MINIMAL EMBEDDING DIMENSIONS OF CONNECTED NEURAL CODES

RAFFAELLA MULAS AND NGOC M. TRAN
MINIMAL EMBEDDING DIMENSIONS OF CONNECTED NEURAL CODES

RAFFAELLA MULAS AND NGOC M. TRAN

A receptive field code is a recently proposed deterministic model of neural activity patterns in response to stimuli. The main question is to characterize the set of realizable codes, and their minimal embedding dimensions with respect to a given family of receptive fields. Here we answer both of these questions when the receptive fields are connected. In particular, we show that all connected codes are realizable in dimension at most three. To our knowledge, this is the first family of receptive field codes for which both the exact characterization and minimal embedding dimension are known.

1. Introduction

A receptive field code is a deterministic model of neural activity patterns in response to stimuli defined by Curto, Itskov, Veliz-Cuba and Youngs [5]. It consists of \( n \in \mathbb{N} \) neurons, each neuron \( i \in [n] = \{1, 2, \ldots, n\} \) has a receptive field \( U_i \subseteq \mathbb{R}^d \). Given a stimulus \( x \in \mathbb{R}^d \), the neurons generate a codeword \( \sigma(x) \in 2^{[n]} \) via

\[
i \in \sigma(x) \iff x \in U_i.
\] (1)

A receptive field code \( C(U) \subseteq 2^{[n]} \) is the set of all possible codewords generated from the collection of receptive fields \( U = (U_1, \ldots, U_n) \). Without loss of generality, we can assume that every receptive field code includes the empty set \( (\emptyset \in C) \), which is equivalent to assuming that \( \bigcup_{i \in [n]} U_i \subseteq \mathbb{R}^d \).

The minimal embedding problem is the following: given a code \( C \subseteq 2^{[n]} \) and a family \( \mathcal{F} = (\mathcal{F}_d, d \geq 1) \), where \( \mathcal{F}_d \) is a collection of sets in \( \mathbb{R}^d \), find the smallest \( d \) such that \( C = C(U) \) for some \( U \subseteq \mathcal{F}_d \). Call this smallest \( d \) the minimal embedding dimension of the code \( C \) with respect to the family \( \mathcal{F} \), denoted \( d^*(C, \mathcal{F}) \).

This paper focuses on connected codes. These are codes realizable by connected sets in \( \mathbb{R}^d \), for some \( d \in \mathbb{N} \), which are either all closed or all open, following the convention of [5]. Our main results completely characterize realizability and minimal embedding dimensions of connected codes. In particular, it is easy to check if a code is connected, and if it is, then the minimal embedding dimension is at most three. We state our main results below. Characterization of the minimal embedding dimension for the case \( d^* = 2 \) is given in terms of the graph of a family of connected sets \( U \) (cf. Definition 9).

**Proposition 1** (Realizability of connected codes). A code \( C \) is connected if and only if for each \( \sigma, \tau \in C \) and for each \( i \in \sigma \cap \tau \), there exists a sequence of distinct codewords \( v_1, \ldots, v_m \in C \) such that:

- \( \sigma = v_1 \);
- \( v_j \subseteq v_{j+1} \) or \( v_{j+1} \subseteq v_j \), for every \( j \in [m-1] \);
- \( v_m = \tau \);
- \( i \in v_j \) for each \( j \in [m] \).

**MSC2010:** 92B20, 05C65, 05C10.

**Keywords:** none, receptive field code, minimal embedding, realizability.
Figure 1. Four receptive fields generating the connected code \( \mathcal{C}(\mathcal{U}) = \{ \emptyset, 1, 2, 3, 4, 12, 13, 23, 24, 123 \} \).

**Theorem 2** (minimal embedding of connected codes). Suppose \( \mathcal{C} \) is a connected code. Let \( d^*(\mathcal{C}) \) denote its minimal embedding dimension with respect to the family of connected sets.

