Investigating an algebraic signature for max intersection-complete codes

Megan K. Franke, Molly Hoch

REU 2017

Abstract

Place cells represent certain animals' location relative to their surrounding environment. From the receptive fields of these cells, we build neural codes. There is particular biological interest in studying which neural codes can be represented by convex receptive fields. One such type of convex code is the max intersection-complete code. In order to better understand neural codes, we associate to each an ideal, called the neural ideal. Here we study the canonical form of the neural ideal in the interest of finding a signature for max intersection-complete codes. We provide an algorithm for determining maximal codewords given a canonical form, and provide sufficient conditions for a code to be non-convex based on its canonical form.

1 Introduction

In 2014 John O'Keefe was a recipient of the Nobel Prize in Physiology or Medicine for his discovery in 1971 of place cells, neurons which signal to an animal its location relative to its environment [2]. Each neuron fires in a convex region of the environment referred to as a place field, or a receptive field. We build a neural code from these place fields and their intersections. Because these place fields appear as convex regions biologically, there is interest in studying which neural codes can be realized convexly.

In [5], Curto et al. developed an algebraic structure for studying neural codes. From a neural code we build its corresponding neural ideal, which has as a minimal generating set the canonical form. Algorithms for determining the canonical form of a neural ideal can be found in [5], [7]. The canonical forms of simplicial complexes and intersection-complete codes were classified in [4].

In [3], Cruz et al. presented the following theorem:

Theorem 1. *Suppose $\mathcal{C}$ is a max intersection-complete code. Then $\mathcal{C}$ is both open convex and closed convex.*

This theorem, along with Curto et al.'s work with the canonical form motivates our goal of finding an algebraic signature in the canonical form for max intersection-complete codes. Max intersection-complete codes make up a broad category of convex codes, and such a signature would decrease computational time needed to determine whether or not a code is max intersection-complete and thus convex open.

In this paper, we provide an algorithm that identifies the facets of a code given its corresponding canonical form. We also state a necessary condition for a canonical form to correspond to a max-intersection complete code. We use definitions from [5], and use the `NeuralIdeals` package presented in [7] in order to calculate the canonical forms of codes throughout [1], [6].

2 Background

In this section we define those terms which give us the necessary algebraic structures used to describe neural codes.

Definition 1. A *neural code* $\mathcal{C}$ on a set of n neurons is a set of subsets of $[n]$. The elements of $\mathcal{C}$ are called *codewords*. We say that a codeword $\sigma \in \mathcal{C}$ is a *maximal codeword*, or *facet*, of $\mathcal{C}$ if σ is not properly contained in any other codeword in $\mathcal{C}$. For simplicity, we write a codeword $\{s_1, \ldots, s_m\}$ as $s_1 \ldots s_m$.

Definition 2. A *realization* of a code $\mathcal{C}$ is a collection of sets $\mathcal{U} = \{U_1, \ldots, U_n\}$ where $U_i \in \mathbb{R}^d$ such that $\mathcal{C} = \mathcal{C}(\mathcal{U}) := \{\sigma \in [n] \mid U_\sigma \setminus \bigcup_{j \in [n] \setminus \sigma} U_j \neq \emptyset\}$. A $U_i \in \mathcal{U}$ is a *place field* of the neuron i.

We define $U_\sigma := \bigcap_{i \in \sigma} U_i$ for $\sigma \in [n]$.

By convention, we assume that every code contains $\emptyset$. Codes which contain all possible intersections of codewords are called *intersection-complete codes*. Codes which contain all possible intersections of facets are called *max intersection-complete codes*. In this work, we are interested in max intersection-complete codes.

Example 1. Let $\mathcal{C} = \{123, 124, 12, 13, 14, \emptyset\}$. The facets of $\mathcal{C}$ are $\{123, 124\}$, and $123 \cap 124 = 12 \in \mathcal{C}$. Thus $\mathcal{C}$ is max intersection-complete however, $12, 13 \in \mathcal{C}$ but $12 \cap 13 = 1 \notin \mathcal{C}$, so $\mathcal{C}$ is not intersection-complete. Figure 1 depicts a realization of the receptive fields for this code.

