involve
a journal of mathematics

# Sparse neural codes and convexity

R. Amzi Jeffs, Mohamed Omar, Natchanon Suaysom,
Aleina Wachtel and Nora Youngs

msp

2019 vol. 12, no. 5

INVOLVE 12:5 (2019)
dx.doi.org/10.2140/involve.2019.12.737

# Sparse neural codes and convexity

R. Amzi Jeffs, Mohamed Omar, Natchanon Suaysom,
Aleina Wachtel and Nora Youngs

(Communicated by Ann N. Trenk)

Determining how the brain stores information is one of the most pressing problems in neuroscience. In many instances, the collection of stimuli for a given neuron can be modeled by a convex set in $\mathbb{R}^d$. Combinatorial objects known as *neural codes* can then be used to extract features of the space covered by these convex regions. We apply results from convex geometry to determine which neural codes can be realized by arrangements of open convex sets. We restrict our attention primarily to sparse codes in low dimensions. We find that intersection-completeness characterizes realizable 2-sparse codes, and show that any realizable 2-sparse code has embedding dimension at most 3. Furthermore, we prove that in $\mathbb{R}^2$ and $\mathbb{R}^3$, realizations of 2-sparse codes using closed sets are equivalent to those with open sets, and this allows us to provide some preliminary results on distinguishing which 2-sparse codes have embedding dimension at most 2.

## 1. Introduction

One of the fundamental problems of convex geometry is understanding the intersection behavior of convex sets. Classical theorems in this area include Helly's theorem and its many variations, which show that the presence of lower-order intersections of convex sets in $\mathbb{R}^d$ can force intersections of higher order; see for example [Amenta et al. 2017; Danzer et al. 1963; Eckhoff 1993; Matoušek 2002]. Recent work [Tancer 2013] on the representability of simplicial complexes provides a sharp bound on the dimension in which intersection patterns of convex sets can be realized. We consider the problem of simultaneously realizing intersection patterns along with other relationships between convex sets, such as containment. This problem is motivated by one of the challenges of mathematical neuroscience: determining how the structure of a stimulus space is represented in the brain.

Many types of neurons respond to stimuli in an environment; the set of all such stimuli is called the *stimulus space* $X$. Usually, we consider $X \subset \mathbb{R}^d$. If we are

*MSC2010:* 05C62, 52A10, 92B20.
*Keywords:* neural code, sparse, convexity.

considering data from $n$ neurons $\{1, \ldots, n\}$ which respond to stimuli in $X$, the *receptive field* for neuron $i$ is the subset $U_i$ of the stimulus space $X$ for which neuron $i$ is highly responsive. Throughout this article, we assume the sets $U_i$ are convex. Indeed, experimental data on many types of neurons, such as place cells [O'Keefe and Dostrovsky 1971] or orientation-tuned neurons [Hubel and Wiesel 1959], make it evident that receptive fields are often well-approximated by convex sets. Hence, for such neurons, the regions of stimulus space in which multiple neurons fire can be modeled by intersections of convex sets, and thus the mathematical theory developed by Helly, Tancer, and others can inform us about the possible arrangements of receptive fields in a given dimension.

Helly's theorem, however, cannot inform us about all types of receptive field arrangements. For example, if $U_i$, $U_j$ are receptive fields which intersect, the neural data will differentiate between $U_i \subseteq U_j$ and $U_i \nsubseteq U_j$, but Helly's theorem merely notes that $U_i$ and $U_j$ intersect. We thus go beyond the usual scope of convex geometry to consider the problem of finding arrangements of convex sets which fully realize the information present in the neural data, including containments. This problem was posed originally in [Curto et al. 2013b], and has been an active area of exploration in recent years. Others such as [Chen et al. 2019; Curto et al. 2017; Cruz et al. 2019; Amzi Jeffs 2018; Amzi Jeffs and Novik 2018] have approached it using methods from algebra, combinatorics, and discrete geometry, but a full solution remains out of reach. In order to address this issue, we first describe how neural data is represented mathematically.

**Definition.** A *neural code* on $n$ neurons is a set of binary firing patterns $\mathcal{C} \subset \{0, 1\}^n$, representing neural activity. Elements of $\mathcal{C}$ are referred to as *codewords*.

The firing of a neuron is an all-or-nothing event, and so a codeword $c \in \mathcal{C}$ represents a data point in which a specific set of neurons are simultaneously firing, with neuron $i$ active if $c_i = 1$ and inactive if $c_i = 0$. For example, the codeword 0011 represents a data point at which neurons 3 and 4 were active, while neurons 1 and 2 were not. In the receptive field context, the presence of this codeword in $\mathcal{C}$ indicates that $(U_3 \cap U_4) \backslash (U_1 \cup U_2) \neq \varnothing$.

**Definition.** Let $\mathcal{U} = \{U_1, \ldots, U_n\}$ be a collection of sets in $\mathbb{R}^d$. The *associated neural code* $\mathcal{C}(\mathcal{U}) \subseteq \{0, 1\}^n$ is the set of firing patterns representing the regions in the arrangement

$$\mathcal{C}(\mathcal{U}) \stackrel{\text{def}}{=} \left\{ c \in \{0, 1\}^n \;\middle|\; \Big(\bigcap_{c_i=1} U_i\Big) \backslash \Big(\bigcup_{c_j=0} U_j\Big) \neq \varnothing \right\}.$$

Any collection of sets $\mathcal{U}$ in $\mathbb{R}^d$ gives rise to an associated neural code. However, as we have mentioned, the receptive fields $U_i$ are generally presumed to be convex. One of our main motivating examples is that of place cells, whose receptive fields

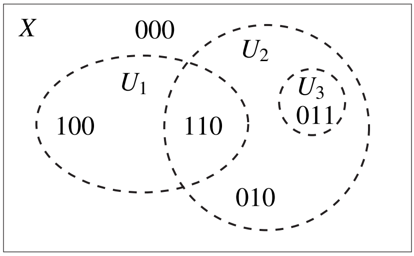


**Figure 1.** An open convex realization of the code $\mathcal{C} = \{000, 100, 010, 110, 011\}$ in $\mathbb{R}^2$, with each region labeled with its corresponding codeword. This shows that $\mathcal{C}$ is an open convex realizable code with $d(\mathcal{C}) \leq 2$. It can be shown that, in fact, $d(\mathcal{C}) = 1$.

are generally seen to be convex, as explained in [Curto et al. 2017]. We additionally assume the receptive fields $U_i$ are open, since by restricting to open sets, we force all sets in our realization to be full-dimensional; furthermore, their intersections, if nonempty, must also be full-dimensional. This allows us to avoid degenerate cases which would not be meaningful in a neural context. These assumptions are consistent with the literature [Curto et al. 2013b; 2017; Lienkaemper et al. 2017]. However, many of our proofs will require that we shift between closed and open convex sets that are associated to the same code. We therefore make the following definition:

**Definition.** If $\mathcal{U} = \{U_1, \dots, U_n\}$ is a collection of open (respectively, closed) convex sets in $\mathbb{R}^d$ for which $\mathcal{C} = \mathcal{C}(\mathcal{U})$, then we say that $\mathcal{C}$ is *open* (*closed*) *convex realizable in* $\mathbb{R}^d$, and that $\mathcal{U}$ is an *open* (*closed*) *convex realization* of $\mathcal{C}$.

Then, for any code $\mathcal{C}$, we define $d(\mathcal{C})$ to be the minimum dimension $d$ such that $\mathcal{C}$ has an open convex realization in $\mathbb{R}^d$, if such a dimension $d$ exists. Figure 1 shows an open convex realization in $\mathbb{R}^2$ for a code $\mathcal{C}$ which has minimum dimension $d(\mathcal{C}) = 1$. If $\mathcal{C}$ is not realizable with open convex sets in *any* dimension, we say $d(\mathcal{C}) = \infty$. Such codes do exist; see Figure 2.

**Definition.** The *support* of a vector $c \in \{0, 1\}^n$, denoted by $\operatorname{supp}(c)$, is the set of indices of value 1, or the set of all firing neurons:

$$\operatorname{supp}(c) \stackrel{\text{def}}{=} \{i \mid c_i = 1\}.$$

The *support of a code* $\mathcal{C} \subseteq \{0, 1\}^n$ is the set of the supports of its codewords:

$$\operatorname{supp}(\mathcal{C}) \stackrel{\text{def}}{=} \{\operatorname{supp}(c) \mid c \in \mathcal{C}\}.$$

We assume that there are instances when none of the neurons of interest are firing; hence, we will always assume that the codeword $00\cdots0$ is present in any code.