- \( d^*(\mathcal{C}) = 1 \) if and only if the sensor graph of \( \mathcal{C} \) is bipartite [22].
- Else, \( d^*(\mathcal{C}) = 2 \) if and only if there exists a realization \( \mathcal{C}(\mathcal{U}) = \mathcal{C} \) by connected sets \( \mathcal{U} \) in dimension 3 such that the graph of \( \mathcal{U} \) is planar.
- Else, \( d^*(\mathcal{C}) = 3 \).

1.1. **Related literature.** The minimal embedding dimension \( d^*(\cdot, \mathcal{F}) \) and the family \( \mathcal{F} \) of receptive fields form a trade-off in measuring the complexity of the signal encoded by the neurons, and is thus of particular interest in receptive field coding. In the extreme case where \( \mathcal{F}_d \) is the collection of all sets in \( \mathbb{R}^d \), then any code is realizable in dimension one [5].

**Lemma 3.** Let \( \mathcal{C} \) be a code on \( n \) neurons. For any \( d \geq 1 \), \( \mathcal{C} = \mathcal{C}(\mathcal{U}) \) for some sequence of sets \( \mathcal{U} \) in \( \mathbb{R}^d \).

There has been a number of work on criteria for realizability and bounds for \( d^*(\cdot, \mathcal{F}) \) when the set \( \mathcal{F} \) consists of (open or closed) convex sets [3; 6; 4; 5; 10; 11; 12; 13; 15; 16; 17; 22; 18]. However, complete characterization and the exact minimal embedding dimension of convex codes remain a problem. Giusti and Itskov [12] found necessary conditions for a code to be realizable with open convex sets, and proved lower bounds on the embedding dimensions of such codes. In [3], Cruz, Giusti, Itskov and Kronholm proved that there exists a family of codes, called max-intersection-complete codes, that are both open convex and closed convex, and they gave an upper bound for their embedding dimension. To the best of our knowledge, connected codes form the first family of receptive field codes for which an intrinsic characterization and the exact embedding dimension is known. Furthermore, our proof gives explicit constructions for the code realization in each dimension.

There are biologically observed receptive fields which are connected but not convex. A prominent example are the center-surround fields in the ganglion cells on the retina [7; 1], which approximately have the shape of a torus in dimension two. In some other cases, such as the grid cells in entorhinal cortex, the receptive field of a single cell is neither convex nor connected [14]. Even amongst place cells of the hippocampus, which generally have a single convex receptive field, it is shown that certain cells have disconnected receptive fields [20]. Such a code with disconnected receptive fields may be realized as the
projection of a connected code in higher dimensions. In comparison, these codes cannot be realized as projections of convex codes in any dimension, because such projections preserve convexity. This gives us motivations for studying connected codes.

In dimension 1, connected codes equal convex codes. This was completely characterized by Rosen and Zhang [22] via the sensor graph and it is included in Theorem 2 for completeness. We do not define the sensor graph of a code here, but note that it is an intrinsic property of the code, independent of any realization $U$.

Jeffs [16, Theorem 4.1] gave an alternative characterization of connected codes, which is distinct from our approach as it is formulated in terms of the links in the code, that we do not define here.

Receptive field codes are closely related to Euler diagrams, which found applications in information systems, statistics and logic [8; 21]. Since their main applications are in visualization, the literature on Euler diagrams focus exclusively on 2 and 3 dimensions. Translated to our setup, an Euler diagram in $\mathbb{R}^3$ is a collection $\mathcal{U} = (U_1, \ldots, U_n)$ of closed, orientable surfaces embedded in $\mathbb{R}^3$. An Euler diagram in $\mathbb{R}^2$ is a similar collection of closed curves embedded in $\mathbb{R}^2$. A diagram description is a code $C$ such that $\emptyset \in C$. The description of an Euler diagram $\mathcal{U}$ is the code $C(\mathcal{U})$, where $\mathcal{U}^o := (U_1^o, \ldots, U_n^o)$ consists of the relative interior of the sets $U_i$’s. The main problem in this literature is realizability: given a code $C$, is there an Euler diagram $\mathcal{U}$ such that $C = C(\mathcal{U})$? Every code $C$ can be realized by an Euler diagram in dimension 2 [21], and by an Euler diagram in dimension 3 with connected sets $U_i$’s [9]. Note the crucial difference to receptive field codes: in an Euler diagram, codewords are generated by intersection of the relative interior of the $U_i$’s. In particular, all codes $C$ which fail the condition of Proposition 1 satisfy $C = C(\mathcal{U})$ for some tuple of closed connected sets $\mathcal{U}$ in $\mathbb{R}^3$, but $C \neq C(\mathcal{U})$ for any tuple of closed connected sets in any dimension.

