Information and Coding Theory **Winter 2021**

Homework 4

Due: March 14, 2021

Note: *You may discuss these problems in groups. However, you must write up your own solutions and mention the names of the people in your group. Also, please do mention any books, papers or other sources you refer to. It is recommended that you typeset your solutions in LaTeX.*

1. **More on linear codes.** **[3 + 3 + 4 = 10 points]**

   Recall that a linear code $C \subseteq \mathbb{F}_q^n$ was a subspace specified by a *generator matrix* $G \in \mathbb{F}_q^{n \times k}$ such that $\forall w \in \mathbb{F}_q^k$, $\mathsf{Enc}(w) = Gw$. The *parity-check* matrix was defined as a matrix H such that the columns of H^T form a basis for the null-space of G^T. Prove the following facts about linear codes.

   (a) Prove that for a linear code C, the distance $\Delta(C)$ can be written as

   $$\Delta(C) \;=\; \min_{x \in C \setminus \{0^n\}} \mathsf{wt}(x)\,,$$

   where 0^n denotes the all-zero vector in $\mathbb{F}_q^n$ and $\mathsf{wt}(x)$ denotes the number of non-zero entries in x.

   (b) Recall that we considered the Hamming code over the field $\mathbb{F}_2$ with block-length $n = 7$ in class, defined by a parity check matrix with the seven columns corresponding to the numbers 1 through 7, written in binary. We now consider the general Hamming code, defined by the parity-check matrix $H \in \mathbb{F}_2^{r \times n}$ where $n = 2^r - 1$, and the i^{th} column of H is given by the number i written in binary using r bits (take the top entry to be the most significant bit and the bottom entry to be the least significant bit). Find the dimension, block-length and the distance for this code.

   (c) For a linear code C with generator matrix G and parity-check matrix H, it's *dual code* $C^{\perp}$ is defined as a code with generator matrix H^T. Prove that G^T is a parity-check matrix for $C^{\perp}$. Find the message length, block length and distance for the *dual code* of the Hamming code defined above.

2. **Good distance codes from linear compression.** **[3 + 3 + 6 = 12 points]**

   In class, we saw that a linear compression scheme can be used to obtain capacity-achieving codes for the binary symmetric channel. Here, we will show that a linear compression scheme with a good *probabilistic* guarantee, also yields codes which have good distance, and can hence be used to correct *worst-case* errors. Let H be an arbitray matrix in $\mathbb{F}_2^{m\times n}$, which yields a good linear compression scheme for $Z \sim (\mathsf{Bern}(p))^n$, i.e., there exists a (deterministic) decompression algorithm $\mathsf{Decom} : \mathbb{F}_2^m \to \mathbb{F}_2^n$ such that

$$\mathop{\mathbb{P}}_{Z\sim(\mathsf{Bern}(p))^n}[\mathsf{Decom}(HZ) \neq Z] \;\leq\; 2^{-t}\,.$$

For the following problem, assume that H has full row-rank i.e., $\mathrm{im}(H) = \mathbb{F}_2^m$. You can also assume that the decompression algorithm always "checks its answer" i.e., if given w it returns z, then we do have that $Hz = w$ (of course, since we are compressing, we have $m < n$, and there might also exist other z' such that $Hz' = w$.) Also, take $p < 1/2$.

Prove the following:

(a) The error probablity for *any* (deterministic) decompression algorithm can be written as

$$\mathop{\mathbb{P}}_{Z\sim(\mathsf{Bern}(p))^n}[\mathsf{Decom}(HZ) \neq Z] \;=\; 1 - \sum_{w\in\mathbb{F}_2^m}\mathop{\mathbb{P}}_{Z\sim(\mathsf{Bern}(p))^n}[Z = \mathsf{Decom}(w)]\;.$$

(b) Conclude from the above expression that the smallest error probability is achieved by the following (maximum-likelihood) decompression map

$$\mathsf{Decom}(w) \;:=\; \mathop{\arg\min}_{x:Hx=w}\{\mathsf{wt}(x)\}\;.$$

(c) Use the above to show that the code $C \subseteq \mathbb{F}_2^n$ with the above matrix H as the parity-check matrix H, i.e.,

$$C \;=\; \{x \in \mathbb{F}_2^n \mid Hx = 0\}\;,$$

has distance at least $t/(\log(1/p))$.

[Hint: Using $x \in C$, for each $z \in \mathbb{F}_2^n$ such that Hz that is correctly decompressed, find a z' such that Hz' is incorrectly decompressed. How do $\mathbb{P}_{Z\sim(\mathsf{Bern}(p))^n}[Z = z']$ and $\mathbb{P}_{Z\sim(\mathsf{Bern}(p))^n}[Z = z]$ compare?]