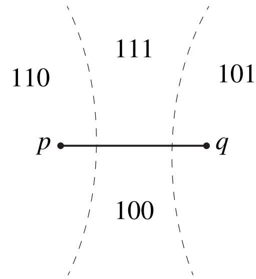


**Figure 2.** The code $\mathcal{C} = \{000, 010, 001, 110, 101\}$ is not open convex realizable in $\mathbb{R}^d$ for any $d < \infty$. If it were, we could pick points $p \in (U_1 \cap U_2) \setminus U_3$ and $q \in (U_1 \cap U_3) \setminus U_2$. The line segment $\overline{pq}$ is contained in $U_1$ by convexity; to move from $p$ to $q$ along $\overline{pq}$, we must leave $U_2$ and enter $U_3$. If we leave $U_2$ before entering $U_3$ that would indicate the presence of codeword 100, which is not in the code; if we enter $U_3$ before leaving $U_2$ that would indicate the codeword 111, which is not in the code. Since all sets are open, these are the only possibilities.

**Example.** Let $\mathcal{C} = \{000, 101, 110, 111\}$. Then $\text{supp}(101) = \{1, 3\}$, $\text{supp}(111) = \{1, 2, 3\}$, and $\text{supp}(\mathcal{C}) = \{\varnothing, \{1, 3\}, \{1, 2\}, \{1, 2, 3\}\}$.

Recent work, for example [Lin et al. 2014], shows the utility and importance of sparsity in neural codes. For practical reasons, our definition of "sparse" differs slightly from the usual low average weight definition often used in coding literature; see for example [Curto et al. 2013a]. We use instead a low maximum weight definition:

**Definition.** A code $\mathcal{C}$ is *$k$-sparse* if $|\text{supp}(c)| \le k$ for all $c \in \mathcal{C}$.

We begin the program of studying $k$-sparse codes by focusing on 2-sparse codes, where there is already rich mathematics to be found. Our fundamental motivating questions are the following:

**Question 1.1.** *Which* 2*-sparse codes are open convex realizable*?

**Question 1.2.** *If $\mathcal{C}$ is an open convex realizable* 2*-sparse code*, *what is its minimum embedding dimension $d(\mathcal{C})$*?

Our main result is the following characterization of which 2-sparse codes have open convex realizations, including a dimensional bound.

**Theorem 1.3.** *A* 2*-sparse code $\mathcal{C}$ has an open convex realization if and only if* $\text{supp}(\mathcal{C})$ *is intersection-complete. Furthermore*, *if $\mathcal{C}$ is realizable then $d(\mathcal{C}) \le 3$.*

This answers our first question in its entirety, and partially answers the second. Note that in this result there is no room for generality in terms of sparsity; there are

3-sparse codes that are realizable but not intersection-complete; see for example the code $\mathcal{C} = \{0, 1\}^3 \backslash \{001\}$ in [Curto et al. 2013b]. In Section 2, we will prove Theorem 1.3 using several lemmas. In particular we show in Lemma 2.6 that for such codes it is equivalent to find a closed convex realization, as it may be transformed to an open convex realization in $\mathbb{R}^2$ or $\mathbb{R}^3$. It immediately follows from this and [Tancer 2013] that any 2-sparse code has a convex open realization in $\mathbb{R}^3$. In Section 3, we consider the second question in more detail, and exhibit a class of 2-sparse codes with $d \leq 2$, as well as a class with $d = 3$.

## 2. Realizability of 2-sparse codes

This section is dedicated to proving Theorem 1.3, which establishes that a 2-sparse code is realizable precisely when its support is intersection-complete and, for such codes $\mathcal{C}$, $d(\mathcal{C}) \leq 3$. In order to prove this theorem, we make use of the simplicial complex of a code, which is introduced below.

**Definition.** A *simplicial complex* on a finite set $S$ is a family $\Delta$ of subsets of $S$ such that if $X \in \Delta$ and $Y \subseteq X$, then $Y \in \Delta$.

In this paper, the set $S$ under consideration will most often be $[n] = \{1, \ldots, n\}$. In a situation where $S = \{v_1, \ldots, v_n\}$, we will typically refer to any set in a simplicial complex on $S$ by its set of indices.

**Definition.** The *simplicial complex of a code* $\mathcal{C}$ is the smallest simplicial complex containing $\operatorname{supp}(\mathcal{C})$; this is denoted by $\Delta(\mathcal{C})$. The *k-skeleton* of a simplicial complex $\Delta$ is the simplicial complex $\Delta_k$ given by the collection of sets in $\Delta$ of size at most $k + 1$; see Figure 3.

If $\mathcal{C}$ is 2-sparse, then $\Delta(\mathcal{C})$ consists only of 0-, 1-, and 2-element sets. We can therefore think of $\Delta(\mathcal{C})$ as a graph, with 1-element sets corresponding to vertices and 2-element sets as edges between them. Note that since $\Delta(\mathcal{C})$ is a simplicial complex, if $\{i, j\} \in \Delta(\mathcal{C})$, then both $\{i\}$ and $\{j\}$ must be in $\Delta(\mathcal{C})$ as well; hence this association is well-defined. The formal relationship between 2-sparse codes and graphs is captured by the following definition.

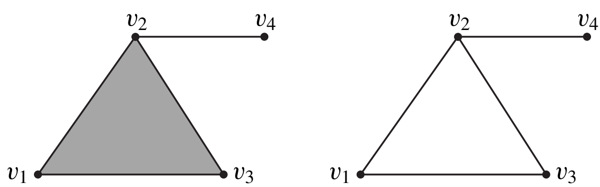


**Figure 3.** At left, a geometric representation of simplicial complex on $S = \{v_1, v_2, v_3, v_4\}$ with $\Delta = \{\varnothing, \{1\}, \{2\}, \{3\}, \{4\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{2, 4\}, \{1, 2, 3\}\}$. At right, a geometric representation of the 1-skeleton of $\Delta$.

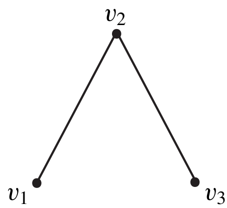


**Figure 4.** The graph $G_{\mathcal{C}}$ for $\mathcal{C} = \{000, 100, 010, 110, 011\}$; see Figure 1 for a realization of $\mathcal{C}$.

**Definition.** Let $\mathcal{C} \subset \{0, 1\}^n$ be a neural code. The *graph of* $\mathcal{C}$, denoted by $G_{\mathcal{C}}$, is the graph whose vertex set is $[n]$, with $i$ adjacent to $j$ if $\{i, j\} \subseteq \operatorname{supp}(c)$ for some $c \in \mathcal{C}$; see Figure 4.

Note that $G_{\mathcal{C}}$ is the 1-skeleton of $\Delta(\mathcal{C})$. In particular, for a 2-sparse code, $\Delta(\mathcal{C})$ and $G_{\mathcal{C}}$ contain exactly the same information because $\Delta(\mathcal{C})$ is equal to its 1-skeleton.

As we saw in Figure 2, there exist 2-sparse codes that are not convex in any dimension. The following lemma generalizes the obstruction presented in that figure.

