TUTORIAL XII

1 Homework 7

1. Give an example of a nontrivial (i.e., $C \neq \{0\}$) linear code C such that $C^{\perp} = C$.

   A: *Example 1: The code generated by* $(1,2) \in \mathbb{F}_5^2$ *of dimension 1, i.e.,* $C = \{(0,0),(1,2),(2,4),(3,1),(4,3)\}$ *is self dual. Example 2: The set of all vectors in* $\{0,1\}^n$ *with even weight (generated by* $[\mathbf{I}_{n/2}|\mathbf{I}_{n/2}]$*).*

2. By constructing the columns of a parity check matrix in a greedy fashion, show that there exists a binary linear code $[n,k,d]_2$ provided that

$$2^{n-k} > 1 + \binom{n-1}{1} + \cdots + \binom{n-1}{d-2}. \tag{1}$$

   This is a small improvement compared to the general Gilbert-Varshamov bound. In particular, it is tight for the $[7,4,3]_2$ Hamming code.

   A: *Construct the columns of the parity check matrix as follows: let* $H_1,\ldots,H_n$ *denote the columns. Choose a random vector* $H_1 \in \{0,1\}^{n-k}$*. For* $i = 2,\ldots,n$*, choose* $H_i \in \{0,1\}^{n-k}$ *such that* H_i *is not spanned by* $d-2$ *or fewer columns in* $\{H_1,\ldots,H_{i-1}\}$*. Suppose this algorithm terminates with* $i \leq n-1$*. Then every vector* $v \in \{0,1\}^{n-k}$ *is a sum of* $d-2$ *or fewer columns from* $\{H_1,\ldots,H_i\}$*. For* $j = 0,\ldots,d-2$*, let* $\mathsf{span}_j(H_1,\ldots,H_i)$ *denote the set of vectors obtained by summing exactly* j *of the vectors in* $\{H_1,\ldots,H_{i-1}\}$*. Then we have* $\{0,1\}^{n-k} \subseteq \cup_{j=0}^{d-2}\mathsf{span}_j(H_1,\ldots,H_i)$*. Since* $|\mathsf{span}_j(H_1,\ldots,H_i)| \leq \binom{i}{j}$ *and* $i \leq n-1$*, we have* $2^{n-k} \leq \sum_{j=0}^{d-2}\binom{n-1}{j}$ *contradicting the given condition. Therefore the algorithm terminates with* $i = n$*. Furthermore, by construction the number of linearly dependent columns of* H *is* d*. Hence the minimum distance of the code is* d *or in other words, the code defined by* H *is an* $[n,k,d]_2$*-code.*

3. (Bonus question) The Hadamard code has a nice property that it can be locally decoded. Let $x \in C_{Had,r}$ and suppose you are interested only in the i-th bit x_i of x ($i \in \{0,1,\ldots,2^r-1\}$). The challenge is that you only have access to y such that $\Delta(x,y) \leq \frac{n}{10}$ and you would like to look only at a few bits of y. Show that by querying only 2 well-chosen positions (the choice will involve some randomisation) of y, you can determine x_i correctly with probability $4/5$ (the probability here is over the choice of the queries, in particular x, y and i are fixed). *Hint:* You might want to query y at the position labelled by $u \in \{0,1\}^r$ at random and the position $u + e_i$ where $e_i \in \{0,1\}^r$ has a 1 at position i and 0 elsewhere.

   A: *Given* y *and* i*, pick* u *at random from* $\{0,1\}^r$ *and query* y *at the position labelled by* u *and* $u + e_i$*. That means we know* $\langle y,u\rangle \pmod 2$ *and* $\langle y,(u+e_i)\rangle \pmod 2$*. We have* $\langle y,u\rangle + \langle y,(u+e_i)\rangle \pmod 2 = \langle y,e_i\rangle \pmod 2 = y_i$*. We can recover* i*-th bit of* x *(or* $y_i = x_i$*) if there were no errors in positions* u *and* $u+e_i$*. Given that* $\Delta(x,y) \leq \frac{n}{10}$*, the probability that an error occurs at a given position is* $\leq \frac{1}{10}$*. By the union bound, the probability that error occurs in both positions* u *and* $u + e_i$ *is* $\leq \frac{2}{10}$*. Hence the probability of determining* x_i *correctly is* $\geq 1 - \frac{2}{10} = \frac{4}{5}$*.*

2 Homework 8

We use the notation introduced in class for Reed-Solomon codes.

