# Class 12 Exercises

CS250/EE387, Winter 2025

In the lecture notes/videos, we saw *list-recovery* and applications of the Guruswami-Sudan. Here we'll see some related notions (*mixture-recovery* and the *Identifiable Parent Property*), and yet another application to *Traitor Tracing*!

1. Consider the following definition.

   **Definition 1.** *Let $\mathcal{C} \subseteq \Sigma^n$ be a code. We say that $\mathcal{C}$ is $(\alpha, \ell, L)$-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 $\ell$ 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 $(\alpha, \ell, \ell)$-mixture-recoverable.

   *Hint: Fix $S \subseteq \mathcal{C}$ of size $\ell$. You want to show that if $c \in \mathcal{C}$ has $c_i \in \{x_i : x \in S\}$ for $\geq \alpha n$ values of $i$, then $c \in S$. Remember that for any MDS code of rate $R$, any $Rn$ positions of a codeword completely determine that codeword.*

   **Solution**

   Let $S \subseteq \mathcal{C}$ have $|S| \leq \ell$. Suppose $c \in \mathcal{C}$ has $c_i \in \{x_i : x \in S\}$ for at least $\alpha n$ values of $i$. We want to show that in fact $c \in S$. (This means that the output list is just $S$ itself, which has size at most $\ell$).

   To see this, notice that since $|S| \leq \ell$, there is some $x \in S$ so that $x$ and $c$ agree in at least $\alpha n/\ell$ places. But if $R < \alpha/\ell$, then since $\mathcal{C}$ is MDS, any $k = \alpha n/\ell$ positions determine a codeword. Therefore we must have $x = c$.

2. **There's a lot of exposition here: we'll do it on the board.** 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 $X^{(1)}, \ldots, X^{(n)}$.

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

- Let $q$ be some parameter, and encode each block of the data under $q$ separate encryption schemes. Let $\Sigma$ 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.

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 their full set of keys is posted on Reddit. But now suppose that there are $\ell$ traitors, colluding together, with associated codewords $c^{(1)}, \dots, c^{(\ell)}$. They assemble and publish 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 $\ell$-(IPP) if for every $z \in \Sigma^n$, there is some $c \in \mathcal{C}$ so that for any $S \subseteq \mathcal{C}$ with $|S| = \ell$ that is consistent with $z$, we have $c \in S$.*

That is, $\mathcal{C}$ has the $\ell$-IPP if for any "franken-word" $z \in \Sigma^n$ made by combining $\ell$ codewords of $\mathcal{C}$, we can identify some $c \in \mathcal{C}$ that was 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.)

**Solution**

Got it!

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

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

**Solution**

Let $\mathcal{C}$ have the $\ell$-IPP. Let $S \subseteq \mathcal{C}$ be a set of size $\ell$. We need to show that any $c \in \mathcal{C}$ that is consistent with $S$ is also in $S$; this is equivalent to the definition of mixture recovery.

But this is implied by the definition of the $\ell$-IPP. Informally, if we saw $z = c$, then even though $S$ is consistent with $z$, we can't trace the traitor to someone in $S$, since $c$ could also have been the traitor.

Slightly more formally, let $z = c$, and suppose that $c \in \mathcal{C}$ is consistent with $S$ but not in $S$. But then consider the set $S$, along with the sets $S_i$ formed by removing the $i$'th element of $S$ and replacing it with $c$. Then all of these sets $S, S_1, \ldots, S_\ell$ have size $\ell$, but there is no $c' \in \mathcal{C}$ that is in all of them, violating the definition.

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

   *Hint: Suppose that $\mathcal{C}$ is $(1/\ell, \ell, \ell)$-mixture-recoverable. Fix $z$ and consider all of the sets $S$ that are consistent with $z$, call them $\mathcal{S} = \{S_1, S_2, \ldots, S_M\}$. Let $c \in S_1$ be the element of $S_1$ that agrees the most with $z$. Can you show that $c \in S_j$ for all $j \geq 1$?*

**Solution**

Following the hint, let $z \in \Sigma^n$, and let $\mathcal{S} = \{S_1, \ldots, S_M\}$ be all the sets that are consistent with $z$. Let $c \in S_1$ be the element of $S_1$ that agrees the most with $z$; notice that $\text{agr}(c, z) \geq n/\ell$.

But then for any $j \in \{1, \ldots, M\}$, we have

$$c \in \{c \in \mathcal{C} \,:\, c_i \in \{x_i : x \in S_j\} \text{ for at least } n/\ell \text{values of } i\}\,,$$

since $z$ is consistent with $S_j$, and $c$ agrees with $z$ (and hence with some element of $S_j$) in at least $1/\ell$ places.

Thus, $(1/\ell, \ell, \ell)$-mixture recovery implies that $c \in S$.

So we can identify $c$ as a "traitor," and the definition of IPP is satisfied.

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

   *Hint: Problem 1 and the previous problem.*

**Solution**

From the above, we want to know when an RS code is $(1/\ell, \ell, \ell)$-mixture recoverable, which by problem 1 happens when it has rate $\alpha/\ell = 1/\ell^2$.

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 $\ell$-IPP. Let $S \subseteq \mathcal{C}$ have $|S| = \ell$ 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 at least one of the traitors.)

   *Hint: Show that for an RS code of this rate, the only codewords $c \in \mathcal{C}$ that agree with $z$ in at least a $1/\ell$ fraction of the places are elements of $S$.*

*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 $n, r, R$.*

**Solution**

Say that the rate of our RS code is $< 1/\ell^2$.

We will first show that

$$\{c \in \mathcal{C} \,:\, \mathrm{agr}(z,c) \geq n/\ell\} \subseteq S$$

for any $z$ that's consistent with $S$. That is, any $c \in \mathcal{C}$ that's close to $z$ is in the traitorous set $S$.

To see this, suppose that $c' \in \mathcal{C} \setminus S$ is some other codeword. We have

$$\mathrm{agr}(z,c') \leq \sum_{c \in S} \mathrm{agr}(c,c') < |S|n/\ell^2 = \frac{n}{\ell}.$$

Indeed, the first inequality is true because for each place that $c'$ agrees with $z$, it has to agree with at least one of the $c \in S$. The second inequality is true since the distance of the code is $n(1-R) > n - n/\ell^2$.

Thus, we can run the Guruswami-Sudan list-decoding algorithm on $z$. We want to choose $1-\sqrt{R(1+1/r)} = 1 - 1/\ell$, which we can do when $r$ is big enough and $R$ is "a bit less than $1/\ell^2$." (That is, take $R = \frac{1}{\ell^2} \cdot \frac{1}{1+1/r}$). The list-decoding algorithm will return all of the codewords $c$ within $1 - 1/\ell$ of $z$. By the above, they are all traitors.

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