ENEE 739C: Advanced Topics in Signal Processing: Coding Theory
Instructor: Alexander Barg
Lecture 3 (**draft;** 9/1/03). Decoding of codes.

Our task of code construction is not complete if we cannot decode the codes we are studying. Decoding another major task in coding theory studied in a variety of scenarios.

Definition 1. *Given a code $\mathcal{C} \in \mathscr{H}_q^n$ and a point $\mathbf{y} \in \mathscr{H}_q^n$,* maximum likelihood, or complete decoding *takes $\mathbf{y}$ to (one of) the closest code vectors by the Hamming distance.*

This procedure may be actually called minimum distance decoding, the name "maximum likelihood" is justified in a short while. Intuitively is seems natural to assume that this decoding is the best we can hope for in terms of minimizing the error rate. Note if a vector $\mathbf{x} \in \mathcal{C}$ (say a binary code) is sent over a binary symmetric channel W, the probability that a vector $\mathbf{y}$ is received on its output $P(\mathbf{y}|\mathbf{x}) = W^n(\mathbf{yx}) = p^w(1-p)^{n-w}$ where $w = \mathrm{d}(\mathbf{x}, \mathbf{y})$. If $\mathbf{c}$ is the decoding result, the probability of a decoding error equals $1 - P(\mathbf{c}|\mathbf{y})$. Therefore, the mapping ψ that minimizes the probability of error for the codeword $\mathbf{c}_i$ is given by

$$\psi(\mathbf{y}) = \mathbf{c}_i \qquad \Longleftrightarrow \qquad P(\mathbf{c}_i|\mathbf{y}) > P(\mathbf{c}_j|\mathbf{y}) \quad \forall j \neq i.$$

Letting $P(\mathbf{y}) = \sum_{\mathbf{c}\in\mathcal{C}} W^n(\mathbf{y}|\mathbf{c})$ we have

$$W^n(\mathbf{y}|\mathbf{c}) = \frac{P(\mathbf{c}|\mathbf{y})P(\mathbf{y})}{P(\mathbf{c})}.$$

If we assume that $P(\mathbf{c}) = 1/M$ for every $\mathbf{c} \in \mathcal{C}$ then maximum likelihood decoding is defined equivalently as

$$\psi(\mathbf{y}) = \mathbf{c}_i \qquad \Longleftrightarrow \qquad W^n(\mathbf{y}|\mathbf{c}_i) > W^n(\mathbf{y}|\mathbf{c}_j) \quad \forall j \neq i.$$

Finally note that maximizing $W^n(\mathbf{y}|\mathbf{c})$ is equivalent to minimizing the distance $\mathrm{d}(\mathbf{y}, \mathbf{c})$, so minimum distance decoding is indeed max-likelihood or ML decoding.

We know from ENEE722 that max-likelihood decoding of a linear binary $[n, k, d]$ code $\mathcal{C}$ can be implemented by the standard array (or the syndrome trellis or by other equivalent means). The problem is that all these methods for codes of large length and rate not very close to 0 or 1 are very computationally involved. Therefore coding theory is also concerned with restricted decoding procedures. One of the most commonly used is *bounded distance decoding*. This name is used loosely for all procedures that find the closest codeword to the received vector in a sphere of some fixed radius t if this sphere contains codewords, and returns "erasure" if it does not. For $t = \lfloor (d-1)/2 \rfloor$ such a codeword (if it exists) is always unique; for greater t the sphere will sometimes contain more than one codeword. In the case of a multitude of codewords in the sphere it is sometimes desirable to output them all as a decoding result: this procedure is called *list decoding* and the decoding result a *list*.

Clearly, minimum distance decoding is the same as bounded distance decoding with $t = r(\mathcal{C})$, the covering radius of the code. In general, intuition about bounded distance decoding is as follows: shrinking the radius increases the probability of erasure and decreases the probability of undetected error. Another extreme (compared to minimum distance decoding, called (pure) error detection is considered toward the end of this lecture.