2. Proof of the main results

We shall first give one definition and state a lemma that can be found in [5] and that will be useful for the proof of our main results.

**Definition 4.** Let $C$ be a code on $n$ neurons. We say that $C$ is realizable by an atom sequence $A = (A_{\sigma} \subseteq \mathbb{R}^d, \sigma \subseteq [n])$ if $A_{\sigma} \neq \emptyset \iff \sigma \in C$. In this case, write $C = C(A)$.

**Lemma 5.** Let $C$ be a code on $n$ neurons. Then $C = C(A)$ if and only if $C = C(\mathcal{U})$, where

$$U_i = \bigcup_{i \in \sigma} A_{\sigma}, \quad (2)$$

or equivalently,

$$A_{\sigma} = \left( \bigcap_{i \in \sigma} U_i \right) \setminus \bigcup_{j \not\in \sigma} U_j, \quad (3)$$

with the convention that $A_{\emptyset} = \mathbb{R}^d \setminus \bigcup_{i \in [n]} U_i$.

In other words, $A$ and $\mathcal{U}$ determine each other.
Lemma 6. A code is realizable with closed connected sets in \( \mathbb{R}^d \) if and only if it is realizable with open connected sets in \( \mathbb{R}^d \).

Proof. Given \( C = C(U) \) where \( U = (U_1, \ldots, U_n) \) is a family of closed connected sets in \( \mathbb{R}^d \), we can always construct a family of open connected sets in \( \mathbb{R}^d \), \( \hat{U} = (\hat{U}_1, \ldots, \hat{U}_n) \), such that \( U_i \subset \hat{U}_i \) for each \( i \in [n] \). The fact that the \( U_i \)'s are contained in the \( \hat{U}_i \)'s implies that the old intersections are preserved and, furthermore, we can take the \( \hat{U}_i \)'s small enough to avoid the formation of new atoms. In this way we get a family of open connected sets in \( \mathbb{R}^d \) with \( C(U) = C(\hat{U}) \).

Vice versa, if \( C = C(U) \) where \( U = (U_1, \ldots, U_n) \) is given by open connected sets, we can let \( \tilde{U} = (\tilde{U}_1, \ldots, \tilde{U}_n) \) be a family of closed connected sets in \( \mathbb{R}^d \) such that \( \tilde{U}_i \subset U_i \) for each \( i \in [n] \). The fact that the \( \tilde{U}_i \)'s are contained in the \( U_i \)'s implies that no new intersections are created, and we can take the \( \tilde{U}_i \)'s big enough to preserve the old atoms. In this way we get a family of closed connected sets in \( \mathbb{R}^d \) such that \( C(U) = C(\tilde{U}) \).

Example 7. As an example of how Lemma 6 works, let \( U \) consists of three closed line segments in the plane meeting at a “T”, so that \( C(U) = \{ \emptyset, 1, 2, 3, 123 \} \). Since we are in \( \mathbb{R}^2 \), open sets \( \tilde{U}_i \)'s containing the line segments \( U_i \)'s must be full-dimensional. Therefore we can take open rectangles \( \tilde{U}_i \)'s containing the old \( U_i \)'s, such that they intersect all together but not pairwise. In this way we have that \( C(U) = C(\tilde{U}) \).

Definition 8. We say that two sets \( A, B \subset \mathbb{R}^d \) are adjacent if \( A \cap B = \emptyset \) and either \( \overline{A} \cap B \neq \emptyset \) or \( A \cap \overline{B} \neq \emptyset \), where \( \overline{A} \) denotes the closure of \( A \) in the Euclidean topology.

