Class 13 Exercises

CS250/EE387, Winter 2025

In the lecture videos/notes, we saw *Folded Reed-Solomon Codes*. Recall that the guarantee of these codes was the following:

Theorem 1. *Let $\varepsilon > 0$. There is a choice of $s = O(1/\varepsilon)$ and $m = O(1/\varepsilon^2)$ so that the following holds.*

Let $\mathcal{C} \subseteq (\mathbb{F}_q^m)^N$ be a Folded RS code with folding parameter m. (So $N = n/m$, where $n \leq q$ is the length of the original RS code). Let R be the rate of $\mathcal{C}$.

The $\mathcal{C}$ is $(1 - R - \varepsilon, L)$-list-decodable, where $L = q^s$. Moreover, for any $z \in (\mathbb{F}_q^m)^N$, the list

$$\mathcal{L} = \{c \in \mathcal{C} : \delta(c, z) \leq 1 - R - \varepsilon\}$$

is contained in a subspace $V \subseteq \mathcal{C}$ of dimension at most s.

In this exercise, we'll see that actually we can improve the list size from $L = q^s$ (which is larger than N^s, since $q \geq n \geq N$) to something that doesn't depend on the length N of the code. (Note: see https://arxiv.org/pdf/2502.14358 for an overview of recent improvements on this! It turns out that the list size can be improved to $O(1/\varepsilon)$, which is optimal!)

1. For this question, we will use the following theorem:

   Theorem 2. *Let $V \subset (\mathbb{F}_q^m)^N$ be any subspace of dimension at most s, so that for any two $c, c' \in V$, $\delta(c, c') \geq 1 - R$.*

   Let $S \subseteq [N]$ be a random set of size t. Then the probability that there exist two $c, c' \in V$ so that $c|_S = c'|_S$ is at most

   $$\mathbb{P}_S[\exists c \neq c' \in V, c|_S = c'|_S] \leq R^t \left(\frac{t}{R}\right)^s.$$

   You don't need to prove the theorem (yet!), but just make sure you understand it.

2. Consider the following (randomized) decoding algorithm for an FRS code of rate R.

   Given $z \in (\mathbb{F}_q^m)^N$:

   - Run the decoder from Theorem 1 to obtain a subspace $V \subseteq \mathcal{C}$ of dimension at most $s = O(1/\varepsilon)$ that contains the list $\mathcal{L} = \{c \in \mathcal{C} : \delta(c, z) \leq 1 - R - \varepsilon\}$.
   - Choose $S \subseteq [N]$ of size t uniformly at random. (In more detail, we will choose t elements of $[N]$, independently with replacement, to be in S. So maybe it happens that $|S| \leq t$ if there are collisions).
   - If there is a unique codeword $c \in V$ so that $c|_S = z|_S$, return c.
   - Otherwise, return FAIL.

   Let $c \in \mathcal{C}$ be such that $\delta(c, z) \leq 1 - R - \varepsilon$. Show that the probability that this algorithm returns c is at least

   $$\Pr[\text{Alg returns } c] \geq (R + \varepsilon)^t - R^t \left(\frac{t}{R}\right)^s.$$

3. Suppose that R is some constant (like, $1/4$ or something like that), and that s is large enough and ε is small enough. Show that if $t \geq \frac{100s}{\varepsilon} \ln(s/\varepsilon)$, then

$$R^t(t/R)^s \leq \frac{1}{e}(R+\varepsilon)^t.$$

Note: It's okay to be super handwavey here. In particular, feel free to use the approximation $e^x \approx 1+x$ for small x as though it were an equality, and feel free to make the constant "100" bigger if you like, and feel free to change $1/e$ to $1/2$ or $9/10$ or any constant in $(0,1)$ that you like.

4. Use the previous two parts to show that, for any z,

$$|\mathcal{L}_z| = \left(\frac{s}{\varepsilon}\right)^{O(s/\varepsilon)} = \left(\frac{1}{\varepsilon}\right)^{O(1/\varepsilon^2)},$$

where

$$\mathcal{L}_z = \{c \in \mathcal{C} \,:\, \delta(c,z) \leq 1-R-\varepsilon\}.$$

In particular, the FRS code $\mathcal{C}$ is actually $(1-R-\varepsilon, (1/\varepsilon)^{O(1/\varepsilon^2)})$-list-decodable, which is asymptotically better than what Theorem 1 gives (assuming N is way way bigger than $1/\varepsilon$).

Note: As before, assume that R is some constant, like $1/4$.

5. **Bonus.** Prove Theorem 2. We'll walk you through a slightly easier version:

Theorem 3. *Let $V \subseteq \mathbb{F}_q^n$ be any subspace of dimension at most s, so that for any two $c, c' \in V$, $\delta(c,c') \geq 1-R$.*

Let $S \subseteq [n]$ be a random (multi-)set of size t (that is, choose t elements of n, independently with replacement). Then the probability that there exist two $c, c' \in V$ so that $c|_S = c'|_S$ is at most

$$\Pr_S[c|_S = c'|_S] \leq R^t \left(\frac{t}{R}\right)^s =: p,$$

where above we are defining p to be that quantity.

(The only difference between this and Theorem 2 is that we are ignoring the folding. The folding doesn't really change the proof, it's just obnoxious to keep track of.)

(a) Let $M \in \mathbb{F}_q^{n \times s}$ be a matrix whose columns form a basis for V. Let $S \subseteq [n]$ be as in the theorem statement. Let $M|_S$ denote M restricted to the columns in S. Explain why it is enough to show that $M|_S$ is rank s with probability at least p.

(b) Say that $S = \{i_1, i_2, \ldots, i_t\}$, and imagine choosing these indices one at a time. Say we have chosen i_1 and are about to choose i_2. Explain why the i_2'th row of M is linearly independent with the i_1'st row of M with probability at least R.

(c) Continuing the line of thought above, suppose we have chosen i_1 and i_2 (and suppose that rows i_1 and i_2 span a space of dimension at most s, which will be true anyway as long as $s > 2$). Explain why the i_3'rd row of M does not lie in the span of the first two, with probability at least R.

(d) Continuing further, let $2 < r \leq t$, and suppose that you have chosen $i_1, i_2, \ldots, i_{r-1}$, **and** that you still don't have a full rank set of rows. Explain why the i_r's row of M does not lie in the span of rows $i_1, \ldots, i_{r-1}$, with probability at least R.

(e) Use the fact that you proved in part (d), along with the sufficient condition in part (a), to prove the theorem.

Hint. If we draw t rows of M and fail to get a full-rank matrix, then there are at least $t-s+1$ rows that we drew that did not increase the dimension of the span of the rows that we have...

Hint. We have $\binom{t}{t-s+1} R^{t-s+1} \leq R^t(t/R)^s$ (why?)