To describe ML decoding in geometric terms, define a *Voronoi region* $D(\mathbf{c}, \mathcal{C})$ of a code point $\mathbf{c} \in \mathcal{C}$ as follows:

$$D(\mathbf{c}, \mathcal{C}) = \{\mathbf{y} \in \mathscr{H}_q^n : \mathrm{d}(\mathbf{y}, \mathbf{c}) < \mathrm{d}(\mathbf{y}, \mathbf{c}') \ \forall \mathbf{c}' \in \mathcal{C} \backslash \{\mathbf{c}\}\}.$$

This defines a partition of $\mathscr{H}_q^n$ into decoding (Voronoi) regions of the code points (with ties broken arbitrarily). We have

$$P_e(\mathbf{c}) = \Pr[\mathscr{H}_q^n \backslash D(\mathbf{c}, \mathcal{C})]$$

where Pr is computed with respect to the binomial probability distribution defined by the channel and with "center" at $\mathbf{c}$.

What to remember:

ML decoding is optimal but generally infeasible (except for short codes, or codes with very few codewords, or codes of rate close to one).

Define the *average* and *maximum* error probability for the code $\mathcal{C}$ as

$$\begin{aligned} P_e(\mathcal{C}) &= \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}} P_e(\mathbf{c}) \\ P_{e,m}(\mathcal{C}) &= \max_{\mathbf{c}\in\mathcal{C}} P_e(\mathbf{c}). \end{aligned}$$

The quantity $P_e(\mathcal{C})$ is one of the main parameters of the code, with lots of attention devoted to it in the literature. For communication applications the error probability is more important than the minimum distance which does not provide much information about the decoding performance except for low noise.

For bounded distance decoding the probabilities of undetected error (miscorrection) and erasure can be defined analogously:

$$\begin{aligned} P_{e,t}(\mathcal{C}) &= \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}} \Pr\Big[\bigcup_{\mathbf{c}'\in\mathcal{C}\setminus\{\mathbf{c}\}} \mathcal{B}_t(\mathbf{c}')\big|\mathbf{c}\Big] \quad \text{(undetected error)} \\ P_{x,t}(\mathcal{C}) &= \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}} \Pr\Big[\mathscr{H}_q^n \setminus \bigcup_{\mathbf{c}'\in\mathcal{C}} \mathcal{B}_t(\mathbf{c}')\big|\mathbf{c}\Big] \quad \text{(erasure)} \end{aligned}$$

Remark. For a linear code $\mathcal{C}$ Voronoi regions of all code points are congruent (the set of correctable errors for any code point is the same), and so $P_e(\mathcal{C}) = P_e(\mathbf{c}), P_{e,t}(\mathcal{C}) = P_{e,t}(\mathbf{c}), \ldots$ for any code point $\mathbf{c}$.

Computing $P_e(\mathcal{C})$ generally is a difficult task. For instance, for a linear code $\mathcal{C}$

$$P_e(\mathcal{C}) = \sum_{\mathbf{x}\text{ a coset non-leader}} \Big(\frac{p}{q-1}\Big)^{\mathrm{wt}(\mathbf{x})}(1-p)^{n-\mathrm{wt}(\mathbf{x})}.$$

However for most codes no reasonable bounds for the weight distribution of coset leaders are known (a few exceptions, apart from the trivial cases like the Hamming code, are the duals of the 2-error-correcting BCH codes and a few related cases [1, 2]). Hence we are faced with a search for bounds on $P_e(\mathcal{C})$ from both sides. This is one of the main topics of coding theory with hundreds of papers devoted to it; this is also the main topic of such books as [4, 9, 3].

Bounds on the error probability $P_e(\mathcal{C})$ of complete decoding

Bounding $P_e(\mathcal{C})$ is a difficult task. Essentially the only technique available relies upon the distance distribution of the code $\mathcal{C}$; it is likely that this is the best way to estimate P_e in the general case. Traditionally, upper bounds on $P_e(\mathcal{C})$ receive more attention than lower bounds. This is justifiable because upper bounds give some level of confidence in estimating performance of communication systems.