Definition 3. The *maximal code* on n neurons is defined to be $\mathcal{C}_{max}(n) := \{\sigma : \sigma \subseteq [n]\}$. Note that the maximal code contains 2^n codewords and for every code $\mathcal{C}$ on n neurons, $\mathcal{C} \subseteq \mathcal{C}_{max}(n)$.

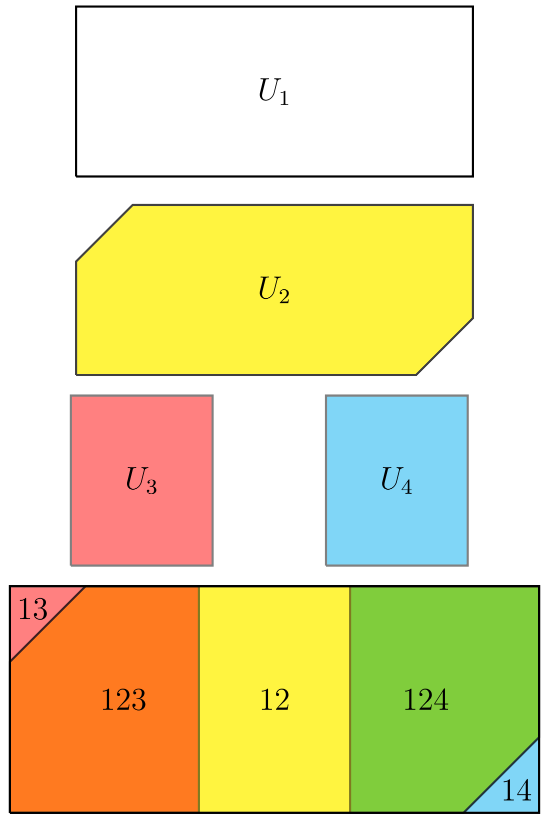

Figure 1: The top four images show the receptive fields $U_1, \dots, U_4$. The last image shows the receptive field code $\mathcal{C}(\mathcal{U})$, with codewords labelled.

In order to better understand the algebraic structure of neural codes, we study the *neural ideal*, which was first defined in [5]. This is an ideal of the *neural ring*, also defined in [5].

Definition 4. The *neural ideal* of a code $\mathcal{C}$ on n neurons is defined to be

$$J_{\mathcal{C}} := \left\langle x_\sigma \prod_{j \in \tau} (1 - x_j) : \sigma \notin \mathcal{C}, \tau = [n] - \sigma \right\rangle.$$

Here $x_\sigma = \prod_{i \in \sigma} x_i$.

We refer to an element of $J_{\mathcal{C}}$ of the form $x_\sigma \prod_{i \in \tau}(1 - x_i)$ where $|\tau| \geq 1$ as a *pseudomonomial*. If $|\tau| = 0$, we call the element a monomial. By the construction of the neural ideal, $\sigma \cap \tau = \emptyset$.

The neural ideal has many different generating sets. We will look in particular at the generating set called the *canonical form*.

Definition 5. The *canonical form* of a neural ideal $CF(J_{\mathcal{C}})$ is the set of minimal monomial and pseudomonomials in $J_{\mathcal{C}}$ with respect to divisibility. The canonical form is split into three types of elements:

- Type 1 relations: x_σ, for $\sigma \neq \emptyset$

- Type 2 relations: $x_\sigma \prod_{i\in\tau}(1-x_i)$, for σ, $\tau \neq \emptyset$
- Type 3 relations: $\prod_{i\in\tau}(1-x_i)$, for $\tau \neq \emptyset$

A Type 3 relation implies that $\emptyset$ is not contained in the code, so we will work only with Types 1 and 2. We write $CF^1(J_\mathcal{C}) = \{m : m$ is a Type 1 relation$\}$ and $CF^2(J_\mathcal{C}) = \{m : m$ is a Type 2 relation$\}$.

