Combinatorics of Uniqueness in Sparse Dictionary Learning

Charles J. Garfinkle, Christopher J. Hillar

Abstract

We study uniqueness in sparse dictionary learning when reconstruction of data is approximate.

Fix positive integers $k < m$ and let $\mathcal{S} = \{S_0, \ldots, S_{m-1}\}$, where

$$S_i = \{i, i+1, \ldots, i+(k-1)\} \text{ modulo } m, \quad \text{for} \quad i = 0, \ldots m-1.$$

Lemma 1. *Fix positive integers $k < m$ and let $\mathcal{S} = \{S_0, \ldots, S_{m-1}\}$, where for $i = 0, \ldots m-1$,*

$$S_i = \{i, i+1, \ldots, i+(k-1)\} \textit{ mod } m.$$

Suppose there exists a map $\pi : \mathcal{S} \to \binom{\mathbb{Z}/m\mathbb{Z}}{k}$ such that for all $\mathcal{I} \in \binom{[m]}{k}$,

$$\bigcap_{i \in \mathcal{I}} S_i = \emptyset \Longrightarrow \bigcap_{i \in \mathcal{I}} \pi(S_i) = \emptyset. \tag{1}$$

Then $\pi(S_i) \cap \cdots \cap \pi(S_{i+(k-1)}) \neq \emptyset$ for all $i \in \mathbb{Z}/m\mathbb{Z}$.

Proof of Lemma 2: Consider the set $T_m = \{(i,j) : i \in \mathbb{Z}/m\mathbb{Z}, j \in \pi(S_i)\}$, which has mk elements. By the pigeon-hole principle, there is some $p \in \mathbb{Z}/m\mathbb{Z}$ and at least k distinct $i_1, \ldots, i_k$ such that $\{(i_1, p), \ldots, (i_k, p)\} \subseteq T_m$. Hence, $p \in \pi(S_{i_1}) \cap \cdots \cap \pi(S_{i_k})$ and by (1) there must be some $v \in \mathbb{Z}/m\mathbb{Z}$ such that $v \in S_{i_1} \cap \cdots \cap S_{i_k}$. This is only possible (given $\mathcal{S}$) if $i_1, \ldots, i_k$ are consecutive modulo $\mathbb{Z}/m\mathbb{Z}$, i.e. $\{i_1, \ldots, i_k\} = \{v-(k-1), \ldots, v\}$.

We now claim there exists no additional $i^* \in \mathbb{Z}/m\mathbb{Z} \setminus \{i_1, \ldots, i_k\}$ such that $p \in \pi(S_{i^*})$. To see why, note that we would then have $p \in \pi(S_{i^*}) \cap \pi(S_{v-(k-1)}) \cap \cdots \cap \pi(S_v)$ and (1) would imply that every k-element subset of $\{i^*\} \cup \{v-(k-1), \ldots, v\}$ is a consecutive set. This is only possible if $m = k+1$; but then there can't have been $k+1$ distinct elements of $\binom{\mathbb{Z}/m\mathbb{Z}}{k}$ all containing p since there are only $\binom{m-1}{m-2} = m-1 = k$ distinct elements of $\binom{\mathbb{Z}/m\mathbb{Z}}{m-1}$ which contain p. Thus, letting $T_{m-1} \subset T_m$ be the set of elements of T_m not having p as a second coordinate, we have $|T_{m-1}| = (m-1)k$ and the proof follows by iterating these arguments. ∎

Lemma 2. *Let $k \geq 2$ and $m > 2k$ (or $m > 3$ if $k = 2$) and suppose there exists a map $\pi : \mathcal{S} \to \binom{\mathbb{Z}/m\mathbb{Z}}{k}$ such that for all $S, S' \in \mathcal{S}$ we have*

$$S \cap S' = \emptyset \Longrightarrow \pi(S) \cap \pi(S') = \emptyset. \tag{2}$$

Then we also have $S \cap S' \neq \emptyset \Longrightarrow \pi(S) \cap \pi(S') \neq \emptyset$ for all $S, S' \in \mathcal{S}$.