It is important to realize right away that, with all the literature devoted to these bounds, the mainstream technique relies on a combination of several trivial observations presented in this section. One simple idea, the *union bound principle*, is to upper bound the error probability as follows:

$$P_e(\mathcal{C}) \le \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}}\sum_{\mathbf{c}'\in\mathcal{C}} P(\mathbf{c}\to\mathbf{c}'), \tag{1}$$

where

$$P(\mathbf{c}\to\mathbf{c}') := \Pr\{\mathbf{x}\in\mathscr{H}_q^n : \mathrm{d}(\mathbf{x},\mathbf{c}') \le \mathrm{d}(\mathbf{x},\mathbf{c})\}$$

is the probability of the half-space cut out by the median hyperplane between $\mathbf{c}$ and $\mathbf{c}'$. It is clear that this approach is generally not optimal because the probability of large parts of $\mathscr{H}_q^n$ is counted many times.

Theorem 1. *Let $\mathcal{C}\subset\mathscr{H}_q^n$ be a code with distance d used over the qSC with crossover probability $p/(q-1)$. Then the error probability of complete decoding*

$$P_e(\mathcal{C}) \le \sum_{w=d}^{n}\mathcal{B}_w \sum_{e=\lceil w/2\rceil}^{n}\pi(e)\sum_{s=0}^{e} p_{e,s}^{w}, \tag{2}$$

where $p_{e,s}^{w}$ is the intersection number of $\mathscr{H}_q^n$ and $\pi(e) = (\frac{p}{q-1})^e(1-p)^{n-e}$.

Proof : In our calculation we rely on (1). Let $\mathbf{c} \in \mathcal{C}$. We need to estimate $\Pr[\overline{D(\mathbf{c},\mathcal{C})}]$. Let $\mathbf{y} \in \overline{D(\mathbf{c},\mathcal{C})}$ and suppose that $\mathrm{d}(\mathbf{y},\mathbf{c}) = e$, then $s := \mathrm{d}(\mathbf{y},\mathbf{c}') \le e$ for some other $\mathbf{c}' \in \mathcal{C}$. Suppose that $\mathrm{d}(\mathbf{c},\mathbf{c}') = w$, then for a given $e \ge \lceil w/2 \rceil$ the number of such vectors $\mathbf{y}$ equals $\sum_{s=0}^{e} p_{e,s}^{w}$. The probability $P(\mathbf{y}|\mathbf{c})$ equals $\pi(e)$. We have

$$P(\overline{D(\mathbf{c},\mathcal{C})}) = \sum_{e=\lceil w/2 \rceil}^{n} \pi(e) \sum_{s=0}^{e} p_{e,s}^{w}.$$

Now to obtain (2) multiply this by the number $\mathcal{B}_w(\mathbf{c})$ of w-neighbors of $\mathbf{c}$ and average over $\mathbf{c}$ as in the proof of the bound on the probability of undetected error. ∎

Note that every error vector of weight greater than and covering radius of $\mathcal{C}$ is uncorrectable, so we can improve (2) as follows:

$$P_e(\mathcal{C}) \le \sum_{w=d}^{2t} \mathcal{B}_w \sum_{e=\lceil w/2 \rceil}^{t} \pi(e) \sum_{s=0}^{e} p_{e,s}^{w} + \sum_{e=t+1}^{n} \binom{n}{e} p^e (1-p)^{n-e}.$$

Generalizing, we obtain the following useful remark which goes back at least to Shannon [7]. It is valid for a BSC or any additive channel W.

Lemma 2. *Let* $\mathbf{e}$ *be the error vector and* $P(\mathbf{e} \in A)$ *be the probability that the error vector falls into some subset* $A \subset \mathscr{H}_q^n$. *Then*

$$P_e(\mathcal{C}) \le \min_{A \subset \mathscr{H}_q^n} [P_e(\mathcal{C}, \mathbf{e} \in A) + P(\mathbf{e} \notin A)] \tag{3}$$

Proof : Let $\mathcal{E}$ be the decoding error event and $\mathbf{c} \in \mathcal{C}$. Then

$$P_e(\mathbf{c}) = \Pr[\mathcal{E}, \mathbf{e} \in A|\mathbf{c}] + \Pr[\mathcal{E}|\mathbf{e} \notin A, \mathbf{c}] P(\mathbf{e} \notin A) \le \Pr[\mathcal{E}, \mathbf{e} \in A|\mathbf{c}] + P(\mathbf{e} \notin A).$$