**Lemma 2.1.** *Let $\mathcal{C}$ be a 2-sparse code. If $\mathcal{C}$ has a convex open realization in any dimension, then* $\operatorname{supp}(\mathcal{C})$ *is intersection-complete.*

*Proof.* Suppose $\mathcal{C}$ is a 2-sparse code with open convex realization $\mathcal{U} = \{U_1, \dots, U_n\}$. Since $\mathcal{C}$ is 2-sparse, $|\operatorname{supp}(c)| \in \{0, 1, 2\}$ for every $c \in \mathcal{C}$. If $|\operatorname{supp}(c)|$ is at most 1, then $\operatorname{supp}(c) \cap \operatorname{supp}(c') \in \operatorname{supp}(\mathcal{C})$ for any $c' \in \mathcal{C}$, because the intersection is either $\varnothing$ or $\operatorname{supp}(c)$. It then remains to show that $\operatorname{supp}(c) \cap \operatorname{supp}(c') = \{i\} \in \operatorname{supp}(\mathcal{C})$ when $\operatorname{supp}(c) = \{i, j\}$ and $\operatorname{supp}(c') = \{i, k\}$ with $j \neq k$. In this case, $U_i \cap U_j$ and $U_i \cap U_k$ are nonempty so there exist points $p \in U_i \cap U_j$ and $q \in U_i \cap U_k$. Consider the line segment $\overline{pq}$ connecting $p$ and $q$. Since $U_i$ is convex, $\overline{pq}$ is contained in $U_i$. For each $m \in [n] \setminus \{i\}$, consider the set $L_m = \overline{pq} \cap U_i \cap U_m$; note that any two such sets are disjoint, and that $L_j$ and $L_k$ are nonempty. If the sets $\{L_m\}$ partition the line $\overline{pq}$, then this would disconnect $\overline{pq}$ in the subspace topology, but as $\overline{pq}$ is connected, this is impossible. Thus, there must be some point on $\overline{pq}$ which is contained in $U_i$ only. The existence of this point implies $\{i\} \in \operatorname{supp}(\mathcal{C})$ as desired. □

The conclusion of the previous lemma is that it is necessary for open convex realizable 2-sparse codes to be intersection complete. In fact, this property characterizes 2-sparse codes with an open convex realization; this is the content of Theorem 1.3. To prove Theorem 1.3, we will use a method of repeatedly making geometric augmentations to existing realizations; in order to make such augmentations without changing the underlying code, we must ensure that subset containment relations between sets are maintained. In the 2-sparse case, the following definition encapsulates the key relationships that must be maintained:

**Definition.** Let $\mathcal{U} = \{U_1, \dots, U_n\}$ be a collection of sets in $\mathbb{R}^d$. For any ordered pair $(U_i, U_j)$ we distinguish three possible relations between $U_i$ and $U_j$:

<u>Type A</u> (disjointness): $U_i \cap U_j = \varnothing$; i.e., $\{i, j\} \not\subseteq \operatorname{supp}(c)$ for any $c \in \mathcal{C}$.

<u>Type B</u> (containment): $U_j \subseteq U_i$; i.e., there exists a codeword $c \in \mathcal{C}(\mathcal{U})$ so that $\{i, j\} \subseteq \operatorname{supp}(c)$ and any codeword whose support contains $j$ must also have $i$ in its support.

<u>Type C</u> (proper intersection): $U_i \cap U_j$ is nonempty and $U_j \setminus U_i$ is nonempty; i.e., there exist codewords $c_1, c_2 \in \mathcal{C}(\mathcal{U})$ so that $\{i, j\} \subseteq \operatorname{supp}(c_1)$, $j \in \operatorname{supp}(c_2)$ and $i \notin \operatorname{supp}(c_2)$.

The type-A, type-B and type-C set relationships effectively characterize the structure of a 2-sparse code; indeed, 2-sparse codes are completely determined by the pairwise relationships of the sets in any realization. We explicitly state this in the following proposition.

**Proposition 2.2.** *Let $\mathcal{U}$ and $\mathcal{U}'$ be collections of sets in $\mathbb{R}^d$ so that $\mathcal{C}(\mathcal{U})$ and $\mathcal{C}(\mathcal{U}')$ are both 2-sparse. Then $\mathcal{C}(\mathcal{U}) = \mathcal{C}(\mathcal{U}')$ if and only if for every ordered pair $(i, j)$ the relation between $U_i$ and $U_j$ is the same as the relation between $U_i'$ and $U_j'$.*

We now introduce the geometric underpinnings of the augmentations we will apply to realizations of codes. In these definitions, we make use of the idea of an $\varepsilon$-ball around a point $p$ ($B_\varepsilon(p) = \{x \in \mathbb{R}^d \mid \|x - p\| < \varepsilon\}$), the interior of a set $A$ ($\operatorname{int}(A) = \{x \in A \mid B_\varepsilon(x) \subseteq A$ for some $\varepsilon > 0\}$), and the closure of a set ($\bar{A} = \{x \in \mathbb{R}^d \mid x$ is a limit point of $A\}$).

**Definition.** Given $\varepsilon > 0$ and $A \subset \mathbb{R}^d$, the *trim* of $A$ by $\varepsilon$ is the set

$$\operatorname{trim}(A, \varepsilon) \stackrel{\text{def}}{=} \operatorname{int}\{p \in \mathbb{R}^d \mid B_\varepsilon(p) \subseteq A\}.$$

The *inflation* of $A$ by $\varepsilon$ is the set

$$\operatorname{inflate}(A, \varepsilon) \stackrel{\text{def}}{=} \{a + x \mid a \in A,\ x \in \mathbb{R}^d \text{ with } \|x\| < \varepsilon\}.$$

If $\mathcal{A} = \{A_1, \ldots, A_n\}$ is a collection of sets, then

$$\operatorname{trim}(\mathcal{A}, \varepsilon) = \{\operatorname{trim}(A_1, \varepsilon), \ldots, \operatorname{trim}(A_n, \varepsilon)\},$$
$$\operatorname{inflate}(\mathcal{A}, \varepsilon) = \{\operatorname{inflate}(A_1, \varepsilon), \ldots, \operatorname{inflate}(A_n, \varepsilon)\}.$$

**Proposition 2.3.** *For any convex set $A \subset \mathbb{R}^d$ and $\varepsilon > 0$, the following statements hold*:

(1) $\operatorname{trim}(A, \varepsilon)$ *is an open convex set.*

(2) $\overline{\operatorname{trim}(A, \varepsilon)}$ *is contained in the interior of $A$.*

(3) $\operatorname{inflate}(A, \varepsilon)$ *is an open convex set.*

*Proof.* For (1), we need only prove convexity, and we may assume $\operatorname{trim}(A, \varepsilon)$ is nonempty. Let $p$ and $q$ be points in $\operatorname{trim}(A, \varepsilon)$; then $B_\varepsilon(p)$ and $B_\varepsilon(q)$ are contained in $A$, and hence so is the convex hull of their union. This convex hull contains the line segment $\overline{pq}$. For (2), note that $\overline{\operatorname{trim}(A, \varepsilon)} \subseteq \operatorname{trim}(A, \varepsilon/2) \subseteq \operatorname{int}(A)$. Finally, (3) follows from the fact that $A$ is convex and $\{x \in \mathbb{R}^d \mid \|x\| < \varepsilon\}$ is open and convex. □

We now show that open convex realizations of 2-sparse codes can be trimmed down to give another open convex realization.

**Lemma 2.4.** *Given a 2-sparse code $\mathcal{C}$ with an open convex realization $\mathcal{U} = \{U_1, \dots, U_n\}$, there exists some $\varepsilon > 0$ so that* $\operatorname{trim}(\mathcal{U}, \varepsilon)$ *is also a realization of $\mathcal{C}$.*

*Proof.* Our method is as follows: For each set $U_i$, we find an $\varepsilon_i$ such that $\operatorname{trim}(U_i, \varepsilon_i) \neq \varnothing$, and for each pair $\{i, j\}$ we find an $\varepsilon_{ij}$ such that $\operatorname{trim}(\{U_i, U_j\}, \varepsilon_{ij})$ preserves their relationship type (type A, type B or type C). We then let $\varepsilon$ be the minimum of all $\varepsilon_i$ and $\varepsilon_{ij}$, and show that $\operatorname{trim}(\mathcal{U}, \varepsilon)$ is a realization of the original code $\mathcal{C}$.

To start, for each $i$ with $U_i$ nonempty, there must be some point $p$ and $\delta_i > 0$ with $B_{\delta_i}(p) \subseteq U_i$. Let $\varepsilon_i = \delta_i/2$. Let $\varepsilon_1 = \min_{i \in [n]} \varepsilon_i$. Now, for each pair $\{i, j\}$, we choose $\varepsilon_{ij}$ depending on the relationship type between $U_i$ and $U_j$:

<u>Type A</u>: If $U_i \cap U_j = \varnothing$, set $\varepsilon_{ij} = \min\{\varepsilon_i, \varepsilon_j\}$.

<u>Type B</u>: If $U_i = U_j$, set $\varepsilon_{ij} = \min\{\varepsilon_i, \varepsilon_j\}$. If $U_i \subsetneq U_j$, note that $U_j \backslash U_i$ has nonempty interior. Thus there exists some point $p$ and some $\delta_{ij} > 0$ with $B_{\delta_{ij}}(p) \subseteq U_j \backslash U_i$. Let $\varepsilon_{ij} = \min\{\delta_{ij}/2, \varepsilon_i\}$.

