Class 12 Exercises

CS250/EE387, Winter 2022

In the lecture notes/videos, we saw *list-recovery*, a generalization of list-decoding. Here we'll explore a related notion called *mixture-recovery*; and another related notion known as the *Identifiable Parent Property*, which comes up in cryptography.

1. Consider the following definition.

   Definition 1. *Let $\mathcal{C} \subseteq \Sigma^n$ be a code. We say that $\mathcal{C}$ is (α, ℓ, L)-mixture recoverable if for all sets $S \subseteq \mathcal{C}$ with $|S| \leq \ell$,*

   $$|\{c \in \mathcal{C} : c_i \in \{x_i : x \in S\} \text{ for at least } \alpha n \text{ values of } i \in [n]\}| \leq L.$$

   That is, mixture-recovery is the special case of list-recovery when the input lists come from ℓ legitimate codewords all scrambled up.

   Let $\mathcal{C} \subset \Sigma^n$ be an MDS code of rate R that satisfies $R < \alpha/\ell$. Show that $\mathcal{C}$ is (α, ℓ, ℓ)-mixture-recoverable.

2. Consider the following problem, known as "traitor tracing." You[1] want to post some data on the internet, encrypted in a way so that only authorized users have access to it. You also want to be able to identify "traitors:" that is, authorized users who share their encryption keys to allow unauthorized users to access the data. Here is one way to do it (you can skip down to the definition of IPP if you don't care about the motivation). Below is a stylized/simplified explanation of the motivation; see http://web.cs.ucla.edu/~miodrag/cs259-security/chor94tracing.pdf) for a more detailed version (which uses something slightly different than IPP codes).

   - Say that your data comes in n blocks $X^{(1)}, \ldots, X^{(n)}$.
   - Let q be some parameter, and encode each block of the data under q separate encryption schemes. Let Σ be some alphabet of size q, and suppose that block $X^{(i)}$ is independently encrypted q times, once for each of the keys $\{K_{i,a} : a \in \Sigma\}$.[2] Post all of the encryptions of all of the data, but keep the keys private.
   - Let $\mathcal{C} \subseteq \Sigma^n$ be a code with N codewords. Suppose you want to authorize N users. Identify each authorized user with a codeword $c \in \mathcal{C}$, and give them the keys

     $$\{K_{i,c_i} : i \in [n]\}.$$

     That way, each authorized user has a key for each block, and can get the whole file.

[1] In this scenario, "you" are "The Authorities."

[2] Anyone with the key $K_{i,a}$ can decrypt the corresponding encryption of $X^{(i)}$.

How does this identify traitors? Suppose there is a single traitor, say the user associated with $c \in \mathcal{C}$, who publishes all of their keys. Then it is easy to identify the traitor, since we can see that the set of published keys was precisely that user's keys. But now suppose that there are ℓ traitors, colluding together, with associated codewords $c^{(1)}, \dots, c^{(\ell)}$. They assemble a set of keys

$$\left\{ K_{i, c_i^{(j_i)}} \, : \, i \in [n] \right\}$$

for some choices of $j_i \in [\ell]$. Now, you essentially see some "mixture" $z \in \Sigma^n$ of the codewords $c^{(j)}$ for $j \in [\ell]$. That is the i'th symbol of z is the i'th symbol of *one* of these codewords, but you don't know which.

We say that a code has the *Identifiable Parent Property* (IPP) if it is possible to identify at least one of the traitors, given z. Formally, we have the following definitions.

Definition 2. *Let $\mathcal{C} \subseteq \Sigma^n$ be a code, and let $S \subseteq \mathcal{C}$. For $z \in \Sigma^n$, we say that z* is consistent with *S if for all $i \in [n]$, there is some $c \in S$ so that $z_i = c_i$. (Notice that z doesn't have to be a codeword!)*

Definition 3. *Let $\mathcal{C} \subset \Sigma^n$ be a code. We say that $\mathcal{C}$ has the ℓ-(IPP) if for every $z \in \Sigma^n$, there is some $c \in \mathcal{C}$ so that for any $S \subseteq \mathcal{C}$ with $|S| \leq \ell$ that is consistent with z, we have $c \in S$.*

That is, $\mathcal{C}$ has the ℓ-IPP if for any "franken-word" $z \in \Sigma^n$ made by combining ℓ codewords of $\mathcal{C}$, there is some $c \in \mathcal{C}$ that *must* have been involved in the combination.

(There is no question for this part, just understand the definitions, and understand the motivation as much as you want to.)

3. Let $\ell \geq 2$. Show that any code with the ℓ-IPP is $(1, \ell, \ell)$-mixture recoverable.

4. Let $\ell \geq 2$. Show that any $(1/\ell, \ell, \ell)$-mixture-recoverable code has the ℓ-IPP.

5. Fill in the blank: Any RS code of rate $R =$ _____ has the ℓ-IPP.

   Hint: Problem 1 and the previous problem.

6. Suppose you have an RS code $\mathcal{C}$ in $\mathbb{F}_q^n$ of rate a bit less than what you got in the previous problem, so it has the ℓ-IPP. Let $S \subseteq \mathcal{C}$ have $|S| \leq \ell$ be our set of "traitors." Suppose that $z \in \mathbb{F}_q^n$ is consistent with S. Given z, explain how to find some element $c \in S$ in polynomial time. (That is, explain how to efficiently identify one of the traitors.)

   Hint: For any positive integer r, recall that an RS code of rate R is $(1 - \sqrt{R(1 + 1/r)}, r/\sqrt{R})$-list-decodable in time polynomial in n, r, R.

7. **(Bonus)** Show that any (MDS) code with the ℓ-IPP must have rate $O(1/\ell^2)$.