Homework 1 Solutions

E2-210, Jan–Apr 2025

1. Given a received word $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, let $J = \{i : y_i = ?\}$ denote the set of erased positions in $\mathbf{y}$. For any $\mathbf{c} = (c_1, c_2, \ldots, c_n) \in \mathcal{C}$, let $\mathbf{c}_{\sim J}$ denote the vector $(c_j : j \in J^c)$ obtained from $\mathbf{c}$ by deleting all the coordinates in J. Let $\mathcal{C}' = \{\mathbf{c}_{\sim J} : \mathbf{c} \in \mathcal{C}\}$. This is the code obtained from $\mathcal{C}$ by deleting all the coordinates in J. Note that if $\mathcal{C}$ has minimum distance d, then $\mathcal{C}'$ has minimum distance at least $d - |J|$.

   Now, consider a decoder $\mathcal{D}$ that operates on a received word $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ as follows:

   Step 1: Determine the set of erased positions $J = \{i : y_i = ?\}$. Delete all the erased positions from $\mathbf{y}$ to obtain $\mathbf{y}'$.

   Step 2: Apply minimum distance decoding to $\mathbf{y}'$ to find the closest (in Hamming distance) vector $\mathbf{c}' \in \mathcal{C}'$.

   Step 3: Extend $\mathbf{c}'$ to a word $\widetilde{\mathbf{y}}$ of length n by inserting ?s in the coordinates in J.

   Step 4: Apply erasure decoding to $\widetilde{\mathbf{y}}$, i.e., identify (if possible) the unique codeword in $\mathcal{C}$ that agrees with $\widetilde{\mathbf{y}}$ in all the unerased positions.

   To see why the above decoder would work, assume that $\mathbf{y}$ is obtained from a codeword $\mathbf{c} \in \mathcal{C}$ after at most s erasures and at most t errors, where $s + 2t \leq d - 1$. Then, $|J| \leq s$, and hence, $\mathcal{C}'$ has minimum distance at least $d - s$. Since $2t + 1 \leq d - s$, in Step 2, minimum distance decoding would uniquely recover $\mathbf{c}' = \mathbf{c}_{\sim J}$. Extend $\mathbf{c}'$ to $\widetilde{\mathbf{y}}$ as in Step 3. Now, since $\rho \leq d - 1$, the original codeword $\mathbf{c}$ would be the unique codeword in $\mathcal{C}$ that agrees with $\widetilde{\mathbf{y}}$ in all unerased coordinates, and Step 4 will recover $\mathbf{c}$.

2. $\mathcal{C}_{\text{ISBN}} = \left\{(c_1, c_2, \ldots, c_{10}) \in \mathbb{F}_{11}^{10} : \sum_{i=1}^{10} (11 - i) \cdot c_i \equiv 0 \pmod{11}\right\}$.

   (a) From the definition, it is clear that a parity-check matrix for $\mathcal{C}$ is:

   $$H = [10 \;\; 9 \;\; 8 \;\; 7 \;\; 6 \;\; 5 \;\; 4 \;\; 3 \;\; 2 \;\; 1].$$

   (b) By the rank-nullity theorem, $\dim(\mathcal{C}_{\text{ISBN}}) = n - \text{rank}(H) = 10 - 1 = 9$.

   Since H has no 0 column, $\mathcal{C}_{\text{ISBN}}$ has no codewords of weight 1, and hence, $d_{\min}(\mathcal{C}_{\text{ISBN}}) \geq 2$. On the other hand, H has rank 1, and so any two columns of H must be linearly dependent. Therefore, $\mathcal{C}_{\text{ISBN}}$ has codewords of weight equal to 2, which proves that $d_{\min}(\mathcal{C}_{\text{ISBN}}) = 2$.

   (c) The parity-check matrix H above is in the form $[A \mid I_{n-k}]$. Hence, $G = [I_k \mid -A^T]$ is a systematic

generator matrix for $\mathcal{C}_{\text{ISBN}}$:

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 9 \end{bmatrix}$$

3. Observe that $\mathcal{C}^{\#}_{\text{ISBN}} = \text{nullspace}(H^{\#})$, where

$$H^{\#} = \begin{bmatrix} 10 & 9 & 8 & 7 & 6 & 5 & 4 & 3 & 2 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

   (a) Clearly, $\text{rank}(H^{\#}) = 2$, hence $\dim(\mathcal{C}^{\#}_{\text{ISBN}}) = 10 - 2 = 8$ by the rank-nullity theorem.

   For the minimum distance, one can argue (say, by examining determinants of 2×2 submatrices) that any two columns of $H^{\#}$ are linearly independent over $\mathbb{F}_{11}$. Hence, $d_{\min}(\mathcal{C}^{\#}_{\text{ISBN}}) \geq 3$. But, since $H^{\#}$ has rank 2, any three columns of H are linearly dependent. From this, we conclude that $d_{\min}(\mathcal{C}^{\#}_{\text{ISBN}}) = 3$. Thus, $\mathcal{C}^{\#}_{\text{ISBN}}$ is a $[10, 8, 3]$ linear code over $\mathbb{F}_{11}$.

   (b) Since the minimum distance is 3, the code is single-error-correcting.

4. The parity-check matrix $H^{\#}$ above gives rise to $q^{n-k} = 11^2 = 121$ distinct syndromes. So, the code $\mathcal{C}^{\#}_{\text{ISBN}}$ can correct at most 120 non-zero error vectors via syndrome decoding. As we shall see, there are far more than 120 error vectors obtainable as $\mathbf{e} = \mathbf{y} - \mathbf{c}$, where $\mathbf{c} \in \mathcal{C}^{\#}_{\text{ISBN}}$ is a codeword, and $\mathbf{y}$ is obtained by interchanging some two digits of $\mathbf{c}$. Indeed, any such error pattern $\mathbf{e}$ is of the form $[0 \ldots 0, c_i - c_j, 0, \ldots, 0, c_j - c_i, 0, \ldots 0] = (c_i - c_j)[0 \ldots 0, 1, 0, \ldots, 0, -1, 0, \ldots 0]$, where c_i and c_j ($c_i \neq c_j$) are symbols that appear in the i-th and j-th coordinates (with $i < j$) of some $\mathbf{c} \in \mathcal{C}^{\#}_{\text{ISBN}}$.

   We claim that $c_i - c_j$ can take any non-zero value in $\mathbb{F}_{11}$. Indeed, for any pair of distinct $\alpha, \beta \in \mathbb{F}_{11}$, and any pair of distinct coordinate indices i, j, there is a codeword $\mathbf{c} \in \mathcal{C}^{\#}_{\text{ISBN}}$ such that $c_i = \alpha$ and $c_j = \beta$. This is because even if c_i and c_j are pre-specified, the remaining coordinates c_m ($m \neq i, j$) can always be chosen so that $H\mathbf{c}^T = 0$.

   Thus, single transposition errors arise from error patterns of the form $\gamma[0 \ldots 0, 1, 0, \ldots, 0, -1, 0, \ldots 0]$, for $\gamma \in \mathbb{F}_{11} \setminus \{0\}$, with the 1 occurring in coordinate i, and the -1 in coordinate j, with $1 \leq i < j \leq 10$. The total number of such error patterns is thus $\binom{10}{2} \times |\mathbb{F}_{11} \setminus \{0\}| = 450$. This is much larger than the number of non-zero syndromes available. We thus conclude that the code $\mathcal{C}^{\#}_{\text{ISBN}}$, under syndrome decoding, does **not** meet the requirements of the enhanced ISBN scheme.