Instructions: Please complete all problems in Section 1. Try to complete 2 of the problems in Section 2. You are welcome to do more than 2, but please indicate which 2 you want graded. No problems in Section 3 are required, but they might be fun to think about (some might be open-ended).

- You may collaborate on problems in Sections 1 and 2 with other members of the class; please acknowledge your collaborators. You may consult lecture notes and posted readings, but please do not use any other resources.

- You may collaborate on Section 3 problems with anyone, whether or not they are in the class; please acknowledge your collaborators. You may also use whatever resources you want.

Typing up your solutions in \LaTeX{} is encouraged (but I don’t type up my lecture notes, so I can’t be too strict). Legibility and complete sentences are required.

Section 1

1. (6 pts) Let $p = 1/8$. At what rate $R$ can you (or rather, can we so far in this class) guarantee that:

   (a) There exists a family of binary codes of rate $R$ that can correct a $p$-fraction of adversarial errors?

   (b) There exists a family of binary codes of rate $R$ that can efficiently correct a $p$-fraction of adversarial errors?

   (c) There exists a family of binary codes of rate $R$ that can correct a $p$-fraction of random errors?

   Explain your answers. (Also, please give your answers in the format “$\approx 0.33$” rather than “$\sqrt{H_2^{-1}(1/2)}$”)

2. (4 pts) Say that a group testing matrix $\Phi \in \{0,1\}^{t \times N}$ is “$d$-good” if it can identify up to $d$ defective items. More precisely, for $d < N$, $\Phi \in \{0,1\}^{t \times N}$ is $d$-good iff the map from sets $T \subset [N]$ with $|T| \leq d$ to outcomes in $\{0,1\}^t$ given by

   $$T \mapsto \left(\bigvee_{i \in T} \bigvee_{i \in T} \bigvee_{i \in T} \bigvee_{i \in T}\right)$$

   is injective.

   In class we proved that if $\Phi \in \{0,1\}^{t \times N}$ is $d$-disjunct, then it is $d$-good.

   (a) Show that for $d = 2$, there are matrices that are $d$-good but not $d$-disjunct. (It’s okay if show this by giving a somewhat silly example).

   (b) Show that any $d$-good matrix is $(d-1)$-disjunct.
Section 2

Section 2 problems are worth 10 points each.

1. \( ((\ell, \ell)\text{-list-recoverability}) \)
   
   (a) Let \( f, g \in \mathbb{F}_q[X] \) be polynomials of degree at most \( k < (q-1)/2 \). Suppose for every \( \alpha \in \mathbb{F}_q \), we are given the sets \( \{f(\alpha), g(\alpha)\} \); these sets come labeled with \( \alpha \), but we do not know which element of the set comes from \( f \) and which comes from \( g \). Give an efficient (aka, polynomial in \( q \)) algorithm for recovering \( f \) and \( g \).

   (Hint: Consider the polynomial \( p(X,Y) = (Y - f(X))(Y - g(X)) \)).

   (Second hint/supplementary resource: You may use the following fact: if you have a polynomial \( p(X,Y) \) as in Hint 1, you can factor it to find \( f(X) \) and \( g(X) \) in polynomial time. In the special quadratic case above, it might be fun to figure out an algorithm to do this from scratch.

   If you are curious to see how it’s done for general bivariate polynomials, see Section 2 here: [http://sites.math.rutgers.edu/~sk1233/courses/ANT-F14/lec10.pdf](http://sites.math.rutgers.edu/~sk1233/courses/ANT-F14/lec10.pdf).

   NOTE: Please do this from first principles (and/or from the second hint above), do not use the GS algorithm.

   (b) Suppose that \( C \subset \mathbb{F}_q^\ell \) be any code, and let \( A \in \{0,1\}^{nq \times |C|} \) be the group testing matrix obtained by the Kautz-Singleton construction we saw in class. (That is, the columns of \( A \) are codewords of \( C \) concatenated with the identity code).

   (b) Later in the course, we will see the following definition in class:

   **Definition 1.** A code \( C \in (\ell, L)\text{-list-recoverable} \) if for all collections of sets \( S_1, \ldots, S_n \in \Sigma \) so that \( |S_i| \leq \ell \) for all \( i \), there are at most \( L \) codewords \( c \in C \) so that \( c_i \in S_i \) for all \( i \).

   In part (a), you saw that full-length RS codes (aka, RS codes over \( \mathbb{F}_q \) with length \( n = q \)) of an appropriate dimension have the following property of a code \( C \subseteq \Sigma^n \):

   For any \( c,c' \in C \), given the unordered sets \( \{c_i, c'_i\} \) for \( i = 1, \ldots, n \), it is possible to recover \( c,c' \).

   What is the relationship between this property and \( (2,2)\text{-list-recoverability} \)? (That is, are they the same? Is one stronger than the other? Are they incomparable?)