<u>Type C</u>: If $U_i \cap U_j \neq \varnothing$, but neither $U_i \subseteq U_j$ nor $U_j \subseteq U_i$ is true, note that $U_i \cap U_j$ is open and therefore there exist a point $p$ and $\varepsilon' > 0$ with $B_{\varepsilon'}(p) \subseteq U_i \cap U_j$. There exist also points $p_i$, $p_j$ in $U_i \backslash U_j$, $U_j \backslash U_i$ respectively, with corresponding $\hat{\varepsilon}$ and $\tilde{\varepsilon}$ such that $B_{\hat{\varepsilon}}(p_i) \subseteq U_i \backslash U_j$ and $B_{\tilde{\varepsilon}}(p_j) \subseteq U_j \backslash U_i$. Pick $\varepsilon_{ij} = \min\{\varepsilon_i, \varepsilon_j, \hat{\varepsilon}/2, \tilde{\varepsilon}/2, \varepsilon'/2\}$.

Let $\varepsilon_2 = \min_{i,j} \varepsilon_{ij}$, and finally, let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Since $\operatorname{trim}(U, \varepsilon) \subset U$, and originally there were no triple intersections, by construction it is impossible for $\operatorname{trim}(\mathcal{U}, \varepsilon)$ to have triple intersections. Thus, $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$ is still 2-sparse. We now show that $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon)) = \mathcal{C}$.

If the codeword with support $\{i, j\}$ is in $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$, then

$$\operatorname{trim}(U_i, \varepsilon) \cap \operatorname{trim}(U_j, \varepsilon) \neq \varnothing.$$

As $\operatorname{trim}(U, \varepsilon) \subset U$, this implies $U_i \cap U_j \neq \varnothing$. Since $\mathcal{C}$ is 2-sparse, the codeword with support $\{i, j\}$ is in $\mathcal{C}$. On the other hand, if the codeword with support $\{i, j\}$ is in $\mathcal{C}$, then $U_i \cap U_j \neq \varnothing$, and so we are in a case of type A, B or C above. By our choice of $\varepsilon$, we ensure that in each case $\operatorname{trim}(U_i, \varepsilon) \cap \operatorname{trim}(U_j, \varepsilon) \neq \varnothing$, and hence (as the code is 2-sparse) the codeword with support $\{i, j\}$ is in $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$.

If a codeword with support $\{i\}$ is in $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$, then

$$\operatorname{trim}(U_i, \varepsilon) \backslash \bigcup_{j \in [n], j \neq i} \operatorname{trim}(U_j, \varepsilon) \neq \varnothing.$$

We then know that $U_i \setminus \bigcup_{j\in[n], j\neq i} U_j \neq \varnothing$. If it were not, then we would have $U_i \subseteq \bigcup_{j\in I} U_j$ for some index set $I$. However, this is impossible: If $|I| = 1$, then $U_i \subseteq U_j$, but then $\operatorname{trim}(U_i, \varepsilon) \subseteq \operatorname{trim}(U_j, \varepsilon)$. If $|I| > 1$, then $U_i \subseteq \bigcup_{j\in I} U_j$. But then the 2-sparsity of $\mathcal{C}$ means we would see the codewords $\{i, j\}$ and $\{i, k\}$ in $\mathcal{C}$ for $j, k \in I$ but not their intersection $\{i\}$, contradicting Lemma 2.1. Hence, the codeword with support $\{i\}$ is in $\mathcal{C}$.

Now, suppose a codeword with support $\{i\}$ is in $\mathcal{C}$, and let $J = \{j \mid U_i \cap U_j \neq \varnothing\}$. If $|J| \leq 1$ then we are in a case of type A, B, or C above, and by our choice of $\varepsilon$ we know there is a codeword with support $\{i\}$ in $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$. If $|J| \geq 2$, let $j, k \in J$. Then by our choice of $\varepsilon$, we know $\operatorname{trim}(U_i, \varepsilon) \cap \operatorname{trim}(U_j, \varepsilon) \neq \varnothing$ and $\operatorname{trim}(U_i, \varepsilon) \cap \operatorname{trim}(U_k, \varepsilon) \neq \varnothing$, and hence the codewords with supports $\{i, j\}$ and $\{i, k\}$ are in $\operatorname{trim}(\mathcal{U}, \varepsilon)$. By Lemma 2.1, we know the codeword with support $\{i\}$ is also in $\mathcal{C}(\operatorname{trim}(\mathcal{U}, \varepsilon))$. □

Next, we show that a closed convex realization of a 2-sparse codes can be inflated to create an open convex realization.

**Lemma 2.5.** *Let $\mathcal{C}$ be a 2-sparse code with a closed convex realization $\mathcal{V} = \{V_1, \dots, V_n\}$ in which every set is bounded. Then there exists some $\varepsilon > 0$ such that* $\operatorname{inflate}(\mathcal{V}, \varepsilon)$ *is an open convex realization of $\mathcal{C}$.*

*Proof.* Consider the partial ordering on $\mathcal{V}$ given by set inclusion. We will use this ordering to inflate the sets in $\mathcal{V}$ iteratively (possibly by different $\varepsilon$ factors) and then argue that we can obtain a uniform $\varepsilon$ for which $\operatorname{inflate}(\mathcal{V}, \varepsilon)$ is an open convex realization of $\mathcal{C}$. In this iterative process, if $V_i = V_j$ for any $i \neq j$, we apply the process simultaneously to $V_i$ and $V_j$. As such, it is sufficient for our proof to assume $V_i \neq V_j$ for any $i \neq j$.

To start, begin with a fixed index $i$ for which $V_i$ is maximal in $\mathcal{V}$ with respect to inclusion. All sets in $\mathcal{V}$ are closed and bounded, so for any $j$ with $V_i \cap V_j = \varnothing$, $V_i$ has positive distance $d_{i,j}$ to $V_j$. Let $\delta_i = \min_{V_i \cap V_j = \varnothing} d_{i,j}$. Now if there are $j, k \neq i$ with $V_j \cap V_k \neq \varnothing$, then $V_i$ has positive distance $d_{i,j,k}$ to $V_j \cap V_k$; take $\delta_i'$ to be the minimum of all such $d_{i,j,k}$. Furthermore, let $\delta_i'' > 0$ be such that for all $j$ with $V_j \not\subseteq V_i$, we have $V_j \not\subseteq \operatorname{inflate}(V_i, \delta_i'')$. Finally, choose $\varepsilon_i < \min\left\{\frac{1}{2}\delta_i, \frac{1}{2}\delta_i', \frac{1}{2}\delta_i''\right\}$. These choices help guarantee that no new pairwise or triple intersections are created, and no new containments are created.

If we replace $V_i$ by $\overline{\operatorname{inflate}(V_i, \varepsilon_i)}$, then the code is still 2-sparse, and the three subset relationship types for the ordered pairs $(V_i, V_j)$ where $j \neq i$ are maintained:

Type A: Disjointness is preserved since $\varepsilon_i$ is at most half the distance from $V_i$ to any set disjoint from it.

Type B: Containment is preserved since we are only making $V_i$ bigger.

Type C: Proper intersection is preserved by our choice of $\varepsilon_i$.

By a similar argument, the subset relationship of the ordered pair $(V_j, V_i)$ for any $j \neq i$ is also preserved after replacing $V_i$ by $\overline{\text{inflate}(V_i, \varepsilon_i)}$. Thus replacing $V_i$ by $\overline{\text{inflate}(V_i, \varepsilon_i)}$ yields a new realization of $\mathcal{C}$.

For any subsequent step in our iterative process, choose a set $V_i \in \mathcal{V}$ for which every member of the set $\{V_j \in \mathcal{V} \mid V_j \supset V_i\}$ has already been inflated. Choose $\varepsilon_i$ in the same way as previously described with the additional caveat that if $V_i \subseteq V_j$ then $\varepsilon_i < \varepsilon_j$. A similar argument shows that replacing $V_i$ by $\overline{\text{inflate}(V_i, \varepsilon_i)}$ yields a new realization of $\mathcal{C}$. Once we have inflated every set in the realization we can let $\varepsilon = \min_{i \in [n]} \varepsilon_i$ and observe that $\text{inflate}\big(\mathcal{U}, \frac{1}{2}\varepsilon\big)$ is an open convex realization of $\mathcal{C}$. □

This result allows us to prove the useful fact that for 2-sparse codes, open and closed convex realizations exist interchangeably, and we can build either type of realization from the other.