Now average over all $\mathbf{c} \in \mathcal{C}$:

$$P_e(\mathcal{C}) = \frac{1}{M} \sum_{\mathbf{c}} \Pr[\mathcal{E}, \mathbf{e} \in A|\mathbf{c}] P(\mathbf{c}) + P(\mathbf{e} \notin A)$$

∎

Let us write out an explicit expression for $P(\mathbf{c}_1 \to \mathbf{c}_2)$.

Lemma 3. (i) *Let* $\mathbf{c}_1, \mathbf{c}_2 \in \mathscr{H}_2^n, \mathrm{d}(\mathbf{c}_1, \mathbf{c}_2) = w$. *Suppose that* $\mathbf{c}_1$ *is sent over a BSC*(p). *Then*

$$P(\mathbf{c}_1 \to \mathbf{c}_2) - \sum_{r=\lceil w/2 \rceil}^{n-\lfloor w/2 \rfloor} \sum_{i=\lceil w/2 \rceil}^{r} \binom{w}{i} \binom{n-w}{r-i} p^r (1-p)^{n-r} \tag{4}$$

Let $n \to \infty, w = \omega n$, *then*

$$P(\mathbf{c}_1 \to \mathbf{c}_2) \cong 2^{n\omega \log_2 2\sqrt{p(1-p)}}. \tag{5}$$

(ii) *One useful property of this exponent is that if*

$$\frac{2\sqrt{p(1-p)}}{1+2\sqrt{p(1-p)}} \le \omega \le \frac{1}{2}$$

then $-\omega \log_2 2\sqrt{p(1-p)} \le D(\omega \| p)$.

Proof : Formula (4) is obtained by simply writing out explicitly definition (2). Note that the upper summation limit $n - \lfloor w/2 \rfloor$ for r equals the covering radius of the code $\{\mathbf{c}_1, \mathbf{c}_2\}$. Now let us use Lemma 2 to rewrite (4) as follows:

$$\begin{aligned} P(\mathbf{c}_1 \to \mathbf{c}_2) &\le \sum_{r=\lceil w/2 \rceil}^{\lfloor n/2 \rfloor} \sum_{i=\lceil w/2 \rceil}^{r} \binom{w}{i} \binom{n-w}{r-i} p^r (1-p)^{n-r} + \sum_{r=\lfloor n/2 \rfloor+1}^{n} \binom{n}{r} p^r (1-p)^{n-r} \\ &:= P_1 + P_2. \end{aligned}$$

We have $P_2 \cong \exp(-nD(\frac{1}{2}\|p)) = \exp(n \log_2 2\sqrt{p(1-p)}\,)$.

Let $\rho = r/n \le 1/2$. The maximum on i of the product $\binom{w}{i}\binom{n-w}{r-i}$ is attained when $i \approx \rho w \le w/2$; hence,

$$P_1 \cong \sum_{r=\lceil w/2 \rceil}^{\lfloor n/2 \rfloor} \binom{w}{w/2}\binom{n-w}{r-w/2} p^r (1-p)^{n-r}.$$

Taking logarithms, we obtain

$$P_1 \cong \max_{\omega/2 \le \rho \le \omega} \exp\left[n(\omega + (1-\omega)h_2(\frac{\rho - \omega/2}{1-\omega}) + \rho \log_2 p + (1-\rho)\log_2(1-p)))\right].$$

The exponent is maximized for $\rho = \omega/2 + (1-\omega)p$. Substituting we obtain

$$P_1 \cong \exp[n\omega \log_2 2\sqrt{p(1-p)}].$$

Of course, $-\omega \log_2 2\sqrt{p(1-p)} < -\log_2 2\sqrt{p(1-p)}$, so (4) is proved.