1. Consider a $[4,2]_5$ Reed-Solomon code over $\mathbb{F}_5$ with $\alpha_i = 2^{i-1}$. Compute a generator matrix and a parity check matrix for this code (you may use an exercise from the tutorial for that, but you can also do it without). What is the minimum distance of this code? Check that your answer is consistent with the parity check matrix.

A: *The set of messages can be constructed as linear combinations of the polynomials X and 1 and so it suffices to construct the codes for these vectors. Since $n = 2$, we have to evaluate these polynomials at the points $\alpha_1 = 1$, $\alpha_2 = 2$, $\alpha_3 = 4$ and $\alpha_4 = 3$. We have $X \mapsto (1, 2, 4, 3)$ and $1 \mapsto (1, 1, 1, 1)$. So the generator matrix is given by*

$$\mathbf{G} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

A message $\mathbf{m} \in \mathbb{F}_5^4$ is mapped to the codeword defined by $\mathbf{mG}$ (assuming the messages are row vectors).

The parity check matrix is given by

$$\mathbf{H} = \begin{pmatrix} 1 & 2 & 4 & 3 \\ 1 & 4 & 1 & 4 \end{pmatrix},$$

constructed using the fact that the codewords vanish at points α and α^2 for some primitive element α of $\mathbb{F}_5$ (here we choose $\alpha = 2$). So for any codeword $\mathbf{c}$, $\mathbf{Hc}^T = (0, 0)^T$. It is straight-forward to verify that the minimum distance of the code is 3.

2. A well-studied family of codes is called cyclic codes. Their defining property is that if $(c_0, \ldots, c_{n-1}) \in C$ then $(c_{n-1}, c_0, \ldots, c_{n-2}) \in C$. Show that if β is a generator of $\mathbb{F}_q^*$ and $\alpha_i = \beta^{i-1}$ with $n = q - 1$, then the $[n, k]_q$ Reed-Solomon code is cyclic.

A: *The generator matrix for the above defined code is given by*

$$\mathbf{G} = \begin{pmatrix} \alpha_1^{k-1} & \alpha_2^{k-1} & \alpha_3^{k-1} & \cdots & \alpha_n^{k-1} \\ \alpha_2^{k-2} & \alpha_2^{k-2} & \alpha_3^{k-2} & \cdots & \alpha_n^{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_n \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \beta^{k-1} & (\beta^2)^{k-1} & \cdots & (\beta^{q-2})^{k-1} \\ 1 & \beta^{k-2} & (\beta^2)^{k-2} & \cdots & (\beta^{q-2})^{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \beta & \beta^2 & \cdots & \beta^{q-2} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}.$$

Let $\mathbf{c} = (c_{n-1}, \ldots, c_1, c_0) \in C$ denote a codeword. Then there is a vector $\mathbf{m} = (m_{k-1}, \ldots, m_0) \in \mathbb{F}_q^k$ such that $\mathbf{c} = \mathbf{mG}$. We need only show that $\mathbf{c}' = (c_0, c_{n-1}, \ldots, c_1)$ is a valid codeword or equivalently there exists a message $\mathbf{m}' \in \mathbb{F}_q^k$ such that $\mathbf{c}' = \mathbf{m}'\mathbf{G}$. Using the fact that $(\beta^{q-2})^j = (\beta^{-1})^j$ for $j = 0, \ldots, k-1$, we have

$$\begin{aligned} \mathbf{c}' &= (m_{k-1}, \ldots, m_0) \begin{pmatrix} (\beta^{q-2})^{k-1} & 1 & \beta^{k-1} & \cdots & (\beta^{q-3})^{k-1} \\ (\beta^{q-2})^{k-2} & 1 & \beta^{k-2} & \cdots & (\beta^{q-3})^{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta^{q-2} & 1 & \beta & \cdots & \beta^{q-3} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \\ &= (m_{k-1} \cdot (\beta^{-1})^{k-1}, \ldots, m_0 \cdot (\beta^{-1})^0) \begin{pmatrix} 1 & \beta^{k-1} & (\beta^2)^{k-1} & \cdots & (\beta^{q-2})^{k-1} \\ 1 & \beta^{k-2} & (\beta^2)^{k-2} & \cdots & (\beta^{q-2})^{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \beta & \beta^2 & \cdots & \beta^{q-2} \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \\ &= \mathbf{m}'\mathbf{G}, \end{aligned}$$

where $\mathbf{m}' = (m_{k-1}(\beta^{-1})^{k-1}, \ldots, m_0(\beta^{-1})^0)$.