**Lemma 2.6.** *Let $\mathcal{C}$ be a 2-sparse code. Then $\mathcal{C}$ has an open convex realization in $\mathbb{R}^d$ if and only if $\mathcal{C}$ has a closed convex realization in $\mathbb{R}^d$.*

*Proof.* First, let $\mathcal{U}$ be an open convex realization of $\mathcal{C}$. Applying Lemma 2.4, there is an $\varepsilon > 0$ such that $\mathcal{U}' = \text{trim}(\mathcal{U}, \varepsilon)$ is an open realization of $\mathcal{C}$. Since the closure of each $U_i'$ is contained in $U_i$ (by Proposition 2.3), $\mathcal{U}'$ is an open convex realization of $\mathcal{C}$ in which two sets intersect if and only if their closures do. Let $\mathcal{V} = \{\overline{U}_1', \dots, \overline{U}_n'\}$. No triple intersections exist in $\mathcal{V}$ since these would correspond to triple intersections in $\mathcal{U}$. Thus by Proposition 2.2 it suffices to show that $\mathcal{V}$ preserves the relations between sets in $\mathcal{U}'$. Disjointness is preserved since sets in $\mathcal{U}'$ intersect if and only if their closures do. Containment is preserved under taking closures. Lastly, proper intersection is preserved, since if $U_i \setminus U_j$ is nonempty then there are limit points of $U_i$ that are not limit points of $U_j$.

For the reverse direction, let $\mathcal{V}$ be a closed convex realization of $\mathcal{C}$. For every nonempty intersection $V_i \cap V_j$, let $p_{i,j}$ be a point in this intersection. Furthermore, if some set $V_i$ is not contained in any other $V_j$, let $p_i \in V_i \setminus \bigcup_{j \neq i} V_j$. Then set $V$ to be the convex hull of all these $p_i's$ and $p_{i,j}'s$. Replacing each $V_i$ by $V_i \cap V$ yields a realization of $\mathcal{C}$ in which every set is closed, convex, and bounded. Applying Lemma 2.5, we obtain an open convex realization of $\mathcal{C}$ in $\mathbb{R}^d$. □

Although it may not be immediately clear from the proof, the condition that $\mathcal{C}$ is 2-sparse is necessary for Lemma 2.6 to hold. The 2-sparse condition is in fact best possible, since there exist 3-sparse codes which have closed convex realizations in $\mathbb{R}^2$, but for which open convex realizations exist only in $\mathbb{R}^3$ or higher. One such example is the code

$$\mathcal{C} = \{0000, 1000, 0100, 0010, 0001, 1110, 1001, 0101, 0011\}.$$

Figure 5 shows a closed realization of this code in $\mathbb{R}^2$, but it has no open realization in $\mathbb{R}^2$; see [Curto et al. 2017] for more details.

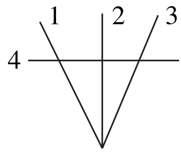


**Figure 5.** A closed realization of a code in $\mathbb{R}^2$ that has no open realization in $\mathbb{R}^2$.

Even more strikingly, there exist codes with a closed convex realization in $\mathbb{R}^2$ that have no open convex realization in *any* dimension; see [Lienkaemper et al. 2017] for an example of such a code on five neurons. This emphasizes how special realizations of 2-sparse codes are.

We can now use the previous lemmas to relate the convexity of a 2-sparse code $\mathcal{C}$ to the convexity of its associated simplicial complex $\Delta(\mathcal{C})$. We first need a technical lemma.

**Lemma 2.7.** *Let $\mathcal{U}$ be an open convex realization of a 2-sparse code $\mathcal{C}$. Then if $U_j \not\subseteq U_k$ for any $k \neq j$, there is a point $p \in \partial U_j \setminus \bigcup_{k \neq j} U_k$.*

*Proof.* Recall that for any set $U \subset \mathbb{R}^d$, $\partial U$ is the boundary of $U$. Consider the sets $\{\partial U_j \cap U_k\}_{k \neq j}$. These sets are disjoint: if not, then there exists $p \in (\partial U_j \cap U_k) \cap (\partial U_j \cap U_\ell)$. As $p \in U_k \cap U_\ell$, there exists $\varepsilon > 0$ with $B_\varepsilon(p) \subseteq U_k \cap U_\ell$. But then $B_\varepsilon(p) \cap U_j \neq \varnothing$, as $p \in \partial U_j$, so $U_j \cap U_k \cap U_\ell \neq \varnothing$ contradicting that $\mathcal{C}$ is 2-sparse.

Now, note that the disjoint sets $\{\partial U_j \cap U_k\}_{k \neq j}$ are open in the subspace topology with respect to $\partial U_j$, and hence they cannot partition $\partial U_j$ since $\partial U_j$ is connected. Thus, there exists $p \in \partial U_j \setminus \bigcup_{k \neq j} U_k$. □

**Lemma 2.8.** *Let $\mathcal{C}$ be a 2-sparse code and let $d \geq 2$. Then $\mathcal{C}$ has an open convex realization in $\mathbb{R}^d$ if and only if* $\operatorname{supp}(\mathcal{C})$ *is intersection-complete and $\Delta(\mathcal{C})$ has an open convex realization in $\mathbb{R}^d$.*

*Proof.* For the forward direction, we know from Lemma 2.1 that if $\mathcal{C}$ has a realization then $\operatorname{supp}(\mathcal{C})$ is intersection-complete. We will show that given a realization $\mathcal{U}$ of $\mathcal{C}$, we can construct a realization of $G_{\mathcal{C}}$. Since $\mathcal{C}$ is 2-sparse, we know $\mathcal{C}$ and $\Delta(\mathcal{C})$ must already contain the same 2-element sets, so we will show that we can adjust the realization of $\mathcal{C}$ to obtain any singletons $\{i\}$ which appear in $\Delta(\mathcal{C})$ but not in $\mathcal{C}$.

Let $\{i\} \in \Delta(\mathcal{C}) \setminus \operatorname{supp}(\mathcal{C})$. If there exist $j, k$ such that $\{i, j\}$ and $\{i, k\}$ are both in $\operatorname{supp}(\mathcal{C})$, then as $\operatorname{supp}(\mathcal{C})$ is intersection-complete, we know $\{i\} \in \operatorname{supp}(\mathcal{C})$. Thus, there must be exactly one $j$ such that $\{i, j\} \in \operatorname{supp}(\mathcal{C})$. Note immediately that in the realization $\mathcal{U}$ we have $U_i \subseteq U_j$ since $\{i, j\}$ is the only set in the support where $i$ appears. It suffices to transform $\mathcal{U}$ so that $U_i$ and $U_j$ intersect, but $U_i$ also contains points not in any other set in the realization.

If we have $U_j \subseteq U_i$, then $U_i = U_j$ so $U_j \cap U_k = \varnothing$ for any other $k$, and we can replace $U_j$ with an open ball properly contained in $U_i$ to obtain the desired result.

Otherwise, $U_j$ may intersect many other sets in the realization, but cannot be contained in them, since this would imply a triple intersection between the containing set $U_k$, $U_j$, and $U_i$. Apply Lemma 2.4 to obtain $\varepsilon > 0$ for which $\mathcal{U}' = \operatorname{trim}(\mathcal{U}, \varepsilon)$ is an open realization of $\mathcal{C}$. Define the sets $V_k = \partial U'_j \cap \overline{U}'_k$; note that each $V_k$ is closed. Furthermore, these sets are disjoint, since if $p \in V_k \cap V_\ell$, then $p \in U_j \cap U_k \cap U_\ell$ in the original realization which is impossible for a 2-sparse code. Since $\partial U'_j$ is connected and the $V_k$ are disjoint closed sets, $\bigcup_{k \neq j} V_k \subsetneq \partial U'_j$; let $p \in \partial U'_j \setminus \bigcup_{k \neq j} V_k$. Then $p$ has positive distance to all sets $U'_k$ with $k \neq j$ so there is some $\varepsilon' > 0$ with $B_{\varepsilon'}(p) \cap U'_k = \varnothing$ for all $k \neq j$. Replacing $U'_i$ with $B_{\varepsilon'}(p)$ will create a realization of a code $\mathcal{C}'$ with $\operatorname{supp}(\mathcal{C}') = \operatorname{supp} C \cup \{i\}$. Repeating this step as many times as necessary, we obtain a realization of $\Delta(\mathcal{C})$.