Let us show that under the condition of the lemma,

$$F(\omega) := -\omega \log_2 2\sqrt{p(1-p)} - D(\omega \| p) < 0.$$

First we check that for $F[2\sqrt{p(1-p)}/(1+2\sqrt{p(1-p)})] < 0$ and $F(1/2) < 0$. Next $F'(\omega) < 0$ for

$$\frac{\sqrt{p}}{\sqrt{p} + 2\sqrt{1-p} - 2p\sqrt{1-p}} < \omega < 1/2.$$

Finally,

$$\frac{\sqrt{p}}{\sqrt{p} + 2\sqrt{1-p} - 2p\sqrt{1-p}} < \frac{2\sqrt{p(1-p)}}{1+2\sqrt{p(1-p)}}$$

for all $0 < p < 1/2$. This shows that $F(\omega) < 0$, so our claim is proved. ∎

Theorem 4. (Bhattacharyya bound) *Let $\mathcal{C}$ be a code with distance enumerator $B(x,y)$ used for transmission over a BSC(p). Then*

$$P_e(\mathcal{C}) \le B(1, 2\sqrt{p(1-p)}) - 1$$

Proof : An asymptotic equivalent of the claimed inequality follows directly from the union bound (1) and the previous lemma (part (i)). Let us prove that a stronger version claimed here is also true. Let $\mathbf{c} \in \mathcal{C}$ be the transmitted vector, which can be assumed all-zero.

$$\begin{aligned}
P_e(\mathbf{c}) &\le \sum_{\mathbf{c}' \in \mathcal{C} \setminus \{\mathbf{c}\}} P(\mathbf{c} \to \mathbf{c}') \le \sum_{\mathbf{c}'} \sum_{\mathbf{y}: \mathrm{d}(\mathbf{c}',\mathbf{y}) \le \mathrm{d}(\mathbf{c},\mathbf{y})} P(\mathbf{y}|\mathbf{c}) \\
&\le \sum_{\mathbf{c}'} \sum_{\mathbf{y}: \mathrm{d}(\mathbf{c}',\mathbf{y}) \le \mathrm{d}(\mathbf{c},\mathbf{y})} \sqrt{P(\mathbf{y}|\mathbf{c})P(\mathbf{y}|\mathbf{c}')} \\
&\le \sum_{\mathbf{c}'} \sum_{\mathbf{y} \in \mathscr{H}_2^n} \sqrt{P(\mathbf{y}|\mathbf{c})P(\mathbf{y}|\mathbf{c}')} = \sum_{\mathbf{c}'} \prod_{i=1}^{n} \sum_{y=0}^{1} \sqrt{P(y|c_i)P(y|c_i')} \\
&= \sum_{\mathbf{c}'} \prod_{i: c_i'=1} \sum_{y=0}^{1} \sqrt{P(y|c_i)P(y|c_i')} = \sum_{\mathbf{c}'} \prod_{i: c_i'=1} 2\sqrt{p(1-p)} \\
&= \sum_{w=1}^{n} B_w(\mathbf{c})(2\sqrt{p(1-p)})^w,
\end{aligned}$$

where $B_w(\mathbf{c})$ is the number of neighbors of $\mathbf{c}$ in $\mathcal{C}$ at distance w. The proof is completed by averaging over $\mathbf{c}$. ∎

The Bhattacharyya bound is good for small values of p relative to the code distance and deteriorates for stronger noise. For (much) more information on this bound see [9].

The main result about the error probability of decoding for the qSC is given in the following theorem.

Theorem 5. *Let $\mathscr{C} = \log_2 q - h_q(p)$. For any rate $0 \le R \le \mathscr{C}$ there exists a sequence of codes $\mathcal{C}_n \subset \mathscr{H}_q^n, n = 1, 2, \ldots$ of length n such that $R(\mathcal{C}_n) \to R$ as $n \to \infty$ and*

$$P_e(\mathcal{C}_n) < \exp(-n(E(R,p) - o(1))), \tag{6}$$

where $E(R,p) > 0$ is a convex, monotone decreasing function of R. Conversely, for $R > \mathscr{C}$ the error probability $P_e(\mathcal{C})$ of any sequence of codes approaches one.