3. Prove that any $[n, k]_q$ Reed-Solomon code C has the following property: for any $\mathbf{v} \in \mathbb{F}_q^k$ and any subset $S \subseteq \{1, \ldots, n\}$ of size k, there exists $\mathbf{c} \in C$ such that $c_S = \mathbf{v}$ where $\mathbf{c}_S \in \mathbb{F}_q^k$ corresponds to the vector $\mathbf{c}$ projected on the coordinates labelled by S.

A: *Fix a vector $\mathbf{v} \in \mathbb{F}_q^k$ and a subset $S \subseteq \{1, \ldots, n\}$ of size k. Suppose there is no codeword $\mathbf{c}$ with $\mathbf{c}_S = \mathbf{v}$. Then by the pigeon hole principle, there exist codewords $\mathbf{c}^1$ and $\mathbf{c}^2$ such that $\mathbf{c}_S^1 = \mathbf{c}_S^2$. So $\Delta(\mathbf{c}^1, \mathbf{c}^2) \leq n - k$ contradicting the fact that C has minimum distance $n - k + 1$.*

3 Secret Sharing

Secret sharing is a cryptographic problem of splitting a *secret* among several participants/players in such a way that the secret cannot be reconstructed unless a sufficient number of *shares* are combined. More formally, an (ℓ, m)-secret sharing scheme takes as input a set of n players $P_1, \ldots, P_n$ and a secret $s \in \mathcal{X}$ to be shared among them. The output is a set of shares $s_1, \ldots, s_n$ where s_i corresponds to P_i. The scheme must satisfy the following properties.

1. For all $A \subseteq \{1, \ldots, n\}$ with $|A| \geq m$, $\{P_i\}_{i \in A}$ can recover s from $\{s_i\}_{i \in A}$.
2. For all $B \subseteq \{1, \ldots, n\}$ with $|B| < \ell$, $\{P_i\}_{i \in B}$ *cannot* recover s from $\{s_i\}_{i \in B}$. By *cannot recover*, we mean that s is information theoretically hidden to all parties in B or equivalently, s is equally likely to take on any value in $\mathcal{X}$.

Shamir's $(\ell, \ell+1)$-secret sharing scheme: Let $\mathcal{X} = \mathbb{F}_q$ with $q \geq n$ and $1 \leq \ell \leq n-1$. Pick a random polynomial $f(x) \in \mathbb{F}_q[X]$ of degree $\leq \ell$ such that $f(0) = s$. Choose distinct $\alpha_i \in \mathbb{F}_q$ and set $s_i = (f(\alpha_i), \alpha_i)$.

1. Show that the properties 1 and 2 hold for this scheme.

   A: *Let $A \subseteq [1, n]$ with $A \geq \ell + 1$. Since f is of degree at most ℓ, we can use polynomial interpolation to recover $f(X)$ and hence $s = f(0)$ from the shares $\{(f(\alpha_i), \alpha_i)\}_{i \in A}$. Now suppose $|A| \leq \ell$ or wlg. consider $|A| = \ell$. The we have evaluations of f at ℓ points and f has $\ell+1$ coefficients. For every fixed value of $f(0)$, we get a different $f(X)$ by polynomial interpolation using the ℓ shares. Hence every possible value of s is equally likely. Hence both properties 1 and 2 hold.*

Linear codes and secret sharing: Consider $\mathcal{X} = \mathbb{F}_q$ with $q \geq n$. Let C be an $[n+1, k, d]_q$-code and $C^\perp$ be its dual $[n+1, n+1-k, d^\perp]_q$-code. Consider the following secret sharing scheme: pick a random codeword $\mathbf{c} = (c_0, c_1, \ldots, c_n) \in C$ and set $s = c_0$ and $s_i = c_i$ for $i \in [1, n]$.

