Handout 28 Information Theory and Coding
Solutions to homework 11 December 20, 2007

PROBLEM 1. Recall that the minimum distance is also given by the weight of the minimum weight codeword (see last problem of Homework 10). Now observe that there exists a codeword x of weight w iff $Hx^T = 0$ where H is the parity-check matrix with n columns. This is equivalent to say that some w columns of H are linearly dependent. We then know that there exist d column that are linearly dependent. However no combination of $d-1$ columns or less are dependent since this case would give rise to a codeword of weight less or equal than $d-1$. This concludes the proof.

PROBLEM 2.

(a) We know that the minimum distance of a linear code is the same as the minimum weight of the $2^L - 1$ non-zero codewords. Since the minimum of a set of number is at most their average, the average weight of the non-zero codewords is an upper bound on the minimum distance. To compute the average weight of the non-zero codewords, we would find the total number of 1's in the non-zero codewords, and divide by the number of non-zero codewords. But from (2c) we know that the total number of 1's is at most $\frac{1}{2}N2^L$. This then yields that the average weight is at most $(N/2)2^L/(2^L-1)$ which is the required bound.

The proof that this bound is valid for all (not necessarily linear) codes is a bit more complicated: Consider a code C with M codewords with minimum distance $d_{\min}$. Since $d_{\min} \leq d(\mathbf{x}, \mathbf{y})$ for all codewords $\mathbf{x}$ and $\mathbf{y}$ such that $\mathbf{x} \neq \mathbf{y}$, we conclude that

$$M(M-1)d_{\min} \leq \sum_{\mathbf{x}\in C} \sum_{\substack{\mathbf{y}\in C \\ \mathbf{y}\neq\mathbf{x}}} d(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{x}\in C} \sum_{\mathbf{y}\in C} d(\mathbf{x}, \mathbf{y})$$

Now, $d(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{N} d(x_i, y_i)$ where x_i and y_i are the ith digit of $\mathbf{x}$ and $\mathbf{y}$ and $d(x_i, y_i)$ is equal to 1 if x_i and y_i are different and zero otherwise. Thus,

$$M(M-1)d_{\min} \leq \sum_{i=1}^{N} \sum_{\mathbf{x}\in C} \sum_{\mathbf{y}\in C} d(x_i, y_i).$$

Suppose out of the M codewords Z_i of them have a 0 as their ith digit. Let S_i denote these codewords. Then, necessarily $M - Z_i$ of the codewords will have a 1 as their ith digit. Note that

$$\begin{aligned}
\sum_{\mathbf{x}\in C} \sum_{\mathbf{y}\in C} d(x_i, y_i) &= \sum_{\mathbf{x}\in S_i} \sum_{\mathbf{y}\in C} d(x_i, y_i) + \sum_{\mathbf{x}\in S_i^c} \sum_{\mathbf{y}\in C} d(x_i, y_i) \\
&= \sum_{\mathbf{x}\in S_i} \sum_{\mathbf{y}\in S_i^c} 1 + \sum_{\mathbf{x}\in S_i^c} \sum_{\mathbf{y}\in S_i} 1 \\
&= Z_i(M - Z_i) + (M - Z_i)Z_i \\
&= 2Z_i(M - Z_i).
\end{aligned}$$

Note that $\max_z 2z(M-z) = M^2/2$, and thus $\sum_{\mathbf{x}\in C}\sum_{\mathbf{y}\in C} d(x_i, y_i) \le M^2/2$. Thus,

$$M(M-1)d_{\min} \le NM^2/2,$$

yielding $d_{\min} \le (N/2)M/(M-1)$. With $M = 2^L$ we obtain the same bound as in the linear case.

(b) Assume that the code is systematic, namely, each codeword consists of L information bits followed by $N-L$ check bits. Let us look at the subset of codewords which start with $L-j$ 0s. Since the rest of the j information bits are arbitrary this subset contains 2^j elements. Since the minimum distance within any subset of the codewords is an upper bound to the minimum distance of the whole code, let us apply the bound in part (a) to this subset. Notice that the codewords within the subset agree in the first $L-j$ positions, so the distance between any two of them will not change if we simply remove these bits. This yields a code of blocklength $N-(L-j)$ with 2^j codewords, and applying the bound in part (a) gives the desired result.

In the general case (when the code is not linear), classify the $M = 2^L$ codewords grouping them according to their initial $L-j$ bits. Since there are only 2^{L-j} groups, one of the groups will contain more than $M/2^{L-j} = 2^j$ codewords. These codewords, by construction agree on their initial $L-j$ bits, so the minimum distance within this group will not change if we remove these bits. Applying the bound in part (a) to this subset of codewords we again obtain the desired bound.

(c) First, rewrite the bound in part (b) as

$$N - L \ge 2d_{\min}\frac{2^j-1}{2^j} - j = 2d_{\min} - j - \frac{2d_{\min}}{2^j}$$

Taking $j = 1 + \lfloor \log_2 d_{\min} \rfloor$, we see that

$$N - L \ge 2d_{\min} - 1 - \lfloor \log_2 d_{\min} \rfloor - \frac{2d_{\min}}{2^j}$$

Since $N-L$ is an integer, we can improve the lower bound by rounding it up to the nearest integer:

$$N - L \ge 2d_{\min} - 1 - \lfloor \log_2 d_{\min} \rfloor - \left\lfloor \frac{2d_{\min}}{2^j} \right\rfloor.$$

But, $\log_2 d_{\min} < j \le 1 + \log_2 d_{\min}$, and thus $d_{\min} < 2^j \le 2d_{min}$. So,

$$2 > \frac{2d_{\min}}{2^j} \ge 1,$$

and $\lfloor 2d_{\min}/2^j \rfloor = 1$. Thus, we obtain the desired result,

$$N - L \ge 2d_{\min} - 2 - \lfloor \log_2 d_{\min} \rfloor.$$

Note that for $d \ge 3$ this is a tighter bound than the Singleton bound we proved in class $N - L \ge d_{\min} - 1$.

Problem 3.

(a) At the first step, we can choose any non-zero column vector with r coordinates. This will be the first column of our $r \times n$ parity-check matrix. Now suppose we have chosen i columns so that no $d-1$ are linearly dependent. They are all non-zero columns. There are at most

$$\binom{i}{1} + \cdots + \binom{i}{d-2}$$

distinct linear combinations of these i columns taken $2-2$ or fewer at a time.

(b) The total number of r-tuples (include the all-zero one) is 2^r. We can then choose a new column different from the previous ones, linearly independent from the previous ones, and keep the property that every $d-1$ columns are independent.

(c) We can iterate the procedure and we keep doing so as long as

$$1 + \binom{i}{1} + \cdots \binom{i}{d-2} < 2^r$$

where the first term counts the all-zero vector. At the last step, we can do so iff

$$1 + \binom{n-1}{1} + \cdots \binom{n-1}{d-2} < 2^r.$$

(d) Multiply both sides of the previous inequality by $M = 2^k$ gives the result since $r = n - k$.