Problem set 2

Date: March 10, 2021 | Error Correcting Codes-S21
Instructor: Mrinal Kumar | Due: April 04, 2021

Instructions

- Please type the soutions neatly in LaTeXand submit the pdf. If you need to draw any pictures, feel free to do them by hand and attach the pictures.
- Discussion on the problems with other members of the class is permitted and to an extent, even encouraged. But, you *must* write the solutions on your own. You *must* also acknowledge any discussions you might have had with others separately for every problem.
- Please do not look up solutions on the internet or in other references. In case you use any sources outside the notes for this course, again properly acknowledge them.
- To get the most out of the problem sets, you are encouraged to think about the problems on your own before discussing them with others, consulting the references or looking at the hints (which some of the problems might have).
- For late submissions, you lose 20% of the points for every day after the deadline.
- The first problem is due to Prahladh Harsha.
- There are some bonus problems here, which you do not have to necessarily turn in, although you are strongly encouraged to think about them.

Problems

1. Let $\mathbb{F}$ be a finite field of size q. A set $K \subseteq \mathbb{F}^m$ is said to be a Kakeya set if K contains a line in every direction, i.e., for every $\mathbf{a} \in \mathbb{F}^m$, there exists a vector $\mathbf{b} \in \mathbb{F}^m$ such that the set

$$L_{\mathbf{a},\mathbf{b}} = \{\mathbf{b} + t \cdot \mathbf{a} : t \in \mathbb{F}_q\}$$

is contained in the set K. Note that the set $L_{\mathbf{a},\mathbf{b}}$ is indeed a *line* (a one dimensional affine space) contained in $\mathbb{F}^m$. Clearly, the whole set $\mathbb{F}^m$ is a Kakeya set. The goal of this problem is to use the polynomial method based techniques we saw in the class to conclude that there cannot be Kakeya sets that are much smaller in size than q^m. Intuitively, the regime of parameters we are interested in is when q is growing, and m is something much smaller than q. We will show that any Kakeya set K has size at least $\binom{q+m-1}{m}$. The proof is via contradiction. Suppose that $\tilde{K}$ is a Kakeya set of size strictly smaller than $\binom{q+m-1}{m}$.

   (a) **(10 points)**. Show that there exists an m variate polynomial $Q \in \mathbb{F}_q[\mathbf{x}]$ of total degree d, with $d < q$ such that $Q(\mathbf{c}) = 0$, for every $\mathbf{c} \in \tilde{K}$.

   (b) **(10 points)**. For an $\mathbf{a} = (a_1, a_2, \ldots, a_m) \in \mathbb{F}^m$, let $\mathbf{b} = (b_1, b_2, \ldots, b_m) \in \mathbb{F}^m$ be such that the line $L_{\mathbf{a},\mathbf{b}}$ is contained in $\tilde{K}$. Now, consider the restriction of Q on the line $L_{\mathbf{a},\mathbf{b}}$, which is the univariate polynomial

$$P_{\mathbf{a},\mathbf{b}}(T) := Q(\mathbf{b} + T \cdot \mathbf{a}) = Q(T \cdot a_1 + b_1, T \cdot a_2 + b_2, \ldots, T \cdot a_m + b_m)\,.$$

   Show that $P_{\mathbf{a},\mathbf{b}}(T)$ must be identically zero.

(c) **(5 points)**. Let Q_d be the homogeneous component of Q of degree equal to d. We know that Q_d is a non-zero homogeneous polynomial of degree $d < q$. Show that the coefficient of T^d in $P_{\mathbf{a},\mathbf{b}}(T)$ is equal to $Q_d(\mathbf{a})$.

(d) **(5 points)**. Conclude that Q_d must be identically zero.

(e) **(5 points)**. Combine the above parts to arrive at a contradiction and conclude that the $\tilde{K}$ cannot have size less than $\binom{q+m-1}{m}$.

2. In this question, we will see a list decoding algorithm for codes which are closely related to Reed-Solomon codes. We have parameters n, k, s and we work over a field $\mathbb{F}$ of size at least n and characteristic larger than k or zero. Let $\alpha_1, \ldots, \alpha_n$ be distinct elements of $\mathbb{F}$. The message space is again the space of univariate polynomials of degree at most $k-1$ over $\mathbb{F}$. The encoding of a polynomial $f \in \mathbb{F}[x]$ is given by the function $\mathrm{Enc} : \mathbb{F}[x] \to (\mathbb{F}^s)^n$, defined as follows:

$$\mathrm{Enc}(f) = \left(f(\alpha_i), \frac{\partial f}{\partial x}(\alpha_i), \ldots, \frac{\partial^{s-1} f}{\partial x^{s-1}}(\alpha_i)\right)_{i=1}^{n} .$$

In other words, the alphabet of the code is $\mathbb{F}^s$ and the encoding outputs an n length vector over $\mathbb{F}^s$ where the i^{th} coordinate contains the evaluation of f and all its derivatives of order up to $s-1$ on the input α_i.