Definition 9 (graph of a realization). Let \( C = C(U) \) be a connected code with realization \( U \). Let \( A \) be the atoms defined via \( U \) in (3). The graph of \( U \), denoted \( G(U) \), is a graph with one vertex for every connected component of each atom \( A_\sigma \) with \( \sigma \neq \emptyset \), and an edge for every pair of connected components of atoms that are adjacent.

Lemma 10. Let \( C = C(U) \) be a connected code with realization \( U \) in dimension \( d \). If \( G(U) \) can be embedded in \( \mathbb{R}^{d'} \), then \( C \) can also be realized by connected sets in dimension \( d' \).

Proof. Take an embedding of \( G(U) \) in \( \mathbb{R}^{d'} \). Let \( A \) be the atoms defined via \( U \) in (3). Let \( v_\sigma \in \mathbb{R}^{d'} \) be the realization of the vertex of \( G(U) \) indexed by the \( j \)-th component of the atom \( A_\sigma \). For each pair of nodes \( v_\sigma \) and \( v_\tau \), let \( e_{\sigma, \tau} \subset \mathbb{R}^{d'} \) be the realization of the edge between these nodes. If they are not connected, define \( e_{\sigma, \tau} = \emptyset \). Now define atoms \( A' \) in \( \mathbb{R}^{d'} \) via

\[
A'_\sigma := \bigcup_j \left( v_\sigma \cup \bigcup_{\tau k \subset \tau} e_{\sigma, \tau} \right) \subset \mathbb{R}^{d'},
\]

for \( \sigma \in C \{ \emptyset \} \), and

\[
A'_\emptyset := \mathbb{R}^{d'} \setminus \bigcup_{\sigma \in C \{ \emptyset \}} A'_\sigma.
\]

It is easy to check that \( C = C(A') \), so \( C \) is realizable in dimension \( d' \), as needed. \( \square \)
2.1. Proof of Proposition 1. Let $C$ be a code on $n$ neurons. By Lemma 3, $C = C(\mathcal{U}) = C(\mathcal{A})$ for some $U_i, A_\sigma \subseteq \mathbb{R}^d, i \in [n], \sigma \subseteq [n]$. For each $i \in [n], U_i$ is connected if and only if for every $\sigma, \tau \subseteq [n]$ such that $i \in \sigma \cap \tau$, from each connected component $C_\sigma$ of $A_\sigma$ to each connected component $C_\tau$ of $A_\tau$, there is a path $C_\sigma = C_{v_1} \rightarrow C_{v_2} \cdots \rightarrow C_{v_{m-1}} \rightarrow C_{v_m} = C_\tau$ in $\mathcal{G}(\mathcal{C}(\mathcal{U}))$, where $C_{v_j} \subseteq A_{v_j}$ is a connected component of $A_{v_j}$, such that $A_{v_j} \subseteq U_i$ for each $j \in [m]$. Note that, in order to have the receptive fields either all open or all closed, two connected components $C_\sigma \subseteq A_\sigma, C_\tau \subseteq A_\tau$ are allowed to be adjacent if and only if either $\sigma \subseteq \tau$ or $\tau \subseteq \sigma$. Hence $U_i$ is allowed to be connected if and only if for every $\sigma, \tau$ such that $i \in \sigma \cap \tau$, there exists a sequence of distinct codewords $v_1, \ldots, v_m \in C$ such that:

- $\sigma = v_1$,
- either $v_j \subseteq v_{j+1}$ or $v_{j+1} \subseteq v_j$, for every $j \in [m-1],$
- $v_m = \tau$ and
- $i \in v_j$ for each $j \in [m]$.

This proves the proposition. □

2.2. Proof of Theorem 2. We split the statement of Theorem 2 into two parts, and prove them separately. The first part, Proposition 11 states that the minimal embedding dimension for a connected code is at most 3. The second part, Proposition 13 gives a characterization for connected codes with $d^* = 2$. For the case $d^* = 1$, see [22, Proposition 1.9 and Theorem 3.4].

