ENEE 739C: Advanced Topics in Signal Processing: Coding Theory
Instructor: Alexander Barg

Lecture 4 (**draft;** 9/3/03). Typical-sets decoding and the Shannon coding theorem

http://www.enee.umd.edu/~abarg/ENEE739C/course.html

Our goal is to derive an upper bound on the minimum attainable error probability of ML decoding for binary codes used over a BSC(p). This bound, as most existence results of coding theory, is obtained via averaging over code ensembles.

Together with ML decoding we will study another decoding mapping, *typical-sets decoding*. The idea is as follows: with probability tending to one, the weight of the error vector created by the channel BSC(p) will be close to pn. So it is natural to look for codevectors which are at distance about pn from the received vector $\mathbf{y}$. Let

$$T = \{\mathbf{x} \in \mathscr{H}_2^n : |\operatorname{wt}(\mathbf{x}) - np| \le \beta\sqrt{np(1-p)}\},$$

where $\beta = o(n^{1/2}), \beta \to \infty$ is an arbitrarily slowly increasing function of n.

Typical-sets decoder. If there exists a unique codevector $\mathbf{c}$ such that $\mathbf{y} - \mathbf{c} \in T$, decode $\mathbf{y}$ to $\mathbf{c}$, otherwise output an arbitrary codevector (random, say).

The interest in typical-sets decoding arises when we want to establish capacity results (Direct coding theorems) without caring about the decrease rate of the error probability. This is the approach taken in the book of J. Wolfowitz [3] (see also [2, §2.1], [1]). Another reason, made clear by Lemmas 1-2, is that it provides an easy, though sometimes coarse tool of analyzing thresholds of particular code families on the BSC and related channels.

Let us analyze error rate of a linear code with weight distribution $A_w, w = 0, d, \dots, n$. We assume that the all-zero vector was transmitted over the channel and denote by $\mathcal{E}$ the event that the decoder outputs a codeword of weight different from zero and by $\mathcal{E}_w$ the event that it outputs a codeword of some fixed weight $w > 0$. Our goal will be to establish conditions for the error probability P_e under typical-sets decoding to decrease with n. This probability can be written as follows:

$$P_e \le \Pr[\mathcal{E}, \mathbf{y} \in T] + \Pr[\mathbf{y} \notin T]. \tag{1}$$

By the de Moivre–Laplace theorem[1]

$$\Pr[y \notin T] = \Pr[|\operatorname{wt}(\mathbf{y}) - np| > \beta\sqrt{np(1-p)}] \le 2e^{-\beta^2/2}$$

which tends to zero by our choice of β. Therefore from now on we forget about the second term in the bound (1) and concentrate on the first one. Note that for $y \in T$ the difference $|\operatorname{wt}(\mathbf{y}) - pn| = o(n)$, so it will not affect the main term of our estimate to replace the condition $\mathbf{y} \in T$ with $\operatorname{wt}(\mathbf{y}) = pn$. Denote by $\mathcal{E}_w(\mathbf{c})$ the event that $\mathbf{y}$ is decoded incorrectly to a given codeword c of weight w. Then for the first term of the bound (1) we have

$$\Pr[\mathcal{E}, \mathbf{y} \in T] = \sum_{w=d}^{2(np+\beta\sqrt{np(1-p)})} A_w \Pr[\mathcal{E}_w(\mathbf{c}) | \mathbf{y} \in T] \Pr[\mathbf{y} \in T].$$

First observe that

$$\Pr[\mathcal{E}_w(\mathbf{c}) | \mathbf{y} \in T] \lesssim \binom{n}{pn}^{-1} \sum_{i=w/2}^{pn} \binom{w}{i}\binom{n-w}{pn-i}.$$

The vector $\mathbf{y}$ is chosen randomly with $p(0) = 1-p, p(1) = p$, so the unconstrained maximum of the summation term $\binom{w}{i}\binom{n-w}{pn-i}$ is attained when $i = wp < w/2$. Hence we write

$$\Pr[\mathcal{E}_w(\mathbf{c}) | \mathbf{y} \in T] \lesssim n\binom{n}{pn}^{-1} \binom{w}{w/2}\binom{n-w}{pn-w/2} \tag{2}$$

[1] W. Feller, An Introduction to Probability Theory and Its Applications (1970), Vol. 1, §§VII.6,VII.7.

and, bounding $P[\mathbf{y} \in T]$ above by one,

$$\Pr[\mathcal{E}, \mathbf{y} \in T] \le n \binom{n}{pn}^{-1} \sum_{w=d}^{2np} A_w \binom{w}{w/2} \binom{n-w}{pn-w/2}. \tag{3}$$

Let $\Pr[\mathcal{E}, \mathbf{y} \in T] = 2^{-nE(1-o(1))}$, $w = \omega n$, then

$$E \ge \max_{\delta \le \omega \le 2p} h_2(p) - a_\omega(\mathcal{C}) - \omega - (1-\omega) h_2\Big(\frac{p - \omega/2}{1-\omega}\Big). \tag{4}$$

We would like to obtain a sufficient condition for the error probability $\Pr[\mathcal{E}, \mathbf{y} \in T]$, and thus for P_e to go to zero as $n \to \infty$. For a fixed channel we have

Lemma 1. *The error probability $P_e(\mathcal{C})$ of a code $\mathcal{C}$ with relative distance δ under typical-sets decoding tends to 0 if the weight profile $a_\omega(\mathcal{C})$ satisfies*

$$a_\omega < S(p, \omega) := h_2(p) - \omega - (1-\omega) h_2\Big(\frac{p - \omega/2}{1-\omega}\Big) \qquad (\delta \le \omega \le 2p).$$

For a fixed code we have

Lemma 2. *Let $\mathcal{C}$ be a code with relative distance δ and weight profile a_ω. Then the error probability $P_e(\mathcal{C})$ under typical-sets decoding tends to 0 for all p such that*

$$S(p, \omega) - a_\omega > 0 \qquad (\delta \le \omega \le 2p).$$

The importance of these results lies in the fact that they give estimates on the threshold weight distribution for codes on a BSC with transition probability p or on the threshold probability for max-likelihood decoding of a given code used over the BSC (recall the definition in Lecture 3).