The following information comes from [8], [5], and clarifies the interpretations of the Type 1 and Type 2 relations:

- For $x_\sigma \in CF^1(J_\mathcal{C})$, we have $U_\sigma = \emptyset$.
- For $x_\sigma \prod_{i\in\tau}(1-x_i) \in CF^2(J_\mathcal{C})$, we have $U_\sigma \subseteq \bigcup_{i\in\tau} U_i$.

Example 2. For our code $\mathcal{C} = \{123, 124, 12, 13, 14, \emptyset\}$ and its neural ideal $J_\mathcal{C}$, the canonical form is $CF(J_\mathcal{C}) = \{x_3x_4, x_4(1-x_1), x_3(1-x_1), x_2(1-x_1), x_1(1-x_2)(1-x_3)(1-x_4)\}$.

Here, $x_3x_4 \in CF(J_\mathcal{C})$ corresponds to $U_3 \cap U_4 = \emptyset$, which is reflected in the code because it does not contain $\{1234, 134, 234, 34\}$. We also have $x_4(1-x_1) \in CF(J_\mathcal{C})$, which means that $U_4 \subseteq U_1$, so any codeword containing 4 must also contain 1, as reflected by $\{14, 124\} \in \mathcal{C}$ and no other codewords containing 4.

Definition 6. A pseudomonomial is *simple* if it is of the form $x_\sigma \prod_{i\in\tau}(1-x_i)$ where $|\tau| = 1$. If $|\tau| > 1$ the pseudomonomial is *complex*.

Definition 7. Let $\mathcal{C}$ be a code on n neurons. We say that $\sigma \subseteq [n]$ is a *missing intersection codeword* if $\sigma \notin \mathcal{C}$ and there exist some collection $\tau_1, \tau_2, \dots \tau_k \in \mathcal{C}$ such that $\bigcap_{i\in[k]} \tau_i = \sigma$.

Example 3. In our example, $x_3(1-x_1)$ is a simple pseudomonomial, $x_1(1-x_2)(1-x_3)(1-x_4)$ is a complex pseudomonomial, and 1 is a missing intersection codeword.

The following proposition comes from [4] and characterizes the canonical forms of intersection-complete codes.

Proposition 1 (Curto et al. 2015)**.** A code is intersection-complete if and only if its canonical form consists only of monomials and simple pseudomonomials.

Our goal in this project was to find a characterization for max intersection-complete codes.

3 Main Results

Let $\mathcal{C}$ be an arbitrary code on n neurons and $\mathcal{U} = \{U_1, \dots, U_n\}$ be a set of open sets, not necessarily convex, with $\mathcal{C} = \mathcal{C}(\mathcal{U})$. Let $J_\mathcal{C}$ denote the neural ideal of $\mathcal{C}$ and $CF(J_\mathcal{C})$ denote the canonical form of $J_\mathcal{C}$. Let $\mathcal{C}_{max}(n) = \mathcal{C}_{max}$.

Starting with $\mathcal{C}_{max}$, suppose we remove all $\tau \in \mathcal{C}_{max}$ such that there exists a monomial $x_{a_1}x_{a_2}\dots x_{a_k} \in CF(J_\mathcal{C})$ with $\{a_1, a_2, \dots, a_k\} \subseteq \tau$. Denote this new subset of $\mathcal{C}_{max}$ with all such τ removed by $\mathcal{C}'_{max}$.

Proposition 2. The facets of $\mathcal{C}'_{max}$ are exactly the facets of $\mathcal{C}$.

