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

Lecture 7 (**draft;** 10/14/03). Random matrices over finite fields. Erasure channel and its error exponent. Complexity issues in coding theory.

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

In lectures 3-6 we looked at decoding of codes from a probabilistic perspective, ignoring the constructive aspect of our systems. Here we wish to change the point of view and study issues related to implementation complexity of decoding of linear codes. We will start with a technical topic of independent interest: properties of random matrices over $\mathbb{F}_q$. The main use of these results will be in analysis of some decoding algorithms; however we also point out another, unrelated direction which we can treat almost "for free", that of the reliability function of the erasure channel.

Let q be a prime power and let $\mathcal{C}$ be a linear q-ary $[n,k]$ code. A k-subset of coordinates $E = \{i_1, \dots, i_k\}$ is called an *information set* (a *message set*) if all the q^k codewords differ in these coordinates. In the notation of the previous lecture, $\dim(\mathcal{C}_E) = k$.

The question about the number U_k^k of information sets for a given code is difficult. Some bounds are given in [5]; however, there are no general bounds which are very different from $\binom{n}{k}$. Our first goal will be to explain the reason for this. In doing so, we will analyze the rank of random matrices over $\mathbb{F}_q$.

What is the probability that a large square binary $k \times k$ matrix is nonsingular (over $\mathbb{F}_2$)? The answer is given by

$$\lim_{k\to\infty} \frac{[k]_k}{2^{k^2}} = \prod_{i=1}^{\infty}(1-2^{-i}) \approx 0.2888.$$

Remark. Generally, the probability that a large $k \times k$ matrix over $\mathbb{F}_q$ has rank $k-c$ (corank c) falls very rapidly as c grows. The following table shows this probability for $0 \le c \le 5$:

q \ c	0	1	2	3	4	5
2	.2888	.5776	.1284	.0052	4.7×10^{-5}	9.7×10^{-8}
3	.5601	.4201	.0197	8.7×10^{-5}	4.1×10^{-8}	2.1×10^{-12}
5	.7603	.2376	.0021	6.7×10^{-7}	8.6×10^{-12}	4.4×10^{-18}

What is the expected corank (i.e., $k-(\text{rank})$) of a $k \times k$ binary matrix? It is given by the sum $\sum_c c\pi(c)$, where $\pi(c)$ are given by the first row of this table. The answer is (very close to) 0.8502.

We will need a generalization of this argument, proved by a more detailed analysis. We begin with a technical lemma.

Lemma 1. *The number of $m \times \ell$ matrices of rank r equals*

$$\begin{bmatrix} m \\ r \end{bmatrix} \begin{bmatrix} \ell \\ r \end{bmatrix} [r]_r.$$

Proof : Let $M = \mathbb{F}_q^m$ be an m-dimensional space and $L \subset M$ be its ℓ-dimensional subspace. We will count the number of linear maps $f : M \to L$ of rank r. We have $\dim\ker f = m - r$, so the number of kernels mapping is $\left[{m \atop r}\right]$. Further,

$$\dim \mathrm{Im} f = m - \dim\ker f = r,$$

so the number of image subspaces L is $\left[{\ell \atop r}\right]$. Suppose the image and the kernel are fixed. Then to every choice of the basis in $\mathrm{Im} f$ there corresponds exactly one matrix $\mathbf{A}$ with the needed properties, and there are $[r]_r$ such choices. ∎

Theorem 2. *Let $\mathbf{G}$ be a $k \times n$ matrix over $\mathbb{F}_q$ with independent equiprobable entries. The probability that it contains a $k \times k$ submatrix of rank $\le \ell - 1$ is*

$$\tau_{n,k}(\ell) < \binom{n}{k} q^{-(k-\ell)^2}.$$

Proof : By the previous lemma, the probability that $\mathbf{G}$ contains a $k \times k$ matrix of rank u equals

$$\pi(k,u) = q^{-k^2} \begin{bmatrix} k \\ u \end{bmatrix}^2 [u]_u = \prod_{j=0}^{u-1} \frac{(q^k - q^j)^2}{q^u - q^j} < q^{-k^2+ku} \prod_{j=0}^{u-1} \frac{q^{k-j}-1}{q^{u-j}-1}$$