This theorem is a simple corollary of Lemma 2 and the Bhattacharyya bound if one substitutes in it the weight profile α_0 of a random linear code. However I prefer another proof which develops geometric intuition about the decoding process (one of the next lectures).

What to remember:

There are code sequences which under complete (ML decoding) have error probability that falls exponentially with the code length for any code rate below capacity. For instance, random linear codes have this property (and achieve the best known error exponent), and generally a good code sequence is expected to have this property.

Analogous results hold for a large class of information transmission channels. The quantity $\mathscr{C} = 1 - h_2(p)$, which depends only on p, is called the *capacity of the channel*. For the Gaussian channel with signal-to-noise ratio A

$$\mathscr{C} = \frac{1}{2}\ln(1 + A). \tag{7}$$

Though Theorem 5 says that there are codes of sufficiently large length for which P_e falls as long as $R(\mathcal{C}) < \mathscr{C}$ and that for any code $P_e \to 1$ if $R(\mathcal{C}) > \mathscr{C}$, it does not imply any conclusions for a particular code $\mathcal{C}$. However, a similar result is still possible. We begin with an auxiliary statement.

Theorem 6. *Suppose a code $\mathcal{C} \subset \mathscr{H}_q^n$ is used for transmission over a qSC with transition probability $p, 0 < p < 1 - q^{-1}$. Then the error probability of complete decoding $P_e(\mathcal{C}) = P_e(\mathcal{C}, p)$ is a continuous monotone increasing function of p and $P_e(\mathcal{C}, 0) = 0$, $P_e(\mathcal{C}, 1 - q^{-1}) > 1/2$.*

Proof : Only the inequality $P_e(\mathcal{C}, 1 - q^{-1}) > 1/2$ is nonobvious. Let $\mathcal{C} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_M)$ be a code and let $D(\mathbf{c}_i, \mathcal{C})$ be the Voronoi region of the codeword $\mathbf{c}_i$. We will prove that $|D(\mathbf{c}_i, \mathcal{C})| < q^n/2$ for any $1 \le i \le M$ if $M \ge 3$. First let $M = 3$. The claim is true by inspection for $n = 2$. Suppose it is true for any code of length $n - 1$. Consider a code of length n. Puncturing it on any fixed coordinate, we obtain a code $\mathcal{C}'$ for which $D(\mathbf{c}'_i, \mathcal{C}') < q^{n-1}/2$ for all $i = 1, 2, 3$. Consider the way the Voronoi regions change in transition from $\mathcal{C}'$ to $\mathcal{C}$. Every vector $\mathbf{y}' \in D(\mathbf{c}'_i, \mathcal{C}')$ can be augmented in q ways to a vector $\mathbf{y}$ of length n; even if all these vectors are in $D(\mathbf{c}_i, \mathcal{C})$, then

$$|D(\mathbf{c}_i, \mathcal{C})| \le q|D(\mathbf{c}'_i, \mathcal{C}')| < \frac{q^{n-1}}{2} q = q^n/2.$$

Thus the fact $|D(\mathbf{c}_i, \mathcal{C})| < q^n/2$ is justified by induction on n for $M = 3$. Now perform induction on M. For the code $C \backslash \mathbf{c}_M$ the claim is true by the induction hypothesis; adding a codeword can only decrease the Voronoi regions of $\mathbf{c}_1, \ldots, \mathbf{c}_{M-1}$. Thus the full claim is justified by symmetry.

Now observe that the qSC with transition probability $1 - q^{-1}$ induces on $\mathscr{H}_q^n$ a uniform distribution. Together with our assumption that decoding $\psi(\mathbf{x})$ always ends in error if there are two or more codewords at an equal distance from $\mathbf{x}$ we obtain

$$P_e(\mathcal{C}) = \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}}(1 - P_c(\mathbf{c})) = \frac{1}{M}\sum_{\mathbf{c}\in\mathcal{C}} q^{-n}(q^n - |D(\mathbf{c}, \mathcal{C})|) > q^{-n}\frac{q^n}{2} = 1/2,$$

where $P_c(\mathbf{c})$ is the probability of *correct* decoding conditioned on the fact that $\mathbf{c}$ is transmitted. ∎

This theorem enables us to give the following definition.