Proof. Our construction of $\mathcal{C}'_{max}$ and the Type 1 relations give us the inclusion $\mathcal{C} \subseteq \mathcal{C}'_{max}$. Hence, we need only show that the set of facets of $\mathcal{C}'_{max}$ is contained in the set of facets of $\mathcal{C}$. Let $\tau \in \mathcal{C}'_{max}$ be maximal under inclusion. Suppose for sake of contradiction that $\tau \notin \mathcal{C}$. Then, $U_\tau = \emptyset$. By [8], we know that $\{x_\sigma : \sigma \text{ is minimal w.r.t } U_\sigma = \emptyset\} \subseteq CF(J_\mathcal{C})$. Then, either $x_\tau \in CF(J_\mathcal{C})$, or a monomial which divides x_τ is in $CF(J_\mathcal{C})$. Let $x_{\tau'} \in CF(J_\mathcal{C})$ where $x_{\tau'}|x_\tau$. Then, $\tau' \subseteq \tau$. Since $x_{\tau'} \in CF(J_\mathcal{C})$ and $\tau'|\tau$, by our construction of $\mathcal{C}'_{max}$, we get that $\tau \notin \mathcal{C}'_{max}$ which is a contradiction. Thus, every facet of $\mathcal{C}'_{max}$ is contained in $\mathcal{C}$. Since $\mathcal{C} \subseteq \mathcal{C}'_{max}$ and τ is maximal under inclusion in $\mathcal{C}'_{max}$, we know that τ must be maximal under inclusion in $\mathcal{C}$. Thus, τ is a facet in $\mathcal{C}$, proving our claim.

□

Algorithm 1 uses the ideas of Proposition 2 to provide a systematic method of finding the facets of a code. It takes as input $CF^1(J_\mathcal{C})$ and returns a list $\mathcal{C}_{fac}$ consisting of the facets of $\mathcal{C}$.

Input : The Type 1 elements of a canonical form $CF^1(J_\mathcal{C}) = \{m_{\sigma_1}, \dots, m_{\sigma_k}\}$ of a neural code $\mathcal{C}$ on n neurons
Output: A list of the facets of $\mathcal{C}$
begin
    $\mathcal{C}'_{max} = \mathcal{C}_{max}(n)$,
    for $\tau \in \mathcal{C}'_{max}$ **do**
        for $m_{\sigma_i} \in CF^1(J_\mathcal{C})$ **do**
            if $\sigma_i \subseteq \tau$ **then**
                $\mathcal{C}'_{max} = \mathcal{C}'_{max} - \{\tau\}$
            end
        end
    end
    $\mathcal{C}_{fac} = \mathcal{C}'_{max}$
    for $\nu \in \mathcal{C}_{fac}$ **do**
        for $\tau \in \mathcal{C}'_{max}$ **do**
            if $\nu \subset \tau$ **then**
                $\mathcal{C}_{fac} = \mathcal{C}_{fac} - \{\nu\}$
            end
        end
    end
    return $\mathcal{C}_{fac}$
end

Algorithm 1: Algorithm which provides a list of a code's facets given the Type 1 relations of its neural ideal.

Theorem 2. *Algorithm 1 determines the facets of a code from the neural ideal.*

Proof. Let $\mathcal{C}$ be an arbitrary code on n neurons. Suppose the codewords of $\mathcal{C}$ are unknown but $J_{\mathcal{C}}$ is known. Let $\mathcal{C}_{max}$ be the maximal code on n neurons, and construct $\mathcal{C}'_{max}$ from $CF(J_{\mathcal{C}})$. Then it follows from Proposition 2 that the facets of $\mathcal{C}$ are exactly the facets of $\mathcal{C}'_{max}$.

□

The following example illustrates this algorithm on a small code.

Example 4. Let $\mathcal{C} = \{123, 124, 12, 13, 14, \emptyset\}$ as in Example 1 and let $J_{\mathcal{C}}$ be its corresponding neural ideal. Recall that $\mathcal{C}$ is max intersection-complete. We have $CF^1(J_{\mathcal{C}}) = \{x_3x_4\}$. Because this is a neural code on four neurons,

$$\mathcal{C}_{max} = \{1234, 123, 124, 134, 234, 12, 13, 14, 23, 24, 34, 1, 2, 3, 4, \emptyset\}.$$

From $CF^1(J_{\mathcal{C}})$, we see that the only monomial in the canonical form is x_3x_4 which gives us that $\mathcal{C}'_{max} = \{123, 124, 12, 13, 14, 23, 24, 1, 2, 3, 4\}$. Thus, the facets of $\mathcal{C}'_{max}$ are $\{123, 124\}$ which are exactly the facets of $\mathcal{C}$.