Let us provide some intuition on the second of the two items, the threshold probability θ of a code $\mathcal{C}$ with the weight distribution $A_0, A_1, \dots, A_n$. Since the errors typically take the transmitted vector $\mathbf{c}$ to the sphere of radius pn around it, the error probability will likely go to zero if the fraction of points on that sphere which are closer to some other codewords than to $\mathbf{c}$ is close to zero. If this is the case, then the threshold probability $\theta(\mathcal{C}) > p$. This informal argument is made rigorous by Lemma 1.

Finally, taking $\mathcal{C}$ in the above argument to be a random linear code, we obtain a direct coding theorem for the BSC (a Shannon theorem). Before stating it, let us prove one technical result.

Lemma 3. (A COVERING LEMMA) *Let*

$$N(w, r) = \sum_{\mathbf{c} \in \mathcal{S}_w} \big|\{\mathbf{y} \in \mathcal{S}_r : \mathrm{d}(\mathbf{y}, \mathbf{c}) \le r\}\big|.$$

Then for $n \to \infty$, $r = \rho n, 0 < \rho < 1/2 :$ $N(w, r) \lesssim \binom{n}{r}^2$ with equality

$$N(w, r) \cong \binom{n}{r}^2$$

only for $\frac{w}{n} \sim 2\rho(1-\rho)$.

Proof :

$$\begin{aligned} N(w, r) &= \binom{n}{w} \sum_{i=w/2}^{r} \binom{w}{i} \binom{n-w}{r-i} \\ &\cong \binom{n}{w} \binom{w}{w/2} \binom{n-w}{r-w/2} \quad \text{(max on } i \text{ is attained for } i \approx w\rho \le w/2) \\ &= \binom{n}{r} \binom{n-r}{w/2} \binom{r}{w/2} \quad \text{(rewrite the multinomial coefficient;} \\ &\qquad\qquad \text{or, verify that } \binom{n}{w} p^w_{r,r} = \binom{n}{r} p^r_{w,r}) \\ &\lesssim \binom{n}{r} \binom{n-r}{r(1-\rho)} \binom{r}{\rho(n-r)}. \end{aligned}$$

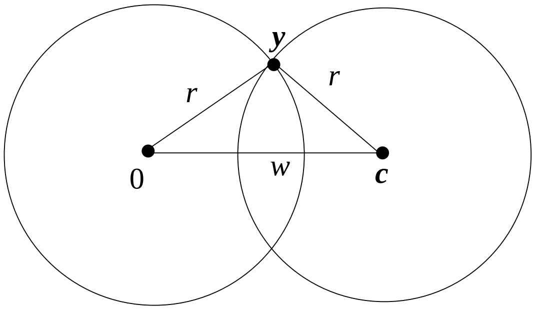

Figure 1. Intersecting spheres: $\mathcal{S}_r(0) \cap \mathcal{B}_r(\mathbf{c})$

Finding the maximum on w in the last step calls for some explanation. For a fixed vector y of weight r we are counting the number of weight-w vectors c with $\mathrm{d}(y,c) = r = \rho n$. This number is maximized if c is a typical vector obtained after n independent drawings from the binomial probability distribution given by $P[y+c=1] = \rho, P[y+c=0] = 1-\rho$. Hence the maximizing argument is given by $w/2 = r(1-\rho)$. Substituting it, we obtain

$$\begin{aligned} N(2r(1-\rho), r) &\cong \exp[n(h_2(\rho) + (1-\rho)h_2(1-\rho) + \rho h_2(1-\rho))] = \exp(2nh_2(\rho)) \\ &\cong \binom{n}{r}^2. \end{aligned}$$

▮

What to remember:

For large n, the maximum of the intersection volume of the ball $\mathcal{B}_r(\mathbf{c})$ and the sphere $\mathcal{S}_r(0)$ is attained for $\mathrm{wt}(\mathbf{c}) \sim 2r(1-\frac{r}{n})$.

Theorem 4. *The threshold probability for random codes of rate R satisfies $p_0 \geq \delta_{\mathrm{GV}}(R)$. Hence, the capacity $\mathscr{C}$ of the binary symmetric channel satisfies $\mathscr{C} \geq 1 - h_2(p)$.*

Proof : Take $\mathcal{C}$ a random linear code of rate R. We need to prove that $P[\mathcal{E}, y \in T] \to 0$ if $p < \delta_{\mathrm{GV}}(R)$. Let us use the Covering Lemma in (3):

$$P[\mathcal{E}, y \in T] \leq \binom{n}{pn}^{-1} 2^{-nh_2(\delta_{\mathrm{GV}}(R))} \sum_{w=d}^{2np} N(w,p) \cong \exp[-n(h_2(\delta_{\mathrm{GV}}(R) - h_2(p))].$$

▮

References

1. T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, New York e.a., 1991.
2. I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless channels*, Akadémiai Kiadó, Budapest, 1981.
3. J. Wolfowitz, *Coding theorems of information theory*, Springer-Verlag, Berlin e. a., 1964.