For the reverse, suppose $\mathcal{U}$ is an open convex realization of $\Delta(\mathcal{C})$. Note that if $\{i, j\} \in \operatorname{supp}(\Delta(\mathcal{C}))$, it is also in $\operatorname{supp}(\mathcal{C})$ since $\mathcal{C}$ is 2-sparse. Now, suppose $\{i\} \in \operatorname{supp}(\Delta(\mathcal{C})) \setminus \operatorname{supp}(\mathcal{C})$. Then there is at most one $j \neq i$ such that $\{i, j\} \in \operatorname{supp}(\mathcal{C})$ as $\mathcal{C}$ is intersection-complete. If there is such a $j$, replace $U_i$ with $U_i \cap U_j$ which is an open convex set; if there is no such $j$, then remove $U_i$ entirely. This gives a convex realization of $\Delta(\mathcal{C}) \setminus \{i\}$, and we can repeat this operation as many times as necessary to obtain a realization of $\mathcal{C}$. □

The above lemma can be summarized as follows: realizing a 2-sparse code and realizing its simplicial complex are equivalent, as long as $\operatorname{supp}(\mathcal{C})$ is intersection-complete. This equivalence is our main tool in proving Theorem 1.3 and obtaining a complete classification of which 2-sparse codes are convex in $\mathbb{R}^3$.

*Proof of Theorem 1.3.* The fact that any open convex realizable 2-sparse code must have $\operatorname{supp}(\mathcal{C})$ that is intersection-complete follows directly from Lemma 2.1. For the reverse direction, since $\operatorname{supp}(\mathcal{C})$ is intersection-complete, we know by Lemma 2.8 that it is sufficient to find an open convex realization for $\Delta(\mathcal{C})$. As $\mathcal{C}$ is 2-sparse, Lemma 2.6 tells us that it suffices to find a closed convex realization for $\Delta(\mathcal{C})$. Since $\Delta(\mathcal{C})$ is a 1-dimensional simplicial complex, a construction of [Tancer 2013] (see the proof of Theorem 3.1 therein) leads to a closed convex realization of a 1-dimensional simplicial complex in $\mathbb{R}^3$. This proves the desired result. □

Theorem 1.3 makes it very straightforward to check whether a 2-sparse code has an open convex realization in $\mathbb{R}^3$. The challenge that lies ahead is determining the minimal embedding dimension for a given 2-sparse code. We begin investigating this problem in the next section.

## 3. Dimension of 2-sparse codes

We noted early on that for 2-sparse codes, the simplicial complex $\Delta(\mathcal{C})$ and the graph $G_{\mathcal{C}}$ of the pairwise intersections of the code capture the same information. In this section, we will make heavy use of this correspondence, and construct realizations of various 2-sparse codes using graph-theoretic methods. Hence, throughout this discussion we will often refer to "realizations" of a graph $G_{\mathcal{C}}$. It is important to note that while a graph is the intersection graph of its realization, finding convex sets whose intersection graph is the graph of concern is not sufficient here. In particular, if a collection of convex sets has a triple with nonempty intersection then it is not, for our purposes, a realization of any graph, since graphs only encode intersections of order 2.

Our main result, Theorem 1.3, shows that any intersection-complete 2-sparse code can be realized in dimension $d \leq 3$. In this section, we begin the program of classifying 2-sparse codes based on minimal embedding dimension. We focus on distinguishing codes of dimension $d = 3$ from codes of dimension $d \leq 2$; note that the general problem of distinguishing 1-dimensional codes has been solved [Rosen and Zhang 2017]. Recall from Lemma 2.8 that realizing a 2-sparse code $\mathcal{C}$ is equivalent to realizing its simplicial complex $\Delta(\mathcal{C})$ (and therefore, its graph $G_{\mathcal{C}}$), so throughout this section we refer to realizing $G_{\mathcal{C}}$ rather than $\mathcal{C}$ itself. Our main contribution is that while the dimension of certain graphs can be bounded, we find that the traditional 2-dimensional graph-theoretic distinction (planarity) is not necessary for $G_{\mathcal{C}}$ to represent a 2-dimensional code. In particular, in Proposition 3.1, we observe $d(\mathcal{C}) \leq 2$ if $G_{\mathcal{C}}$ is planar, and in Proposition 3.2 if $G_{\mathcal{C}}$ is not planar, one can construct a *related* graph whose code has minimal embedding dimension 3. However, planarity does not strictly govern minimal embedding dimension, as any complete or complete bipartite graphs are realizable in $\mathbb{R}^2$.

The following proposition describes some common graphs which do have 2-dimensional convex realizations, including planar graphs.

**Proposition 3.1.** *The following graphs have an open convex realization in* $\mathbb{R}^2$:

(1) *planar graphs*,

(2) *the complete k-partite graph* $K_{n_1,n_2,\dots,n_k}$ *with part sizes* $n_1, n_2, \dots, n_k$,

(3) *any graph G with vertex set* $\{v_1, v_2, \dots, v_n, u_1, \dots, u_k\}$ *where the induced subgraph on the vertices* $v_1, v_2, \dots, v_n$ *is complete and* $\{v_1, v_2, \dots, v_n\} \supseteq N_G(u_k) \supseteq N_G(u_{k-1}) \supseteq \cdots \supseteq N_G(u_1)$.

*Proof.* In all cases, we find a closed convex realization of the given graph $G$, which by Lemma 2.6 implies the existence of an open convex realization. For (1), we first recall the circle packing theorem, which says that for any planar graph $G$ with vertex set $\{v_1, \dots, v_n\}$, there exist disjoint disks $C_1, C_2, \dots, C_n$ in $\mathbb{R}^2$ such that $C_i$ is tangent to $C_j$ if and only if $v_i$ is adjacent to $v_j$, and $C_i \cap C_j = \varnothing$ otherwise. See Figure 6 for an illustration of how these disks are constructed.

For (2), we first find a realization for the complete graph $K_n = K_{1,1,\dots,1}$ ($n$ copies of 1 here). Consider the line segments $\ell_1, \ell_2, \dots, \ell_n$, where $\ell_i$ has endpoints $(i, 0)$

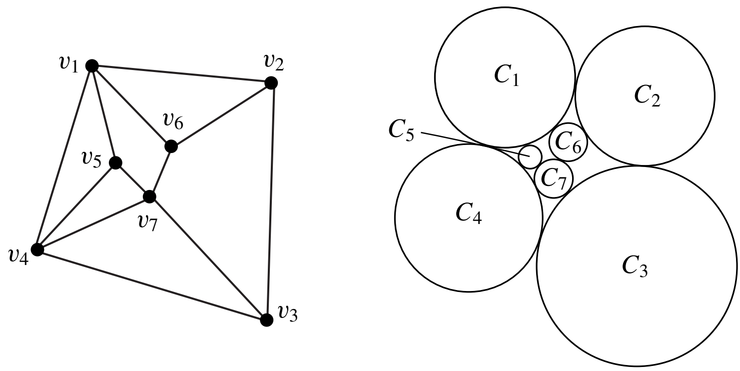


**Figure 6.** A planar graph $G$ and the corresponding closed realization using the circle packing theorem.

and $(0, n+1-i)$, and observe that $\ell_i \cap \ell_j \neq \varnothing$ for any $i \neq j$. Moreover, no three of these lines are concurrent. This gives a closed convex realization of $K_n$. Now to realize $K_{n_1,n_2,\dots,n_k}$, start with a closed convex realization of $K_k$ as constructed in the realization of (2). Replace each line segment $\ell_i$ with $n_i$ disjoint parallel translates of $\ell_i$ that are arbitrarily close in distance to $\ell_i$, and call these segments $s_{i1}, s_{i2}, \dots, s_{in_i}$. Observe that by construction, $s_{ij} \cap s_{ij'} = \varnothing$ for any $j \neq j'$. Moreover, $s_{ij} \cap s_{i'j'} \neq \varnothing$ for $i \neq i'$ because $l_i \cap l_{i'} \neq \varnothing$ and $s_{ij}$ and $s_{i'j'}$ are arbitrarily close and parallel to $\ell_i$ and $\ell_{i'}$ respectively. Moreover, if any three line segments $s_{ij}, s_{i'j'}, s_{i''j''}$ had a point in common, then $l_i, l_{i'}, l_{i''}$ would, which they don't. Hence the union of the sets $\{s_{i1}, s_{i2}, \dots, s_{in_i}\}_{i=1}^k$ gives a closed convex realization of $K_{n_1,n_2,\dots,n_k}$. See Figure 7 for examples of the constructions in the proof of (2).

