Exercises about “compressed sensing over discrete domains” (aka, group testing and syndrome decoding)

MSRI Summer School: Representations of High-Dimensional Data

July 12, 2018

Number of stars indicate difficulty (more stars mean more difficult).

1. ($**$) (**Missing proof.**) In this exercise you'll prove the fact (that we stated but did not prove in the lecture) that the matrix A we constructed out of Reed-Solomon codes was a good group testing matrix. Recall:

   - The code $\mathcal{C}_{RS} \subset \mathbb{F}_p^{p-1}$ is defined as the kernel of some matrix $H \in \mathbb{F}_p^{r \times (p-1)}$ which has the property that every $r \times r$ minor of H has full rank.
   - The matrix $A' \in \mathbb{F}_p^{(p-1)\times|\mathcal{C}_{RS}|}$ is the matrix with all the $c \in \mathcal{C}_{RS}$ as columns.
   - The matrix $A \in \mathcal{B}^{p(p-1)\times N}$ is formed from A' by replacing $i \in \{0, \ldots, p-1\}$ with a column vector $e_i \in \{0,1\}^p$. Let's require that $N \leq |\mathcal{C}_{RS}|$, and if $N < |\mathcal{C}_{RS}|$ we'll just throw out columns to make a matrix that has N columns.

   Follow the steps below to show that the matrix A can identify s sick people out of N in the group testing problem we defined in lecture.

   (a) Say that a matrix $A \subseteq \mathcal{B}^{t\times N}$ is *s-disjunct* if the following holds:

   > For all sets $S \subset [N]$ of size s, and for all $i \in [N] \setminus S$, there exists some $j \in [t]$ so that $A_{ji} = 1$ and $A_{jk} = 0$ for all $k \in S$.

   Show that if A is s-disjunct, then A can identify s sick people out of N in the group testing problem. Moreover, show this can be done in time $O(t \cdot N)$.

   (b) Suppose that A is derived from a matrix $A' \subset \mathbb{F}_p^{(p-1)\times|\mathcal{C}_{RS}|}$ as above. Show that if $\mathcal{C}_{RS}$ has the property:

   > (†) For all sets $S \subset \mathcal{C}_{RS}$ of size s, and for all $c \in \mathcal{C}_{RS} \setminus S$, there exists some $j \in [p-1]$ so that $c_j \notin \{c'_j \,:\, c' \in S\}$.

   then A is s-disjunct.

   (c) Use the property about $r \times r$ minors of H to show that for any two $c \neq c' \in \mathcal{C}_{RS}$, $\sum_{i=1}^{n} \mathbf{1}_{c_i = c'_i} \leq p - r - 2$.

   (d) Show that $\mathcal{C}_{RS}$ has the property (†) provided that $s < (p-2)/(p-r-2)$.

   (e) Choose $p \approx s \log_s(N)$ and $r \approx p(1 - 1/s)$ and conclude that there exists a group-testing matrix which can identify s sick people out of N with $t = O(s^2 \log_s^2(N))$ tests.

2. ($*$) (**Disjunctness ctd.**) Problem 1a shows that s-disjunctness is sufficient to solve the group testing problem. Is it necessary? Prove or give a counterexample.

3. (**Optimality?**) In Problem 1a, you analyzed an s-disjunct matrix $A \in \{0,1\}^{t \times N}$ with $t = O(s^2 \log_s^2(N))$.

   (a) ($**$) Show that one can do better for small s (say, $s = O(\sqrt{\log(N)})$). (Hint: Try a random matrix.)

   (b) ($*****$) Can you do better for general s?

4. (**Hamming Codes.**) In the lecture we saw *Hamming codes,* which was the solution for $s = 1$ in the syndrom decoding problem. Formally, we have the following definition:

   Definition 1. *Let $n = 2^r - 1$ for some integer r. The Hamming code $\mathcal{H}_r$ of length n is defined as the kernel of H_r, where $H_r \in \mathbb{F}_2^{r \times n}$ is the matrix which has every nonzero vector in $\{0,1\}^r$ as its columns.*

   (a) ($*$) Argue that the size $|\mathcal{H}_r|$ of the Hamming code is 2^{n-r}.

   (b) ($*$) The *distance* of a code $\mathcal{C} \subset \mathbb{F}^n$ is defined to be

   $$\min_{c \neq c' \in \mathcal{C}} \Delta(c, c')$$

   where $\Delta(a,b) := \sum_{i=1}^n \mathbf{1}_{a_i \neq b_i}$ is the *Hamming distance.* Argue that any code of distance 3 can correct one error. (Meaning that if Bob sees $c + x$ where $\|x\|_0 = 1$, then Bob can recover c).

   (c) ($*$) Argue that $\mathcal{H}_r$ has distance 3 for all r.

   (d) ($**$) Is there a code $\mathcal{C} \subseteq \mathbb{F}_2^n$ with distance 3 that is bigger than $\mathcal{H}_r$? Either give a construction of such a $\mathcal{C}$ or prove it's not possible. (Hint: Think about the sets $\{v \in \mathbb{F}_2^n : \Delta(v,c) \leq 1\}$ for $c \in \mathcal{C}$.)

   (e) ($***********^*$) For general $d \in [n]$, how big is the biggest code $\mathcal{C} \subset \mathbb{F}_2^n$ with distance d?

5. ($***$) (**Coordinated Failure.**) Consider the following n-player cooperative game, for $n = 2^r - 1$. The players $1, \ldots, n$ are each given black or white hats, uniformly at random. Each player can see the other players' hats, but not their own hat. The players are allowed to strategize before the game begins, *but are not allowed to communicate during the game.* Simultaneously, the players must all say "black," "white," or "pass." The players collectively lose if:

   - Everyone passes, **OR**
   - Any player says a color that is *not* the color of their own hat.

   The players collectively win if they do not lose.

   (a) Find a strategy where the players win with probability $1 - 1/2^r$.

   (b) Prove that your strategy is optimal.

   (c) Explain what this question has to do with the lecture and/or problem 4.

   Hint: Try it first for $n = 3$ to get some intuition.

   Hint 2: The fact that $n = 2^r - 1$ is not an accident.

6. (**) (**Dual view of Reed-Solomon Codes.**) Let $\mathbb{F}_p$ be a finite field of order p.

(a) Show that

$$\sum_{i=0}^{p-2} \alpha^i = 0 \qquad \forall \alpha \in \mathbb{F}_p \setminus \{0,1\}.$$

(Hint. You may use the fact that $\alpha^{p-1} = 1$ for all $\alpha \in \mathbb{F}_p \setminus \{0\}$.)

(b) Let $\gamma \in \mathbb{F}_p$ be a primitive element. (Recall, this means that $\{0, 1, \ldots, p-1\} = \{\gamma, \gamma^2, \ldots, \gamma^{p-1}\}$). In lecture we defined $\mathcal{C}_{RS}$ to be the kernel of the matrix

$$\begin{bmatrix} 1 & \gamma & \gamma^2 & \cdots & \gamma^{p-2} \\ 1 & \gamma^2 & \gamma^4 & \cdots & \gamma^{2(p-2)} \\ \vdots & & & & \\ 1 & \gamma^r & \gamma^{2r} & \cdots & \gamma^{r(p-2)} \end{bmatrix}$$

Show that

$$\mathcal{C}_{RS} = \left\{ (f(1), f(\gamma), \ldots, f(\gamma^{p-2})) \, : \, f \in \mathbb{F}[x], \deg(f) < n - r \right\}.$$

7. (**) (**Distance**) Show that the distance of $\mathcal{C}_{RS}$ as above is exactly $r + 1$.