setminus \{124, 134\}\).

**Example 13** (A network that does not satisfy the nesting property, with less extreme firing thresholds.). Let us consider again the activation function \(\Phi(x) = \ln(1 + e^x)\). Let

\[
W = \begin{pmatrix}
0 & -4.56 & -21.32 & 21.51 \\
-10.57 & 0 & 21.76 & 10.63 \\
9.53 & -10.27 & 0 & -21.74 \\
-11.65 & -11.02 & 3.47 & 0 \\
\end{pmatrix},
\]

which has norm \(1/\|W\|_2 \approx 0.028\). Let \(r_1 = 0.2\) and \(r_2 = 0.6\). Following a similar procedure as in the previous example, we find that

\(\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34, 1234\}\),

so the associated combinatorial code does not satisfy the nesting property.

**Example 14** (Another network that does not satisfy the nesting property.). Consider the network with activation function \(\Phi(x) = 1/(1 + e^{-x})\)—that is, the standard logistic
function—whose derivative is $\Phi'(x) = e^{-x}/(1 + e^{-x})^2 = 1/(2 + e^x + e^{-x})$. Consider

$$W = \begin{pmatrix}
0 & -45.8509 & 20.6376 & -0.7266 \\
28.1654 & 0 & -29.8098 & -40.5420 \\
-38.0833 & 17.2577 & 0 & 21.9630 \\
31.9149 & 3.2307 & 31.4733 & 0
\end{pmatrix},$$

which has norm $1/||W||_2 \approx 0.0141$. Note that $\Phi'(I) \leq 0.25$ for all $I \in \mathbb{R}$. Let $r_1 = 0.01$ and $r_2 = 0.15$. Then

$$\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 1234\}.$$

We end this chapter by exhibiting an example of a network with rank-two synaptic weight matrix such that $\mathcal{P}_\Phi(W)$ does not satisfy the nesting property.

**Example 15** (A rank-two network that whose combinatorial neural code does not satisfy the nesting property). Consider again the standard logistic activation function and

$$W = \begin{pmatrix}
3.3507 & -32.7518 & 38.0737 & 24.9745 \\
9.2577 & 8.7499 & -20.1370 & -3.0590 \\
-13.4458 & 2.7231 & 9.7583 & -6.7624 \\
21.8722 & 34.7394 & -65.3406 & -17.4414
\end{pmatrix}.$$

Here $W$ is a rank-two matrix and $1/||W||_2 \approx 0.0102$. Let $r_1 = 0.01$ and $r_2 = 0.15$. Then

$$\mathcal{P}_\Phi(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 1234\}.$$

**Example 16.** We remark that a logistic activation and soft-plus functions can agree on the existence of gaps in the associated combinatorial neural codes. To illustrate this point, we revisit two examples above where we exhibited instances of four-neuron networks whose combinatorial neural code admits no codewords with three neurons, i.e., sets of with three neurons are forbidden. Let $\Phi_1(x) = \log(1+e^x)$, $\Phi_2(x) = 4/(1+e^{-x}) = 4\Phi_1'(x)$, $\Psi_1(x) = \log(1 + e^x)/4$, and $\Psi_2(x) = 1/(1 + e^{-x}) = \Phi_2(x)/4$. This implies that $\Phi_1'(x) < 1$, $\Phi_2'(x) = 4/(2 + e^{-x} + e^x) \leq 1$, $\Psi_1'(x) = \Phi_1' < 1/4$, and $\Psi_2'(x) \leq 1/4$ for all $x \in \mathbb{R}$. Then
• When

\[
W = \begin{pmatrix}
0 & -4.5617 & -21.3156 & 21.5149 \\
-10.5739 & 0 & 21.7604 & 10.6261 \\
9.5333 & -10.2711 & 0 & -21.7423 \\
-11.6547 & -11.0190 & 3.4710 & 0
\end{pmatrix},
\]

we saw that \( ||W||_2 \approx 0.028 \), and we set \( r_1 = 0.2 \) as well as \( r_2 = 0.6 \). We found that \( \mathcal{P}_{\Phi_1}(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 34, 1234\} \) for the network whose activation function is \( \Phi_1 \). When the activation function is \( \Phi_2 \), our calculations suggest that \( \mathcal{P}_{\Phi_2}(W) = \mathcal{P}_{\Phi_1}(W) \). (For the latter activation function, we used the same activation thresholds \( r_1, r_2 \).)

• When

\[
W = \begin{pmatrix}
0 & -45.8509 & 20.6376 & -0.7266 \\
28.1654 & 0 & -29.8098 & -40.5420 \\
-38.0833 & 17.2577 & 0 & 21.9630 \\
31.9149 & 3.2307 & 31.4733 & 0
\end{pmatrix},
\]

whose norm is \( ||W||_2 \approx 0.0141 \). After setting \( r_1 = 0.01 \) and \( r_2 = 0.15 \), we determined \( \mathcal{P}_{\Phi_2}(W) = \{1, 2, 3, 4, 12, 13, 14, 23, 24, 1234\} \). On the other hand, changing the activation function to \( \Psi_1 \) yields \( \mathcal{P}_{\Psi_1}(W) = \mathcal{P}_{\Psi_2}(W) \).
Chapter 6

Conclusion

Now we summarize our findings.