Proof of Lemma 2: Consider the set $T_m = \{(i,j) : i \in \mathbb{Z}/m\mathbb{Z}, j \in \pi(S_i))\}$, which has mk elements. By the pigeon-hole principle, there is some $p \in \mathbb{Z}/m\mathbb{Z}$ and k distinct $i_1, \ldots, i_k$ such that $(i_1, p), \ldots, (i_k, p) \in T_m$. Letting $\mathcal{I} = \{i_1, \ldots, i_k\}$, we have $p \in \pi(S_i)$ for all $i \in \mathcal{I}$ and by (2) we must have $S_i \cap S_j \neq \emptyset$ for all $i, j \in \mathcal{I}$. We claim that $\mathcal{I}$ must therefore consist of consecutive integers modulo m. To see why, suppose w.l.o.g. that $k-1 \in \mathcal{I}$. Then $\mathcal{I} \subset [0, 2k-2]$. If $k = 2$ then we are done; otherwise, suppose that for some $i, j \in \mathcal{I}$, $i > j$ we have $(S_i \cap S_j) \cap \mathcal{I} = \emptyset$. Then $j + (k-1) < i$, i.e. $j \in [0, k-2]$ whereas $i + (k-1) \geq j \geq m > 2k$, i.e. $i \in [k+2, 2k-2]$. Hmm... **[Proof idea: You can't have three pairwise intersecting sets which don't share a common element.]** Hence there exists some $v \in \mathbb{Z}/m\mathbb{Z}$ such that $\mathcal{I} = \{v-(k-1), \ldots, v\}$ and $p \in \pi(S_{v-(k-1)}) \cap \cdots \cap \pi(S_v)$.

Suppose now that there exists some additional $i^* \in \mathbb{Z}/m\mathbb{Z} \setminus \{v-(k-1), \ldots, v\}$ such that $p \in \pi(S_{i^*})$. Then $p \in \pi(S_{i^*}) \cap \pi(S_i)$ for all $i \in \{v-(k-1), \ldots, v\}$. Hence by (2) we have $S_{i^*} \cap S_i \neq \emptyset$ for all $i \in \{v-(k-1), \ldots, v\}$ which is impossible since $m \geq 2k$. Thus there can be no such i^*. Letting $T_{m-1} \subset T_m$ be the set of elements of T_m not having p as a second coordinate, we have $|T_{m-1}| = (m-1)k$ and the proof follows by iterating the previous arguments. ∎

Lemma 3. *Suppose there exists a map $\pi : \mathcal{S} \to \binom{[\mathbb{Z}/m\mathbb{Z}]}{k}$ such that for $k' \in \{r, r+1\}$,*

$$|\cap_{\ell=1}^{k'} \pi(S_{i_\ell})| \leq |\cap_{\ell=1}^{k'} S_{i_\ell}| \tag{3}$$

The research of Garfinkle and Hillar was conducted while at the Redwood Center for Theoretical Neuroscience, Berkeley, CA, USA; e-mails: cjg@berkeley.edu, chillar@msri.org.

for any set of distinct $i_1, \dots, i_{k'} \in [m]$. Then π is injective and $|\pi(S_v) \cap \cdots \cap \pi(S_{v+(r-1)})| = k-(r-1)$ for all $v \in \mathbb{Z}/m\mathbb{Z}$.

Lemma 4. *Suppose that $m \geq 2k-1$ and there is a function $\pi : \{S_0, \dots, S_{m-1}\} \to \binom{\mathbb{Z}/m\mathbb{Z}}{k}$ such that for $k' \in \{k, k+1\}$,*

$$\bigcap_{i=1}^{k} S_{i_j} = \emptyset \Longrightarrow \bigcap_{i=1}^{k} \pi(S_{i_j}) = \emptyset.$$

Then π is injective and we have:

$$|S_{i_1} \cap S_{i_2}| = 1 \Longrightarrow |\pi(S_{i_1}) \cap \pi(S_{i_2})| = 1.$$

[We need to make a table for small k, m of the .]

Let G be a k-uniform hypergraph on m nodes (each edge has exactly k elements). What is the smallest collection of edges

Problem 1. *Find the smallest family $\mathcal{F} = \{S_j : j \in J\}$ of k-element subsets of $\mathbb{Z}/mZ$ having the proeprty that for all $v \in \mathbb{Z}/mZ$, we have*

$$\{v\} = \bigcap_{i \in I} S_i, \quad \textit{for some } I \subseteq J.$$

Lemma 5. *Given a family above of size $|\mathcal{F}|$, one can find $N = k\binom{m}{k}|\mathcal{F}|$ k-sparse $\mathbf{a}_1, \dots, \mathbf{a}_N$ such that for any A satisfying the spark condition, the dataset $Y = \{A\mathbf{a}_1, \dots, A\mathbf{a}_N\}$ has a unique sparse coding.*