Corollary 1. *Let $\mathcal{C}$ be a code on n neurons, $\tau \subseteq [n]$, and $\sigma \subseteq [n] - \tau$. If $x_\tau \in CF(J_{\mathcal{C}})$ and $x_\sigma \prod_{i\in\tau}(1 + x_i) \in CF(J_{\mathcal{C}})$, then $\mathcal{C}$ is not convex open and thus not max intersection complete.*

Proof. Let $\mathcal{C}$ be an arbitrary code on n neurons, let $J_{\mathcal{C}}$ denote the neural ideal of $\mathcal{C}$, and let $CF(J_{\mathcal{C}})$ denote the canonical form of $J_{\mathcal{C}}$.

Suppose there is $\tau \subseteq [n]$ such that $x_\tau \in CF(J_{\mathcal{C}})$. Then we have that $\bigcap_{i\in\tau} U_i = \emptyset$, but $\bigcap_{i\in\tau-j} U_i \neq \emptyset$, $\forall j \in \tau$.

Suppose further that there is $\sigma \subseteq [n] - \tau$ such that $\prod_{k\in\sigma} x_k \prod_{i\in\tau}(1 - x_i) \in CF(J_{\mathcal{C}})$. Then $U_\sigma \subseteq \bigcup_{i\in\tau} U_i$, where σ and τ are minimal with respect to this inclusion. Then, we can write U_σ as

$$U_\sigma = \Big(\bigcup_{i\in\tau} U_i\Big) \cap U_\sigma.$$

Let $j \in \tau$ be an arbitrary element. Then we can write

$$U_\sigma = \Big[\Big(\bigcup_{i\in\tau\setminus j} U_i\Big) \cap U_\sigma\Big] \cup \Big[U_j \cap U_\sigma\Big].$$

However,

$$\Big(\bigcup_{i\in\tau\setminus j} U_i\Big) \cap U_j = \emptyset.$$

Together, the above equalities imply that U_σ is the union of two disjoint sets. Then in order for U_σ to be a connected set, $\Big(\bigcup_{i\in\tau\setminus j} U_i\Big)$ and U_j cannot both be open sets. Thus, $\mathcal{C}$ is not convex open. By Cruz et al. [3], since $\mathcal{C}$ is not convex open, it is also not max intersection-complete.

□

The corollary gives us a quick method for checking for connectedness and therefore convexity of codes using only the canonical form, bypassing the need to determine facets.

Example 5. Let $\mathcal{C} = \{1234, 1235, 124, 125, 134, 135, 234, 235, 14, 15, 24, 15, 34, 35, 4, 5, \emptyset\}$, and let $J_\mathcal{C}$ be its neural ideal. We find $CF(J_\mathcal{C}) = \{x_4x_5, x_1(1-x_4)(1-x_5),$ $x_2(1-x_4)(1-x_5), x_3(1-x_4)(1-x_5)\}$. The presence of x_4x_5 implies that U_4 and U_5 are disjoint, but we have that $U_1 \subseteq U_4 \cup U_5$, $U_2 \subseteq U_4 \cup U_5$, and $U_3 \subseteq U_4 \cup U_5$. In order for U_1, U_2, or U_3 to be connected sets, one or both of U_4 and U_5 cannot be open. From Figure 2 we see that it is impossible to have the receptive field U_i connected while keeping U_4 and U_5 disjoint, hence our code is not max intersection-complete. Certainly, because 1234, 1235 are facets of $\mathcal{C}$ but $1234 \cap 1235 = 123 \notin \mathcal{C}$, we know $\mathcal{C}$ is not max intersection-complete.

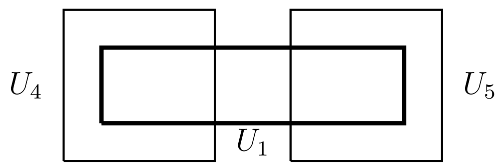

Figure 2: The squares represent U_4 and U_5, and the rectangle represents U_1. Because U_4 and U_5 are both convex open, U_1 cannot be connected and the code cannot be convex open.