We started with threshold-linear networks, \( D\dot{x} + x = [Wx + b]_+ \), and their collection of sets permitted sets, \( \mathcal{P}[\cdot]_+ (W) \). Aside from the \( \mathcal{P}[\cdot]_+ (W) \) forming a simplicial complex when \( W \) is symmetric and Lemma 3, the “2-by-2 Minor Lemma,” not much known about constraints on the combinatorial neural code that must be satisfied so that the given code can arise as the neural code of a threshold-linear network. In that regime, given \( \sigma \in [N] = \{1, 2, \ldots, N\} \), stability of a fixed point of the dynamics comes down to determining whether \((W - I)_\sigma\) is stable. The spectrum of such a matrix is far too general to run a detailed theoretical analysis that will yield further properties about \( \mathcal{P}[\cdot]_+ (W) \). Our way of addressing this issue was to assume that \( W \) is a rank-one matrix, in which case the relationship between the spectrum of \((W - I)_\sigma\) and its principal submatrices is tractable. The rank-one hypothesis, therefore, although not biologically plausible, gave us insights into the structure of \( \mathcal{P}[\cdot]_+ (W) \), which in the scope of this dissertation involved proving convex coding and nesting properties. Furthermore, we can bootstrap from rank-one matrices due the fact that eigenvalues of a matrix are continuous functions its entries.

Inspired by the results in the rank-one, threshold-linear networks scenario, we used similar ideas to prove that \( \mathcal{P}_\Phi (W) \) is a convex code and satisfies the nesting property. The hypotheses the synaptic weight matrices satisfies are the same as for the threshold-linear regime. Before extending the techniques to general firing rate models we define a way of stipulating which neurons are inactive and active, and this enabled us to
introduce a combinatorial neural code for firing rate models, namely the collection of permitted sets, which generalizes the notion of permitted sets from threshold-linear networks.

We remark that although the results in the threshold-linear regime are interesting, it is a simplified scenario and it would be difficult to apply theoretical results to actual neuronal data. The reason why this is the case is that neurons’ responses are graded (i.e., analog), and the level of activity where one might want to consider the neuron active may vary. Our results for $C^1$ activation functions (with finitely many discontinuities) enable one to choose thresholds for determining whether a neuron is responsive. Since our results are independent of activity threshold, one could apply the theoretical results to network based on the choices of threshold.

Let us now discuss natural next directions. The theory of permitted sets of threshold-linear networks, which was introduced in the early 2000s, is well established. A natural direction to pursue is using our criterion for activation thresholds of neurons and determine whether results holding for threshold-linear networks are applicable for more general firing rate models. Neurobiologically relevant phenomena that have already been modeled in the context of threshold-linear networks are the formation of flexible memories and learning rules for encoding a given neural code.
References


Appendix A

A.1 Proof of Lemma 4

Lemma 12. Let $D \in \mathbb{R}^{N \times N}$ be a diagonal matrix. Then

$$|||D||| = \max_{1 \leq i \leq N} |d_i| = ||D||_{\text{max}},$$

where $d_i$ denotes the $i$th entry along the diagonal of $D$.

Proof. Since $|||D||| = \max_{||x||=1} ||Dx||$, where $|| \cdot ||$ denotes the Euclidean norm, we will prove the claim by using Lagrange multipliers: We will find the extrema of $||Dx||^2 = Dx \cdot Dx$ subject to the constraint $x \cdot x - 1 = 0$. Let $x_0 = (x_0^{(1)}, \ldots, x_0^{(N)}) \in \mathbb{R}^N$ be a point where an extremum of $||Dx||^2$ is attained. Let $f(x) = ||Dx||^2$ and $g(x) = x \cdot x - 1$. Then we must find all $x_0$ satisfying the system of equations

$$\begin{align*}
\nabla f(x) &= \lambda \nabla g(x), \\
g(x) &= 0,
\end{align*}$$

where $\nabla f, \nabla g$ denote the gradients of $f, g$, respectively. Since

$$\nabla f(x) = (2d_1^2 x_0^{(1)}, \ldots, 2d_N^2 x_0^{(N)}) \quad \text{and} \quad \nabla g(x) = (2x_0^{(1)}, \ldots, 2x_0^{(N)}),$$