1. Argue that the scheme is correct (that is, any $s \in \mathbb{F}_q$ corresponds to some codeword).

   A: *If the generator matrix of C consists of an all-zero column then we can omit it and obtain a code with shorter block length. So, assume that there exists a codeword $\mathbf{c}$ with $c_0 \neq 0$. Then $\alpha\mathbf{c} \in C$ for all $\alpha \in \mathbb{F}_q$. Therefore, for every $\alpha \in \mathbb{F}_q$, there is a codeword $\mathbf{c}$ with $c_0 = \alpha$.*

2. Show that it is an (ℓ, m)-secret sharing scheme with $\ell \leq d^\perp - 2$ and $m \geq n - d + 2$.

   A: *Suppose we know $m \geq n - d + 2$ symbols of the codeword. Then $n + 1 - (n - d + 2) = d - 1$ symbols are unknown. Treating these as erasures, we can determine the codeword uniquely and hence the secret. Recall that a linear code with minimum distance d can correct $d - 1$ erasures.*

   To show the bound on ℓ, suppose we are given $\leq d^\perp - 2$ symbols of a codeword $(c_0, c_1, \ldots, c_n) \in C$ at positions determined by a set $A \subseteq [1, n]$. What we know about the c_i's is a set of linear constraints $\sum_{i=0}^{n} y_i c_i = 0$ for all codewords $(y_0, \ldots, y_n)$ in $C^\perp$. In order to reconstruct c_0, we need a constraint with $y_0 \neq 0$ and $y_i = 0$ for every $i \in A$. But a codeword in $C^\perp$ representing such a constraint would have weight $\leq d^\perp - 2 + 1 = d^\perp - 1$ contradicting the fact that $C^\perp$ has minimum distance $d^\perp$. What remains is to show that the secret can take every value in $\mathbb{F}_q$ with the same probability. Any $d^\perp - 1$ or fewer columns of the generator matrix of C are linearly independent (this is because $d^\perp$ is the minimum number of dependent columns in the parity check matrix of $C^\perp$ which is nothing but the generator matrix of C). So c_0 can take any value.

Correspondence to Reed-Solomon?

1. Show that $RS[n, k]^\perp = RS[n, n-k]$.

   A: *In the previous tutorial we have seen that if $G \in \mathbb{F}_q^{n \times k}$ generates $RS[n, k]$ and $H \in \mathbb{F}_q^{n \times (n-k)}$ generates $RS[n, n-k]$ then $G^T H = 0$. So H is a parity check matrix of $C = RS[n, k]$ and hence $C^\perp = RS[n, n-k]$. We only need to show that the minimum distance of the code is exactly $k+1$. Suppose there exists a codeword $\mathbf{c} \in C^\perp$ with weight $\leq k$. Then at least $n - k$ components of $\mathbf{c}$ are zero. Consider the $(n-k) \times (n-k)$ matrix formed by the rows of H corresponding to these $n-k$ positions. This matrix has rank $< n - k$ i.e., $n - k$ rows of H are linearly dependent. So there exists some non-trivial linear combination of the $n - k$ rows of H that sums to zero. The coefficient vector has at least k zero's (corresponding to the other k rows) and must correspond to a valid codeword generated by G, thus contradicting the fact that C has minimum distance $n - k + 1$. Therefore, $C^\perp$ must have minimum distance $k + 1$.*

2. Can you represent Shamir's $(\ell, \ell+1)$-scheme as a linear code-based scheme with $C = RS[n', k']_q$ for some n', k'?

A: *The number of players is n in Shamir's scheme and so we should consider a code with block length $n' = n+1$. $RS[n+1, \ell+1]$ has minimum distance $d = n-\ell+1$ so that $m \geq (n+1)-d+2 = \ell+1$ and so choose $k' = \ell+1$. From the previous question, we have $RS[n+1, \ell+1]^{\perp} = RS[n+1, n-\ell]$ which has distance $d^{\perp} = \ell+2$ so that $\ell \leq d^{\perp} - 2$.*