4 Discussion

We have been able to take the first steps towards finding a signature for max intersection-complete codes in the canonical form of the neural ideal. Corollary 1 provides a necessary condition for the canonical form of a max intersection-complete code. We imagine that it would be simple to write an algorithm to seek out this property.

We also have a conjecture about how to find the missing intersection codewords of a code given its corresponding canonical form.

Conjecture 1. It is possible to construct some of a code's missing intersection codewords from the complex pseudomonomials present in the canonical form.

Our proposed method for this construction uses the following procedure:

1. Given a code and the canonical form of its neural ideal, pick one of the complex pseudomonomials present: $x_{a_1} \dots x_{a_m}(1-x_{b_1}) \dots (1-x_{b_\ell})$.

2. Write the intersection $\bigcap_{i \in [\ell]} a_1 \dots a_m b_i = a_1 \dots a_m$.

3. Find those neurons or sets of neurons which appear together at all times. For example, if the canonical form contains both $x_1(1-x_2)$ and $x_2(1-x_1)$, wherever 1 or 2 appears in the intersection from step 2, the other should be added.

4. Add any neurons not prevented by monomials present in the canonical form. For example, if $x_1x_2x_4$ is in the canonical form, we could not add 2 to 14.

Example 6. Let $\mathcal{C} = \{12345, 1236, 2345, \emptyset\}$.

1. Select the pseudomonomial $x_2(1-x_4)(1-x_6) \in CF(J_\mathcal{C})$.

2. Write $24 \cap 26 = 2$. Observe that $24, 26 \notin \mathcal{C}$.

3. In $CF(J_{\mathcal{C}})$, we have $x_2(1-x_3), x_3(1-x_2), x_4(1-x_5), x_5(1-x_4)$, so neurons 2 and 3 always appear together, and neurons 4 and 5 always appear together. We amend the intersection to be $2345 \cap 236 = 23$. Observe that while $2345 \in \mathcal{C}$, 236 is not.

4. The only monomials present in $CF(J_{\mathcal{C}})$ are x_4x_6 and x_5x_6, so we add 1 to both sides of the intersection to get $12345 \cap 1236 = 123$. Here we have both 12345 and $1236 \in \mathcal{C}$, and we have found a missing intersection codeword.

Further work in this topic could include proving the above conjecture, and show that if there exist missing intersections of facets, it will find at least one of them. Additionally, it would be interesting to try to find a class of nonconvex codes which fall into the category covered by Corollary 1, or to find other signatures for nonconvexity.

5 Acknowledgements

This work was conducted at the Texas A&M Mathematics REU from June-July 2017, and was supported by NSF DMS-1460766. We would like to thank our research mentor Dr. Anne Shiu for her support and guidance. We also would like to thank Nida Obatake for her assistance and encouragement.

References

[1] NeuralIdeals. Available at github.com.

[2] Neil Burgess. The 2014 Nobel Prize in Physiology or Medicine: A spatial model for cognitive neuroscience. *Neuron*, 84(6):1120–1125, Dec 2014.

[3] Joshia Cruz, Chad Giusti, Vladimir Itskov, and Bill Kronholm. On open and closed convex codes, May 2017.

[4] Carina Curto, Elizabeth Gross, Jack Jeffries, Katie Morrison, Mohamed Omar, Zvi Rosen, Anne Shiu, and Nora Youngs. What makes a neural code convex? *SIAM Journal on Applied Algebra and Geometry*, 1(1):222–238, 2017.

[5] Carina Curto, Vladimir Istkov, Alan Veliz-Cuba, and Nora Youngs. The neural ring: An algebraic tool for analyzing the intrinsic structure of neural codes. *Bulletin of Mathematical Biology*, 75(9):1571–1611, Sep 2013.

[6] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 7.1)*, 2015. `http://www.sagemath.org`.

[7] Ethan Petersen, Nora Youngs, Ryan Kruse, Dane Miyata, Rebecca Garcia, and Luis David Garcia Puente. Neural Ideals in SageMath. Sep 2016.

[8] Nora Youngs. The neural ring: Using algebraic geometry to analyze neural codes. pages 1–104, 2013.