we have $x_0^{(i)} (d_i^2 - \lambda) = 0$ for all $i \in \{1, 2, \ldots, N\}$. Since $x_0 \neq 0$ (for $g(x_0) = 0$ would not be satisfied otherwise), there is $i_0 \in [N]$ such that $x_0^{(i_0)} \neq 0$. Therefore, $d_{i_0}^2 = \lambda$. Define
\( S = \{ i \in [N] : d_i^2 = d_{i0}^2 \} \). Then
\[
||Dx_0||^2 = \sum_{i=1}^{N} d_i^2 (x_0^{(i)})^2 = \sum_{i \in S} d_i^2 \sum_{i \in S} (x_0^{(i)})^2 = d_{i0}^2
\]
because \( x_0 \cdot x_0 = \sum_{i=1}^{N} (x_0^{(i)})^2 = \sum_{i \in S} (x_0^{(i)})^2 = 1 \) as \( x_0^{(j)} = 0 \) for all \( j \notin S \). Therefore, the maximum of \( ||Dx_0||^2 \) will be whichever is the largest \( d_i^2 \), which implies that \( ||Dx|| = \max_{1 \leq i \leq N} |d_i| \).

### A.2 Examples of Nonconvex Codes

In Section 2.4.2, we said that \( \mathcal{C} = \{\emptyset, 1, 2, 3, 4, 23, 24, 123, 124\} \) is an instance of a non-convex code (due to Vladimir Itskov, personal communication). Let us see why \( \mathcal{C} \) is nonconvex: Suppose that \( \mathcal{U} = \{U_1, U_2, U_3, U_4\} \) is a collection of subsets of some stimuli space \( X \). Observe that
\[
\mathcal{C} = \{0000, 1000, 0100, 0010, 0001, 0110, 0101, 1110, 1101\},
\]
and we can deduce from the missing codewords that \( U_1 \) is nonconvex. (Recall that there is a correspondence between binary strings and subsets, so a subset of \( \{1, 2, 3, 4\} \) such as \( \{1\} \) can be also expressed as 1000, and so on.) Specifically, since 1100 is missing, we know that
\[
(U_1 \cap U_2) \setminus (U_3 \cup U_4) = \emptyset,
\]
by definition of \( \mathcal{C}(\mathcal{U}) \). so \( U_1 \cap U_2 \subseteq U_3 \cup U_4 \). As a result,
\[
U_1 \cap U_2 = (U_1 \cap U_2 \cap U_3) \cup (U_1 \cap U_2 \cap U_4).
\]
Notice that \( U_1 \cap U_2 \neq \emptyset \) because 123 \( \in \mathcal{C} \) (so \( U_1 \cap U_2 \cap U_3 \neq \emptyset \)). Furthermore, \( U_1 \cap U_2 \cap U_3 \) and \( U_1 \cap U_2 \cap U_4 \) are disjoint because 1111 \( \notin \mathcal{C} \). Hence, \( U_1 \cap U_2 \) is nonconvex because we have just shown that \( U_1 \cap U_2 \) is disconnected. Since the intersection of two convex sets is convex, either \( U_1 \) or \( U_2 \) is nonconvex.
In section 3.2, we asked whether there were threshold-linear networks such that \( \mathcal{P}_{[\cdot]}(W) \) was a nonconvex code and \( W \) satisfied Dale’s law. We said that \( \mathcal{P}_{[\cdot]}(W) = \{\emptyset, 1, 2, 3, 4, 12, 14, 23, 34, 123, 134\} \) is the collection of permitted sets of threshold-linear network with synaptic weight matrix

\[
W = \begin{pmatrix}
0 & 0.5 & 2 & 1 \\
2 & 0 & -0.1 & -0.9 \\
0.5 & -0.7 & 0 & 0.5 \\
1 & -0.7 & 2 & 0
\end{pmatrix}.
\]

The argument for showing that \( \mathcal{P}_{[\cdot]}(W) \) is nonconvex is identical to the one presented for \( \mathcal{C} \), but here it is: Assume \( \mathcal{U} = \{U_1, U_2, U_3, U_4\} \) is a collection of subsets of some stimuli space \( X \). Since \( 13 \notin \mathcal{P}_{[\cdot]}(W) \),

\[
(U_1 \cap U_3) \setminus (U_2 \cup U_4) = \emptyset,
\]

so \( U_1 \cap U_3 \subseteq U_2 \cup U_4 \), which implies that \( U_1 \cap U_3 = (U_1 \cap U_2 \cap U_3) \cup (U_1 \cap U_3 \cap U_4) \). The fact that \( 134 \in \mathcal{P}_{[\cdot]}(W) \) implies that \( U_1 \cap U_2 \neq \emptyset \). Furthermore, \( U_1 \cap U_2 \cap U_3 \) and \( U_1 \cap U_2 \cap U_4 \) are disjoint because \( 1234 \notin \mathcal{P}_{[\cdot]}(W) \). Hence, we have shown that \( U_1 \cap U_3 \) is disconnected, so \( U_1 \cap U_3 \) is not convex. Since \( U_1 \) or \( U_2 \) must be nonconvex, it follows that \( \mathcal{P}_{[\cdot]}(W) \) is not a convex code.