Definition 2. *Suppose a code $\mathcal{C}$ is used over a qSC with complete decoding. The transition probability θ is called the* threshold probability *of $\mathcal{C}$ if $P_e(\mathcal{C}, \theta) = 1/2$.*

The term "threshold" suggests that θ separates in some way the values of $P_e(\mathcal{C}, p)$. This is indeed the case as shown by the following theorem [8], stated for binary codes.

Theorem 7. *Let $\mathcal{C}$ be a binary linear code of any length and distance d used over a BSC with crossover probability p. Then for any codeword $\mathbf{c} \in \mathcal{C}$*

$$P_e(\mathbf{c}) \leq 1 - \Phi\big[\sqrt{d}\big(\sqrt{-\ln(1-\theta)} - \sqrt{-\ln(1-p)}\big)\big] \quad \textit{for} \quad 0 < p < \theta$$

$$P_e(\mathbf{c}) \geq 1 - \Phi\big[\sqrt{d}\big(\sqrt{-\ln(1-\theta)} - \sqrt{-\ln(1-p)}\big)\big] \quad \textit{for} \quad \theta < p < 1,$$

where $\Phi(x) = \int_{-\infty}^{x} \exp(-z^2/2)dz/\sqrt{2\pi}$.

To see that $P_e(\mathcal{C})$ exhibits a threshold behavior, recall the asymptotics

$$1 - \Phi(x) \sim \frac{1}{\sqrt{2\pi}x} e^{-x^2/2} \quad (x \to +\infty).$$

Let $s = \sqrt{-\ln(1-\theta)} - \sqrt{-\ln(1-p)}$. Now assume that $d \to \infty$, then

$$P_e(\mathcal{C}) \lesssim \frac{1}{\sqrt{2\pi}s} e^{-ds^2/2}.$$

Similarly, for $\theta < p$, the error probability $P_e(\mathcal{C}) \geq 1 - o(1)$.

Note that if the relative distance of a code is separated from zero then for a very low noise the error probability $P_e(\mathcal{C})$ will fall as an exponential function of n. Therefore in this case the nontrivial part of this theorem is the fact that P_e exhibits a threshold behavior for large n, jumping from almost 0 to almost 1 for a very small change in the input noise level. However, the result is even more general: it shows that for any code sequence such that $\ln n = o(d)$ the error probability still has a threshold behavior.

The proof is difficult and will not be given. It involves a very interesting general method of modern probabilistic combinatorics, so you are encouraged to consult the original.

What to remember:

Every code sequence with relative distance separated from zero has a sharp threshold.

Error detection

Finally, consider a very simple decoding scheme called *error detection*. The vector $\mathbf{y}$ received from a qSC(p) is checked for containment in the code: if $\mathbf{y} \in \mathcal{C}$ then this is the decoding result, if not, the decoder detects an error. This is particularly simple for linear codes: decoding amounts to computing $|bfH\mathbf{y}^t$ and comparing the result to zero. The probability of undetected error for various classes of codes was studied extensively in the literature [5]. Given a code $\mathcal{C}$ with a distance enumerator $B(x, y)$ this probability can be written as follows:

$$P_u(\mathcal{C}, p) = B(\frac{p}{q-1}, 1-p) - (1-p)^n$$

Proof : Let $\pi(i) = (\frac{p}{q-1}^i)(1-p)^{n-i}$. We compute

$$\begin{aligned} P_u(\mathcal{C}, p) &= \frac{1}{M} \sum_{\mathbf{c} \in \mathcal{C}} \sum_{\mathbf{c}' \in \mathcal{C}} \pi(d(\mathbf{c}, \mathbf{c}')) \\ &= \frac{1}{M} \sum_{\mathbf{c} \in \mathcal{C}} \sum_{i=1}^{n} \mathcal{B}_i(\mathbf{c}) \pi(i) \\ &= \sum_{i=1}^{n} \mathcal{B}_i(\mathcal{C}) \pi(i) \\ &= \mathcal{B}\big(1-p, \frac{p}{q-1}\big) - (1-p)^n. \end{aligned}$$

∎

Let us use this result together with random linear codes to derive a lower bound on the maximum attainable exponent of the probability of undetected error for binary codes of rate R used over a BSC(p).