It remains to prove (3). Without loss of generality, we assume $N_G(u_k) = \{v_1, v_2, \dots, v_r\}$, indexed in such a way that each set $N_G(u_j)$ is $\{v_1, v_2, \dots, v_s\}$

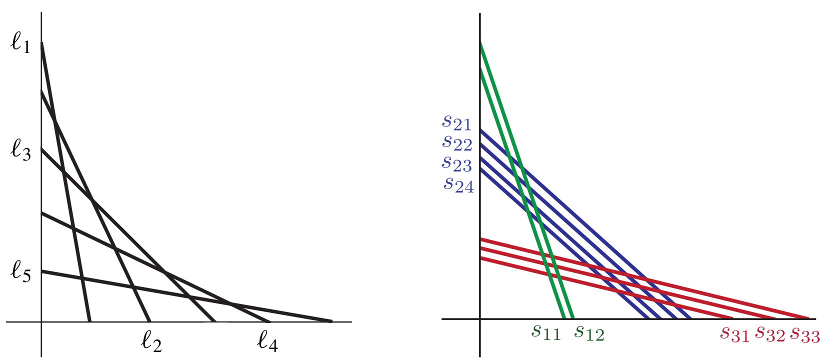


**Figure 7.** A closed convex realization of $K_5$ (left) and a closed convex realization of $K_{2,4,3}$ (right), as constructed in the proof of Proposition 3.1.

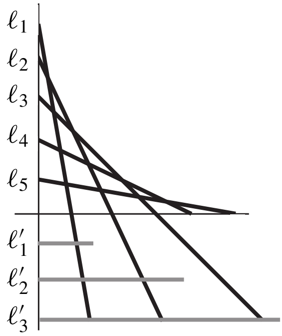


**Figure 8.** A closed convex realization of the graph $G$ with vertices $v_1, v_2, v_3, v_4, v_5, u_1, u_2, u_3$, where the induced graph on $v_1, \dots, v_5$ is complete, and $N(u_3) = \{v_1, v_2, v_3\}$, $N(u_2) = \{v_1, v_2\}$ and $N(u_1) = \{v_1\}$.

for some $s$. To realize $G$, first start with a realization of $K_n$ as in the proof of (2), where $v_j$ is represented by $\ell_j$ for each $j$. Now, extend each line segment $\ell_j$ for $1 \le j \le r$ so that $(0, j)$ remains as an endpoint, the slope remains the same, but the lower endpoint has $y$-coordinate $-k$. Then, for each $s$ with $1 \le s \le k$, introduce a line segment $\ell'_s$ that lies on the line in the $xy$-plane given by $y = s$, and only intersects the line segments in the set $\{\ell'_j \mid j \in N_G(u_s)\}$. The line segments $\ell_1, \dots, \ell_n, \ell'_1, \dots, \ell'_k$ give a closed realization of $G$. See Figure 8 for an example of this construction. □

Thus far, we have exhibited classes of graphs that can be realized in $\mathbb{R}^2$, including any planar and some nonplanar graphs $G_{\mathcal{C}}$. We now show how to adjust any nonplanar graph by edge subdivision to create a new graph that cannot be realized in $\mathbb{R}^2$.

**Proposition 3.2.** *Let $G$ be a nonplanar graph. Let $G'$ be the graph obtained from $G$ by replacing each edge $v_i v_j$ by a length-2 path $v_i, v_{ij}, v_j$ (we refer to this as the* edge subdivision *of $G$ throughout). Then $G'$ does not have an open convex realization in $\mathbb{R}^2$, and hence its minimal embedding dimension is* 3.

*Proof.* Suppose by contradiction that $G'$ has an open convex realization in $\mathbb{R}^2$. Let the graph $G$ have vertex set $\{v_1, v_2, \dots, v_n\}$, so $G'$ has as its vertices $\{v_i \mid i = 1, \dots, n\}$ together with vertices $\{v_{ij} \mid v_i v_j \in E(G)\}$, where for any $i, j$, the vertex $v_{ij}$ is adjacent only to $v_i$ and $v_j$. Suppose the open convex realization $\mathcal{U}$ of $G'$ consists of the sets $\{U_i\}$ and $\{U_{ij}\}$, where for any $i$, $U_i$ is the open convex set corresponding to $v_i$, and for any $i \ne j$ with $v_i v_j \in E(G)$, $U_{ij}$ is the open convex set corresponding to $v_{ij}$.

First, for all $i = 1, \dots, n$ select a point $p_i$ in $U_i$ that does not lie in any other sets in $\mathcal{U}$. Then, for every pair $i, j$ such that $v_i$ and $v_j$ are adjacent in $G$, note that $U_i \cap U_{ij}$ and $U_j \cap U_{ji}$ are nonempty, so we can also select points $x_{ij}$ and $x_{ji}$ in $U_i \cap U_{ij}$ and $U_j \cap U_{ij}$, respectively. Let the line segment $x_{ij} x_{ji}$ intersect $\partial U_i$

and $\partial U_j$ at points $p_{ij}$ and $p_{ji}$, respectively. Define the path $P_{ij}$ from $p_i$ to $p_j$ by concatenating the line segments $p_i p_{ij}$, $p_{ij} p_{ji}$, and $p_{ji} p_j$ in that order.

Now consider another pair of indices $k, l$. We claim that two different paths $P_{ij}$ and $P_{kl}$ can only intersect at the points $p_i$,$p_j$,$p_k$ or $p_l$, if anywhere. To show this, it is enough to show that among any pair of line segments, one chosen from $\{p_i p_{ij}, p_{ij} p_{ji}, p_{ji} p_j\}$ and one from $\{p_k p_{kl}, p_{kl} p_{lk}, p_{lk} p_l\}$, their intersection (if it exists), must be one of the points $p_i$, $p_j$, $p_k$ or $p_l$. We split this into three cases:

First, consider the intersection of $p_i p_{ij}$ and $p_k p_{kl}$. If $i = k$ then the two segments can only intersect at $p_i$, unless $j = l$, in which case the segments were the same segments to begin with. If $i \neq k$, then observe that $p_i p_{ij} \in U_i$, $p_k p_{kl} \in U_k$ and $U_i \cap U_k$ is empty because $v_i$ and $v_k$ are not adjacent in $G'$. A similar argument establishes our desired result when the pair of segments in question are $\{p_i p_{ij}, p_{kl} p_k\}$, $\{p_{ij} p_i, p_{kl} p_k\}$ and $\{p_{ij} p_i, p_k p_{kl}\}$.

Second, consider the intersection of $p_{ij} p_{ji}$ and $p_{kl} p_{lk}$. Notice that $p_{ij} p_{ji} \subseteq U_{ij}$ and $p_{kl} p_{lk} \subseteq U_{kl}$. Since $v_{ij}$ and $v_{kl}$ are not adjacent in $G'$, $U_{ij} \cap U_{kl}$ is empty, so the two paths in question cannot intersect.

Finally, consider the intersection of $p_i p_{ij}$ and $p_{kl} p_{lk}$. Suppose that $i = k$. When $j = l$, the segments in question are $p_i p_{ij}$, $p_{ij} p_{ji}$ but these are from the same path $P_{ij}$ so we need not consider this situation. When $j \neq l$, $p_i p_{ij} \subseteq U_i \cup U_{ij}$, and $p_{il} p_{li} \subseteq U_{il} \backslash U_i$. Since $j \neq l$, $U_{ij} \cap U_{il} = \varnothing$, and hence $(U_i \cup U_{ij}) \cap (U_{il} \backslash U_i) = \varnothing$, so the two segments in question do not intersect. A similar argument establishes the result when $j = l$. It remains to establish the desired result when $i \neq l, k$. Suppose for a contradiction that $p_i p_{ij}$ intersects $p_{kl} p_{kl}$. Since $p_i p_{ij} \subseteq U_i \cup \partial U_i$, and $p_{kl} p_{lk} \subseteq U_{lk}$, this implies $(U_i \cup \partial U_i) \cap U_{lk}$ is nonempty. However, this is impossible because $U_i \cap U_{lk} = \varnothing$ (because $v_i$ and $v_{lk}$ are not adjacent in $G'$) and $\partial U_i \cap U_{lk} = \varnothing$.