3. **Scrambled Reed-Solomon Codes [by Venkat Guruswami].** **[4 + 4 = 8 points]**

Let $\{a_1, \ldots, a_n\}$ be distinct elements of $\mathbb{F}_q$ used to define a Reed-Solomon code $C \subseteq \mathbb{F}_q^n$ with dimension k. Assume that $k < n/6$. Recall that a message $(m_0, \ldots, m_{k-1})$ is encoded by thinking of it as a polynomial $f(X) = \sum_{j=0}^{k-1} m_j \cdot X^j$ and taking the encoding $\mathsf{Enc}(m) = (f(a_1), \ldots, f(a_n))$.

For the following parts, assume the fact (used in class) that for a bivariate polynomial $h(X, Y)$, we can find all its factors of the form $Y - f(X)$.

(a) Suppose we sent two codewords according to the polynomials f and f' (of degree at most $k-1$) but they got mixed up. Thus, we now have two lists $(b_1, \ldots, b_n)$ and $(c_1, \ldots, c_n)$ and we know for each $i \in [n]$

$$\text{either} \quad f(a_i) = b_i \text{ and } f'(a_i) = c_i \qquad \text{or} \qquad f(a_i) = c_i \text{ and } f'(a_i) = b_i$$

Note that each coordinate could be independently scrambled i.e., it may happen that for some i, $f(a_i) = b_i$ and $f'(a_i) = c_i$ and for some $j \neq i$, $f(a_j) = c_j$ and $f'(a_j) = b_j$. Also, we don't know which is the case for which coordinate i. Give an algorithm to find both f and f'. [**Hint**: First find $f + f'$ and $f \cdot f'$.]

(b) Now, suppose that instead of getting both the values $f(a_i)$ and $f'(a_i)$ for each i, we only got one value β_i, such that for each i we either have $\beta_i = f(a_i)$ or $\beta_i = f'(a_i)$. Again, it might happen that for some i, $\beta_i = f(a_i)$ while for some other $j \neq i$, $\beta_j = f'(a_j)$ and we don't know which is the case for which i. However, we are given the promise that

$$\frac{n}{3} \le |\{i \in [n] \mid \beta_i = f(a_i)\}| \le \frac{2n}{3} \quad \text{and} \quad \frac{n}{3} \le \left|\{i \in [n] \mid \beta_i = f'(a_i)\}\right| \le \frac{2n}{3}.$$

Give an algorithm to find both f and f'.

4. **Codes and pseudorandomness.** **[Just for fun: no need to submit]**

In this problem, we will use codes to construct pseudorandom objects known as t-wise independent distributions. Let $C \subseteq \mathbb{F}_2^n$ be a linear code with distance $\Delta(C) = d$, and let $H \in \mathbb{F}_2^{(n-k)\times n}$ be the parity-check matrix of this code.

(a) First consider z uniformly distributed in $\mathbb{F}_2^{n-k}$. Using the fact that z is a random binary string of length $n-k$, prove that for any $a \in \mathbb{F}_2^{n-k} \setminus \{0^{n-k}\}$

$$\mathop{\mathbb{E}}_{z \in \mathbb{F}_2^{n-k}} \left[(-1)^{a\cdot z}\right] \;=\; 0 \qquad \text{where } a \cdot z \;=\; a^T z = \sum_{i=1}^{n-k} a_i z_i \mod 2\,.$$

[**Hint:** The bits of z are independent. Consider what happens to $(-1)^{a\cdot z}$ when you change one bit in z?]

(b) Prove that the code can be used to extend this property of the uniform distribution over length $n-k$ strings, to a distribution over n bits i.e., we can "stretch" the pseudorandomness. Consider the distribution obtained by choosing $z \in \mathbb{F}_2^{n-k}$ at random and taking $x = H^T z$. Note that $x \in \mathbb{F}_2^n$. Prove that for any $b \in \mathbb{F}_2^n \setminus \{0^n\}$ with $\mathrm{wt}(b) < d$, we have

$$\mathop{\mathbb{E}}_{\substack{x = H^T z \\ z \in \mathbb{F}_2^{n-k}}} \left[(-1)^{b\cdot x}\right] \;=\; \mathop{\mathbb{E}}_{z \in \mathbb{F}_2^{n-k}} \left[(-1)^{b\cdot (H^T z)}\right] \;=\; 0\,.$$

Such distributions are called $(d-1)$-wise independent distributions on n bits, since they "look like" the uniform distributions as long as one looks at at most $(d-1)$ bits at a time.

(c) Show that the Hamming code can be used to produce a 2-wise independent distribution on $n = 2^r - 1$ bits, starting with the uniform distribution on just r bits.