Let us first give a formal definition of the exponent:

$$\begin{aligned} E_u(n,R,p) &= \max_{\mathcal{C} \text{ an } (n,nR) \text{ code}} -\frac{1}{n}\log_2 P_u(\mathcal{C},p) \\ E_u(R,p) &= \lim_{n\to\infty} E_u(n,R,p) \end{aligned}$$

The existence of this limit again is an unknown fact, so this definition should be handled similarly to the definition of $R(\delta)$.

Theorem 8. [6] *Let $T(\delta_{\mathrm{GV}}(R),p) = h_2(\delta_{\mathrm{GV}}(R)) + D(\delta_{\mathrm{GV}}(R)\|p)$. We have*

$$\begin{aligned} E_u(R,p) &\geq T(\delta_{\mathrm{GV}}(R),p) \qquad 0 \leq R \leq 1-h_2(p) \\ E_u(R,p) &\geq 1-R \qquad\qquad 1-h_2(p) < R \leq 1. \end{aligned}$$

Proof : Take a random $[n, k = Rn, d]$ linear code $\mathcal{C}$ with weight distribution given by

$$A_0 = 1, A_w \leq n^2 \binom{n}{w} 2^{k-n}, \quad w \geq d.$$

For large n the asymptotic behavior of the weight profile is given by Corollary 5 of Lecture 1. Substituting this into the expression for P_u, we obtain

$$P_u(\mathcal{C}) \leq \sum_{w=d}^{n} n^2 \binom{n}{w} 2^{k-n}(1-p)^{n-w}p^w.$$

Switching to exponents, we obtain

$$E_u(R,p) \geq \max_{\omega \geq \delta_{\mathrm{GV}}(R)} (1-R+D(\omega\|p)).$$

Since $D'_\omega(\omega\|p) = \log\frac{(1-\omega)p}{\omega(1-p)}$, the unrestricted maximum in the exponent is attained for $\omega = p$. Thus if $p > \delta_{\mathrm{GV}}(R)$, or equivalently, $R \geq 1-h_2(p)$, we can substitute $\omega = p$ and obtain $1-R$ in the exponent (since $D(p\|p) = 0$. If $p \leq \delta_{\mathrm{GV}}(R)$ the dominating term in the sum for P_u is the first one, i.e., $w = d \to n\delta_{\mathrm{GV}}(R)$. The number of codewords of minimum weight in the code is nonexponential in n and the exponent of P_u is equal to the exponent of $p^d(1-p)^{n-d}$ which is $T(\delta_{\mathrm{GV}}(R),p)$. ∎

References

1. P. Charpin, *Tools for coset weight enumerators of some codes*, Finite fields: theory, applications, and algorithms (Las Vegas, NV, 1993), Contemp. Math., vol. 168, Amer. Math. Soc., Providence, RI, 1994, pp. 1–14.
2. ———, *Weight distributions of cosets of two-error-correcting binary BCH codes, extended or not*, IEEE Trans. Inform. Theory **40** (1994), no. 5, 1425–1442.
3. I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless channels*, Akadémiai Kiadó, Budapest, 1981.
4. R. G. Gallager, *Information theory and reliable communication*, John Wiley & Sons, New York e.a., 1968.
5. T. Kløve and V. I. Korzhik, *Error detecting codes*, Kluwer, Boston e. a., 1995.
6. V. I. Levenshtein, *Bounds on the probability of undetected error*, Problems of Information Transmission **13** (1977), no. 1, 3–18.
7. C. E. Shannon, *Probability of error for optimal codes in a Gaussian channel*, Bell Syst. Techn. Journ. **38** (1959), no. 3, 611–656.
8. J.-P. Tillich and G. Zémor, *Discrete isoperimetric inequalities and the probability of decoding error*, Combinatorics, Probability and Computing **9** (2000), 465–479.
9. A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*, McGraw-Hill, 1979.