Coding Theory	CMU: Fall 2014

Notes 14: Existence of good binary linear codes for list decoding

October 2014

Lecturer: Venkatesan Guruswami *Scribe: Venkat Guruswami*

In these notes, we prove the following theorem on the existence of binary linear codes for list decoding, which we covered in class. The proof is from [1, Section IV].

Theorem. *Fix $p \in (0, 1/2)$ and an integer $c \geq 1$. Then, for all large enough n, there is an $[n, k]_2$ binary linear code $\mathbf{C}$ with $k = \lfloor (1 - h(p) - 1/c)n \rfloor$ that is (p, c)-list decodable (meaning that for all $y \in \{0, 1\}^n$, $|B(y, pn) \cap C| \leq c$).*

Proof: For each fixed integer $c \geq 1$ and $0 < p < 1/2$, we use the probabilistic method to guarantee the existence of a binary linear code $\mathbf{C}$ of blocklength n, with at most c codewords in any ball of radius $e = pn$, and whose dimension is $k = \lfloor (1 - h(p) - 1/c)n \rfloor$, for all large enough n. This clearly implies the lower bound on U_c^{const} claimed in the statement of the Theorem.

The code $\mathbf{C} = C_k$ will be built iteratively in k steps by randomly picking the k basis vectors in turn. Initially the code C_0 will just consist of the all-zeroes codeword $b_0 = 0^n$. The code C_i, $1 \leq i \leq k$, will be successively built by picking a random (non-zero) basis vector b_i that is linearly independent of $b_1, \ldots, b_{i-1}$, and setting $C_i = \text{span}(b_1, \ldots, b_i)$. Thus $\mathbf{C} = C_k$ is an $[n, k]_2$ linear code. We will now analyze the list of c decoding radius of the codes C_i, and the goal is to prove that the list of c decoding radius of $\mathbf{C}$ is at least e.

The key to analyzing the list of c decoding radius is the following potential function S_C defined for a code C of blocklength n:

$$S_C = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} 2^{\frac{n}{c} \cdot |B(x,e) \cap C|} . \tag{1}$$

For notational convenience, we denote S_{C_i} be S_i. Also denote by T_x^i the quantity $|B(x, e) \cap C_i|$, so that $S_i = 2^{-n} \sum_x 2^{nT_x^i/c}$.

Let $B = |B(0, e)| = |B(0, pn)|$; then $B \leq 2^{h(p)n}$ where $h(p)$ is the binary entropy function of p. Clearly

$$S_0 = 1 - B/2^n + B2^{n/c}/2^n \leq 1 + 2^{n\left(h(p)-1+1/c\right)} . \tag{2}$$

Now once C_i has been picked with the potential function S_i taking on some value, say $\hat{S}_i$, the potential function S_{i+1} for $C_{i+1} = \text{span}(C_i \cup \{b_{i+1}\})$ is a random variable depending upon the choice of b_{i+1}. We consider the expectation $\mathbf{E}[S_{i+1} | S_i = \hat{S}_i]$ taken over the random choice of b_{i+1} chosen uniformly from outside $\text{span}(b_1, \ldots, b_i)$.

$$\begin{aligned}
\mathbf{E}[S_{i+1}] &= 2^{-n} \sum_x \mathbf{E}[2^{n/c \cdot T_x^{i+1}}] \\
&= 2^{-n} \sum_x \mathbf{E}[2^{n/c \cdot \left(|B(x,e) \cap C_i| + |B(x,e) \cap (C_i + b_{i+1})|\right)}] \\
&= 2^{-n} \sum_x \left(2^{n/c \cdot T_x^i} \mathop{\mathbf{E}}_{b_{i+1}} [2^{n/c \cdot T_{x+b_{i+1}}^i}] \right) \qquad (3)
\end{aligned}$$

where in the second and third steps we used the fact that if $z \in B(x,e) \cap C_{i+1}$, then either $z \in B(x,e) \cap C_i$, or $z + b_{i+1} \in B(x,e) \cap C_i$. To estimate the quantity (3), first note that if we did not have the condition that b_{i+1} was chosen from outside $\mathrm{span}(b_1, \ldots, b_i)$ (3) would simply equal $\hat{S}_i^2$. This follows from the fact that x and $x + b_{i+1}$ are independent and the definition of $\hat{S}_i$. Now we use the simple fact that the expectation of a positive random variable taken over b_{i+1} chosen randomly from outside $\mathrm{span}(b_1, \ldots, b_i)$ is at most $(1 - 2^{i-n})^{-1}$ times the expectation taken over b_{i+1} chosen uniformly at random from $\{0,1\}^n$. Hence, we get that

$$\mathbf{E}[S_{i+1}] \leq \frac{\hat{S}_i^2}{(1 - 2^{i-n})}. \tag{4}$$

Applying (4) repeatedly for $i = 0, 1, \ldots, k-1$, we conclude that there exists an $[n,k]$ binary linear code $\mathbf{C}$ with

$$\begin{aligned} S_{\mathbf{C}} = S_k &\leq \frac{S_0^{2^k}}{\prod_{i=0}^{k-1} (1 - 2^{i-n})^{2^{k-i}}} \\ &\leq \frac{S_0^{2^k}}{(1 - 2^{k-n})^k} \leq \frac{S_0^{2^k}}{1 - k 2^{k-n}} \end{aligned} \tag{5}$$

since $(1-x)^a \geq 1 - ax$ for $x, a \geq 0$. Combining (5) with (2), we have

$$S_k \leq (1 - k2^{k-n})^{-1} \left(1 + 2^{n(h(p) - 1 + 1/c)}\right)^{2^k}$$

and using $(1+x)^a \leq (1 + 2ax)$ for $ax \ll 1$, this gives

$$S_k \leq 2(1 + 2 \cdot 2^{k + (h(p) - 1 + 1/c)n}) \leq 6, \tag{6}$$

where the last inequality follows since $k = \lfloor (1 - h(p) - 1/c) n \rfloor$. By the definition of the potential S_k (1), this implies that

$$2^{n/c \cdot |B(x,e) \cap \mathbf{C}|} \leq 6 \cdot 2^n < 2^{n+3},$$

or

$$|B(x,e) \cap \mathbf{C}| \leq (1 + \frac{3}{n}) c$$

for every $x \in \{0,1\}^n$. If $n > 3c$, this implies $|B(x,e) \cap \mathbf{C}| < c + 1$ for every x, implying that the list of c decoding radius of $\mathbf{C}$ is at least e, as desired. □

References

[1] Venkatesan Guruswami, Johan Håstad, Madhu Sudan, and David Zuckerman. Combinatorial bounds for list decoding. *IEEE Transactions on Information Theory*, 48(5):1021–1035, 2002.