2. **(Efficient group testing algorithms)** Let \( C \subset \mathbb{F}_q^n \) be any code, and let \( A \in \{0,1\}^{nq \times |C|} \) be the group testing matrix obtained by the Kautz-Singleton construction we saw in class. (That is, the columns of \( A \) are codewords of \( C \) concatenated with the identity code).

   (a) Suppose that \( C \) is \( (d,d)\text{-list-recoverable} \), as per Definition 1, and suppose that \( |C| > d \). Show that \( A \) can identify up to \( d \) defective items. (That is, it is \( d\text{-good} \) in the sense of Section 1 Problem 2).

   **NOTE:** For partial credit, (say, 8 out of the 10 points for this problem, assuming you do part (b)) you can prove the (easier) statement that \( A \) can identify any set of \( d \) defectives. But the statement is true for sets of size at most \( d \), so it might be fun to try to that.

   (b) Suppose that \( C \) has a \( (d,d)\text{-list-recovery} \) algorithm that runs in time \( \text{poly}(n) \). Give a sublinear-time algorithm (sublinear in \(|C|\)) for the corresponding group testing scheme. (That is, given the outputs of the tests, give an algorithm to identify the \( \leq d \) defective items in time \( \text{poly}(nq) \), where \( nq \) is the number of tests in \( A \); notice this is much less than \(|C|\) which is the total number of items that were pooled.)

3. **(Codes which are good for random errors aren’t terrible for worst-case errors.)** Let \( C \subset \{0,1\}^n \) be a code of rate \( k/n \), with encoding and decoding algorithms \( E : \{0,1\}^k \to \{0,1\}^n \) and \( D : \{0,1\}^n \to \{0,1\}^k \). Fix any constant \( \gamma > 0 \) and \( p \in (0,1/2) \). Suppose that \( C \) has error probability at most \( 2^{-\gamma n} \) on the \( BSC_p \) channel: for all \( x \in \{0,1\}^k \),

   \[ \mathbb{P}_{BSC_p} \{ D(BSC_p(E(x))) \neq x \} \leq 2^{-\gamma n} \]

   Show that there is some constant \( C_p \), which depends only on \( p \), so that the relative distance of \( C \) is at least \( C_p \cdot \gamma \).
4. (Capacity of the BEC) In this problem you’ll show (from scratch, without information theory) that the capacity of the binary erasure channel (with erasure probability $p$) is $1 - p$: that is, there exist codes of rate $1 - p - \varepsilon$ that can communicate reliably over $BEC_p$, but any code of rate $1 - p + \varepsilon$ cannot.

**NOTE:** This problem is on the long side and may be difficult if your probability theory is rusty. If you are trying to find quick problems to do, maybe look at the other ones.

(a) Suppose that $G$ is a random matrix in $\mathbb{F}_2^{k \times n}$. Show that the probability that $G$ has rank less than $k$ is at most $2^{k-n}$.

(b) Define a decoder $D : \{0, 1, \perp\}^n \to \{0, 1\}^k \cup \text{FAIL}$ to be the decoding algorithm that, on input $y \in \{0, 1, \perp\}^n$, returns $x$ if $Gx$ agrees with $y$ on all of the un-erased symbols if such an $x$ exists and is unique; otherwise it returns $\text{FAIL}$.

Let $\varepsilon > 0$ and suppose that $k \leq (1 - p - \varepsilon)n$. Let $G \in \mathbb{F}_2^{k \times n}$ be a random binary matrix. Show that there is some constant $\gamma$ so that

$$\mathbb{E}_G \left[ \mathbb{P}_{BEC_p} \{ D(BEC_p(Gx)) \neq x \} \right] \leq 2^{-\gamma n},$$

where the randomness in the $\mathbb{E}$ is over the choice of the matrix $G$, and the randomness in the $\mathbb{P}$ is over the channel $BEC_p$. (Hint: Let $J \subset [n]$ be the set of erased positions, and consider $\mathbb{E}_J \mathbb{P}_G \{ D(Gx \text{ with the set } J \text{ erased}) \}$.)

(c) Conclude that for all $\varepsilon$, there is a code of rate at least $1 - p - \varepsilon$ so that the error probability on $BEC_p$ is at most $2^{-\gamma n}$, for some $\gamma$ (which depends on $p$ and $\varepsilon$).

(d) Show that for any code with rate at least $1 - p + \varepsilon$, the error probability on $BEC_p$ must be at least $1/2$.

**Section 3**

1. What is the smallest $T$ so that there exists a $T \times N$ $d$-disjunt matrix? (In terms of $N$ and $d$)? We saw an upper bound of $T = O(d^2 \log_d^2(N))$. Find an explicit matrix that has $T = O(d^2 \log(N))$? Can you find any matrix (explicit or otherwise) that has $T = O(d^2 \log_d(N))$? What’s the best lower bound you can prove?

2. In Section 2 we defined $(\ell, \ell)$-list-recoverability. Show that any linear code that is $(\ell, \ell)$-list recoverable has rate at most $1/\ell$. Can you show this for non-linear codes?