Proposition 11. Let $C$ be a connected code on $n$ neurons. Then $C$ is realizable by connected sets in dimension 3.

Proof: For all $\sigma \in C \setminus \{\emptyset\}$, choose disjoint balls $B_\sigma \subset \mathbb{R}^3$ and for all $\sigma, \tau \in C$ such that $\sigma \subset \tau$, let $T_{\sigma, \tau} \subset \mathbb{R}^3$ be a tube that connects $B_\sigma$ and $B_\tau$. Since we are in $\mathbb{R}^3$, the tubes can always be arranged so that they do not intersect with each other and this can be proved by induction the number of tubes. Given $m$ disjoint tubes between $|C \setminus \{\emptyset\}|$ balls, suppose we need to construct a tube $T_{\sigma, \tau}$ joining the balls $B_\sigma$ and $B_\tau$. Since the number $m$ of existing tubes is finite, we can pick a point $s \in B_\sigma$ and a point $t \in B_\tau$ such that their projections in the $(0, 0, 1)$ direction is larger than that of any other point on the $m$ existing tubes. Now join $s$ and $t$ by a tube $T_{\sigma, \tau}$ such that its projection onto the $(0, 0, 1)$ direction is larger than that of all other tubes. Thus, $T_{\sigma, \tau}$ is disjoint from the first $m$ tubes, completing the induction argument. Now, let

$$A_\sigma := B_\sigma \cup \bigcup_{\sigma \subset \tau} T_{\sigma, \tau}$$

for $\sigma \in C \setminus \{\emptyset\}$ and let

$$A_\emptyset := \mathbb{R}^3 \setminus \bigcup_{\sigma \in C \setminus \{\emptyset\}} A_\sigma.$$ 

Then $C = C(A)$. Define $\mathcal{U}$ from $\mathcal{A}$ via (2). By Lemma 5, $C = C(\mathcal{U})$. By construction of $\mathcal{A}$ and since we are assuming that $C$ satisfies Proposition 1, the $U_i$’s are connected. This completes the proof. □

Remark 12. The proof of Proposition 11 uses a classical technique that has been used for example also in [3, Lemma 2.2], [16, Lemma 4.4] and [9, Lemma 4.1].
Proposition 13. Let $C$ be a connected code on $n$ neurons. Then $d^*(C) = 2$ if and only if there exists a realization $C = C(U)$ by connected sets in $\mathbb{R}^3$ such that $G(U)$ is planar.

Proof: Suppose $d^*(C) = 2$. Then there exists a realization $C = C(U)$ with $U$ a collection of connected sets $U$ in $\mathbb{R}^2$. The graph of $U$, $G(U)$, is by construction also embedded in $\mathbb{R}^2$. One can trivially embed a realization in $\mathbb{R}^2$ into $\mathbb{R}^3$ without changing the graph $G(U)$, so we are done. Conversely, suppose that $C = C(U)$ for some $U$ in $\mathbb{R}^3$ such that $G(U)$ is planar. By Lemma 10, $C$ is realizable in dimension 2. □

We conclude our paper with two examples.

Example 14 (connected code with $d^* = 3$). Consider the following code

$$C = \{\emptyset, 1, 2, 3, 4, 5, 12, 13, 14, 15, 23, 24, 25, 34, 35, 45\}. \quad (4)$$

This satisfies Proposition 1, so $C$ is a connected code. It’s easy to see that every graph $G(U)$ associated to this code must be a subdivision of the graph in Figure 3; i.e., if $C = C(U)$, then $G(U)$ must be either the graph in in Figure 3 or it can be obtained from it by subdividing some of its edges into two new edges, which must be connected to a new vertex. This is due to the fact that $C$ is the code that contains exactly every $i \in [5]$ and every pair $ij$ of distinct $i, j \in [5]$. This implies, by Kuratowski’s theorem [2], that every graph associated to $C$ is not planar. By Theorem 2, $C$ has minimal embedding dimension 3.