We will now see a list decoding for these codes closely related to the algorithm that we saw in the class for Folded Reed-Solomon codes. As an input, we have a received word $\mathbf{b} \in (\mathbb{F}^s)^n$, where for every $i \in \{1, 2, \ldots, n\}$ the ith coordinate of $\mathbf{b}$ is denoted by $b_i = (b_{i,0}, \ldots, b_{i,s-1})$.

(a) **(5 points)**.What is the rate and distance of this code, as a function of n, k, s ?

(b) **(15 points)**. Let $m < s$ be a parameter. As a first step, show that for every choice of field constants $\gamma = \{\gamma_{j_0,j_1} : j_0, j_1 \in \{0, 1, \ldots, s-m\}\} \subseteq \mathbb{F}$, there is a non-zero polynomial $Q_\gamma(x, y_0, y_1, \ldots, y_{m-1})$ of the form $Q_\gamma := Q_0(x)y_0 + \ldots + Q_{m-1}(x)y_{m-1}$ such that the following conditions hold.

- Degree of Q_γ is at most $D \le \frac{n(s-m+1)}{m}$
- For every $i \in \{1, 2, \ldots, n\}$

$$Q_0(\alpha_i)b_{i,0} + \ldots + Q_{m-1}(\alpha_i)b_{i,m-1} = 0$$

- For every $i \in \{1, 2, \ldots, n\}$ and $\ell \in \{1, 2, \ldots, s-m\}$, we have

$$\sum_{j=0}^{m-1}\left(\sum_{\ell'=0}^{\ell} \gamma_{\ell',\ell} \cdot \frac{\partial^{\ell'} Q_j}{\partial x^{\ell'}}(\alpha_i) \cdot b_{i,j+\ell-\ell'}\right) = 0 ,$$

where, we follow the notation that $\frac{\partial^0 Q_j}{\partial x^0} = Q_j$.

(c) **(10 points)**. Show that there exists a choice of the constants γ for which the following is true: if f is a polynomial in $\mathbb{F}[x]$ of degree at most $k-1$ such that there exists an $i \in \{1, 2, \ldots, n\}$, with

$$b_i = \left(f(\alpha_i), \frac{\partial f}{\partial x}(\alpha_i), \ldots, \frac{\partial^{s-1} f}{\partial x^{s-1}}(\alpha_i)\right) ,$$

then, the univariate polynomial $R(x) = Q_\gamma\left(x, f, \frac{\partial f}{\partial x}, \ldots, \frac{\partial^{s-1} f}{\partial x^{s-1}}\right)$ vanishes with multiplicity at least $s-m$ at α_i.

Moreover, note that this correct choice of constants γ can be efficiently computed.

(d) **(5 points)**. Conclude that if the number of coordinates i, where $\mathbf{b}$ and $\mathrm{Enc}(f)$ agree is at least $\frac{D+k-1}{s-m}$, then $R(x)$ must be identically zero.

All that remains now for algorithmic list decoding of these codes is to be able to solve equations of the form $\sum_{j=0}^{m-1} Q_j(x)\frac{\partial^j f}{\partial x^j} = 0$ to recover all possible solutions f of degree at most $k-1$. Similar to what we saw for Folded Reed-Solomon codes, note that the space of solutions of this equation is again a linear space over $\mathbb{F}$, and it suffices to prove an upper bound on the dimension of this space, and get our hands on a basis for the subspace. We will again try to recover f one coefficient at a time, but slightly differently to what we did for Folded Reed-Solomon codes. First observe that since Q_γ is a non-zero polynomial, there must exist a j such that Q_j is non-zero. Let m_0 be the largest integer such that Q_{m_0} is non-zero. Now, observe that there must be an $a \in \mathbb{F}$ such that $Q_{m_0}(a)$ is non-zero. So, instead of recovering $f(x)$ directly, we will recover $f(x+a)$ one coefficient at a time. From $R(x) = 0$, we can replace x by $x+a$ everywhere, to get

$$R(x+a) = Q_\gamma\left(x+a, f(x+a), \frac{\partial f}{\partial x}(x+a), \ldots, \frac{\partial^{m-1} f}{\partial x^{m-1}}(x+a)\right) = 0\,.$$

(e) **(Bonus)**. In this equation, chase down the coefficients of monomials of degree $0, 1, \ldots, k-1$ and notice that each of them must be equal to zero. Using this, argue that given the coefficients of degree $0, 1, \ldots, m-2$ in $f(x+a)$, we can recover $f(x+a)$ uniquely. As a consequence, conclude that the linear space of solutions has dimension at most $m-1$.

(f) **(Bonus)**. Combine the parts together to conclude that given any $\varepsilon > 0$, there is a choice of s, m such that the resulting code as constructed above is algorithmically list decodable even when the fraction of errors is $\delta - \varepsilon$, where δ is the relative distance of the code.

(f) **(Bonus)**. Where did the characteristic of the field play a role in the list decoding algorithm described above ? Consider the small chatacteristic analog of the above codes, where we replace the derivatives in the definition by Hasse derivatives. Are the resulting codes list decodable from $\delta - \varepsilon$ errors ?