The above argument establishes that two distinct paths $P_{ij}$, $P_{kl}$ can only intersect at their endpoints. Construct a graph $G''$ on the same vertex set as $G$ with two vertices adjacent precisely when they are adjacent in $G$, but with each edge $v_i v_j$ drawn precisely along the path $P_{ij}$. The graph $G''$ is a planar embedding of $G$, contradicting that $G$ is not planar. □

## 4. Future directions

This paper initiated the program of studying $k$-sparse codes, with a full characterization of the structure of 2-sparse codes. Section 2 was dedicated to a topological and analytic investigation of such codes in order to achieve a full characterization of realizability through Theorem 1.3, which additionally told us that any realizable 2-sparse code has minimal embedding dimension at most 3. Section 3 then began the study of differentiating 2-sparse codes by embedding dimension through Propositions 3.1 and 3.2. The most pressing questions are how these investigations, which relied heavily on the graph-like structure of these codes, could generalize when $k > 2$.

**Question 4.1.** *For a particular* $k$, *how can we characterize which* $k$*-sparse codes are realizable*? *More specifically*, *given a positive integer* $\ell$, *for which* $k$*-sparse codes is* $d(\mathcal{C}) = \ell$?

In investigating the minimum embedding dimension of a $k$-sparse code, certain dimension bounds can be used. For example, suppose $\mathcal{C}$ is a $k$-sparse code with $\Delta = \Delta(\mathcal{C})$, and let $f_d(\Delta)$ be the number of codewords in $\Delta$ with support size $d+1$. Then, by applying the fractional Helly theorem, we find $k > f_d(\Delta)/\binom{n-1}{d}$; this was noted in [Curto et al. 2017]. Similar to this, many known bounds rely solely on the combinatorial information in the code and in particular the simplicial complex $\Delta(\mathcal{C})$. While often dimension bounds are the best known results, a more specific investigation in [Rosen and Zhang 2017] gives a full characterization of 1-dimensional codes. Our work thus focuses on distinctions between dimensions 2 and 3 for 2-sparse codes, as a beginning step towards a characterization of 2-dimensional codes.

However, in addressing the question of whether a $k$-sparse code is realizable at all, an investigation into the topology can provide insight beyond what is apparent from the combinatorics. This is especially evident from the developments in Section 2. The key idea there was shifting from one realization of a code to another by shrinking or expanding sets. Indeed, this method has been applied with more generality and great success in [Cruz et al. 2019]. The question then for $k$-sparse codes for $k > 2$ is what analogous topological operations to realizations preserve the underlying code.

**Question 4.2.** *Given a convex realization* $\mathcal{U} = \{U_1, \dots, U_n\}$ *of a code* $\mathcal{C}$ *in* $\mathbb{R}^d$, *what topological maps can be applied to the sets* $U_i$ *so that the resulting sets still form a convex realization of* $\mathcal{C}$?

## Acknowledgments

The authors thank the Dean's Office and the Department of Mathematics at Harvey Mudd College for their summer research support, and thank Carina Curto, Chad Giusti, Elizabeth Gross, Vladimir Itskov, Bill Kronholm and Anne Shiu for fruitful conversations.

## References

[Amenta et al. 2017] N. Amenta, J. A. De Loera, and P. Soberón, "Helly's theorem: new variations and applications", pp. 55–95 in *Algebraic and geometric methods in discrete mathematics* (San Antonio, 2015), edited by H. A. Harrington et al., Contemp. Math. **685**, Amer. Math. Soc., Providence, RI, 2017. MR Zbl

[Amzi Jeffs 2018] R. Amzi Jeffs, "Morphisms of neural codes", preprint, 2018. arXiv

[Amzi Jeffs and Novik 2018] R. Amzi Jeffs and I. Novik, "Convex union representability and convex codes", preprint, 2018. arXiv

[Chen et al. 2019] A. Chen, F. Frick, and A. Shiu, "Neural codes, decidability, and a new local obstruction to convexity", *SIAM J. Appl. Algebra Geom.* **3**:1 (2019), 44–66. MR

[Cruz et al. 2019] J. Cruz, C. Giusti, V. Itskov, and B. Kronholm, "On open and closed convex codes", *Discrete Comput. Geom.* **61**:2 (2019), 247–270. MR Zbl

[Curto et al. 2013a] C. Curto, V. Itskov, K. Morrison, Z. Roth, and J. L. Walker, "Combinatorial neural codes from a mathematical coding theory perspective", *Neural Comput.* **25**:7 (2013), 1891–1925. MR

[Curto et al. 2013b] C. Curto, V. Itskov, A. Veliz-Cuba, and N. Youngs, "The neural ring: an algebraic tool for analyzing the intrinsic structure of neural codes", *Bull. Math. Biol.* **75**:9 (2013), 1571–1611. MR Zbl

[Curto et al. 2017] C. Curto, E. Gross, J. Jeffries, K. Morrison, M. Omar, Z. Rosen, A. Shiu, and N. Youngs, "What makes a neural code convex?", *SIAM J. Appl. Algebra Geom.* **1**:1 (2017), 222–238. MR Zbl

[Danzer et al. 1963] L. Danzer, B. Grünbaum, and V. Klee, "Helly's theorem and its relatives", pp. 101–180 in *Convexity* (Seattle, 1961), edited by V. Klee, Proc. Symp. Pure Math. **7**, Amer. Math. Soc., Providence, RI, 1963. MR Zbl

[Eckhoff 1993] J. Eckhoff, "Helly, Radon, and Carathéodory type theorems", pp. 389–448 in *Handbook of convex geometry, A*, edited by P. M. Gruber and J. M. Wills, North-Holland, Amsterdam, 1993. MR Zbl

[Hubel and Wiesel 1959] D. H. Hubel and T. N. Wiesel, "Receptive fields of single neurons in the cat's striate cortex", *J. Physiology* **148**:3 (1959), 574–591.

[Lienkaemper et al. 2017] C. Lienkaemper, A. Shiu, and Z. Woodstock, "Obstructions to convexity in neural codes", *Adv. Appl. Math.* **85** (2017), 31–59. MR Zbl

[Lin et al. 2014] A. C. Lin, A. M. Bygrave, A. de Calignon, T. Lee, and G. Miesenböck, "Sparse, decorrelated odor coding in the mushroom body enhances learned odor discrimination", *Nature Neurosci.* **17**:4 (2014), 559–568.

[Matoušek 2002] J. Matoušek, *Lectures on discrete geometry*, Graduate Texts in Math. **212**, Springer, 2002. MR Zbl

[O'Keefe and Dostrovsky 1971] J. O'Keefe and J. Dostrovsky, "The hippocampus as a spatial map: preliminary evidence from unit activity in the freely-moving rat", *Brain Res.* **34**:1 (1971), 171–175.

[Rosen and Zhang 2017] Z. Rosen and Y. X. Zhang, "Convex neural codes in dimension 1", preprint, 2017. arXiv

[Tancer 2013] M. Tancer, "Intersection patterns of convex sets via simplicial complexes: a survey", pp. 521–540 in *Thirty essays on geometric graph theory*, edited by J. Pach, Springer, 2013. MR Zbl

Received: 2017-10-26 Revised: 2018-10-24 Accepted: 2018-12-05

rajeffs@uw.edu *Department of Mathematics, University of Washington, Seattle, WA, United States*

omar@g.hmc.edu *Department of Mathematics, Harvey Mudd College, Claremont, CA, United States*

nsuaysom@g.hmc.edu *Harvey Mudd College, Claremont, CA, United States*

awachtel@g.hmc.edu *Harvey Mudd College, Claremont, CA, United States*

neyoungs@colby.edu *Department of Mathematics and Statistics, Colby College, Waterville, ME, United States*

# involve

msp.org/involve


INVOLVE YOUR STUDENTS IN RESEARCH

*Involve* showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.




See inside back cover or msp.org/involve for submission instructions. The subscription price for 2019 is US $195/year for the electronic version, and $260/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFLOW® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

# involve

2019 vol. 12 no. 5