Figure 3. The graph of a connected realization of a code $C$ with $d^*(C) = 3$ in Example 14.
Figure 4. The realization of a connected code $C$ with $d^*(C) = 2$ in Example 15 and its graph.

Example 15 (connected code with $d^* = 2$). Let $C = \{\emptyset, 1, 2, 3, 12, 123\}$ be a code on 3 neurons. By Proposition 1, this code is connected. Figure 4 shows its realization by connected sets in $\mathbb{R}^2$, and the corresponding graph. We claim that the minimal embedding dimension of this code is 2. One could verify by computing the sensor graph of $C$. Alternatively, note that for the code to be realizable by connected sets, we must have $U_1 \cap U_2 \cap U_3 \neq \emptyset$ and $U_i$ can not be contained in $U_j$ for every $i, j \in [3]$, $i \neq j$. This is clearly not possible in dimension 1.

3. Open directions

In practice, neural firing is stochastic. One could incorporate noise to the receptive field code by replacing the deterministic equation (1) with some parametrization of the firing probability $P(i \in \sigma(x) | x \in U_i)$. To be well-defined, this model needs further specifications, such as the distribution of the signal on $\mathbb{R}^d$. In this formulation, the minimal embedding dimension is a difficult and poorly formed statistical problem. Furthermore, it is clear that the minimal embedding dimension depends heavily on such details. However, underlying such models the assumption that there is a set of true receptive fields $\mathcal{U}$. Knowing the minimal embedding dimension for the deterministic model ensures that the neuroscientists do not have excessively many parameters, which can lead to ill-defined estimation problems. From this view, Theorem 2 states that if the true receptive fields are only required to be connected, one can assume that they are in dimension 3.

Apart from connected and convex sets, there are many biologically relevant models for receptive fields. Finding the minimal embedding dimension of receptive field codes realizable by any given family is an interesting and challenging problem. To be concrete, we propose another simple family motivated by observations from neuroscience. In experiments, one often encounter a group of neurons which all have the same receptive field up to translation, such as the retinal ganglion cells, head direction cells [7], place cells and grid cells [19; 14]. This corresponds to the case where $\mathcal{F}_d$ consists of all possible translations of some set $S \subset \mathbb{R}^d$. We call this the shift code. Thus, a concrete open problem is: which shift codes can be realized, and what would be their minimal embedding dimensions?

Acknowledgments

The authors thank the anonymous referee for constructive comments.
References


RAFFAELLA MULAS: raffaella.mulas@gmail.com
MPI MiS-Leipzig, Germany

NGOC M. TRAN: ntran@math.utexas.edu
University of Texas at Austin, Austin, TX, United States
and
Hausdorff Center for Mathematics, Bonn, Germany

mathematical sciences publishers
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Editorial: A new beginning</td>
<td>1</td>
</tr>
<tr>
<td>T H O M A S  K A H L E  a n d  S O N J A  P E T R O V I Ć</td>
<td></td>
</tr>
<tr>
<td>Maximum likelihood estimation of toric Fano varieties</td>
<td>5</td>
</tr>
<tr>
<td>Estimating linear covariance models with numerical nonlinear algebra</td>
<td>31</td>
</tr>
<tr>
<td>Expected value of the one-dimensional earth mover’s distance</td>
<td>53</td>
</tr>
<tr>
<td>R E B E C C A  B O U R N  a n d  J E B  F .  W I L L E N B R I N G</td>
<td></td>
</tr>
<tr>
<td>Inferring properties of probability kernels from the pairs of variables they involve</td>
<td>79</td>
</tr>
<tr>
<td>L U I G I  B U R I G A N A  a n d  M I C H E L E  V I C O V A R O</td>
<td></td>
</tr>
<tr>
<td>Minimal embedding dimensions of connected neural codes</td>
<td>99</td>
</tr>
<tr>
<td>R A F F A E L L A  M U L A S  a n d  N G O C  M .  T R A N</td>
<td></td>
</tr>
</tbody>
</table>