The last product can be bounded above as

$$\begin{aligned} \begin{bmatrix} k \\ u \end{bmatrix} &< \prod_{j=0}^{u-1} \frac{q^{k-i}}{q^{u-i}-1} = q^{u(k-u)} \prod_{j=0}^{u-1} \frac{q^{u-i}}{q^{u-i}-1} \\ &= q^{u(k-u)} \prod_{i=1}^{u} \Big(1 + \frac{1}{q^i - 1}\Big). \end{aligned}$$

Denote the last product by Π. We have

$$\Pi = \frac{q}{q-1}\Big(1 + \frac{1}{q^2-1} + \dots\Big)$$

For $q \geq 2$ the sum of the omitted terms is less than 1, so $\Pi < 5$. Then

$$\pi(k,u) < 5q^{u(k-u)-k^2+ku} = 5q^{-(k-u)^2}.$$

Now let us use the union bound. For a given matrix $\mathbf{G}$ there are $\binom{n}{k}$ choices of the $k \times k$ submatrix; thus, the probability that there exists a submatrix $\mathbf{A}$ of rank $\mathrm{rk}(\mathbf{A}) = u < \ell$ does not exceed

$$\begin{aligned} \tau_{n,k}(\ell) &\leq \sum_{u=0}^{k-\ell} 5q^{-(k-u)^2} \binom{n}{k} = 5\binom{n}{k} q^{-(k-\ell)^2} \sum_{u=0}^{\ell-1} q^{-(k-u)^2+(k-\ell)^2} \\ &\leq \Big\{u = \ell - i\Big\} \leq 5\binom{n}{k} q^{-(k-\ell)^2} \sum_{i=1}^{\ell} q^{-(i+1)^2+1}. \end{aligned}$$

The last sum is maximum for $q = 2$. It can be checked not to exceed 0.2. This proved the claimed inequality. ∎

Corollary 3. *Let $n \to \infty, k/n \to R > 0, c \geq h_q(R)/R$. For almost all choices of $\mathbf{G}$ the rank of every $k \times k$ submatrix is at least*

$$k - \sqrt{ck}.$$

Proof : Let $E \subset [n]$ be a k-subset. The probability that $\mathbf{G}$ contains a submatrix $\mathbf{G}(E)$ of rank $\leq \ell - 1$ is at most $\tau_{n,k}(\ell)$. Let $k - \ell = \sqrt{ck}$. Compute

$$\log_q \tau_{n,k}(\ell) < -ck + \log_q \binom{n}{k} < -k\Big(c - \frac{h_q(R)}{R}\Big).$$

Thus for $c > h_q(R)/R$ the probability $\tau_{n,k}(\ell) \to 0$ exponentially fast. In other words, for almost all matrices $\mathbf{G}$ the rank of every $k \times k$ submatrix is $\geq k - \sqrt{ck}$. ∎

By a variation on the above argument we can prove

Lemma 4. *Let $\mathbf{G}$ be a $k \times n$ random matrix over $\mathbb{F}_q$, $n \to \infty, k/n$ fixed. Then with probability $\to 1$ over the choice of $\mathbf{G}$ every submatrix formed by $k + O(\sqrt{k})$ columns of $\mathbf{G}$ has rank k.*

Lemma 5. *Let $\mathbf{G}$ be as above. Then with probability $\to 1$ every submatrix formed by $k + \lceil 2\log_q n \rceil$ cyclically consecutive columns is nonsingular.*

Erasure channel. Consider a memoryless binary-input channel $\mathscr{W}$ with $\mathcal{X} = \{0,1\}$, $\mathcal{Y} = \{0,1,*\}$, where $*$ represents erasure. The transition probabilities are given by $P(a|a) = 1-p, P(*|a) = p$.

Let $\mathcal{C}$ be a binary $[n,k]$ code and let $\mathbf{x} \in \mathcal{C}$ be a codevector transmitted over $\mathscr{W}$. Let $E \subset \{1,\dots,n\}$ be the subset of nonerased positions. The received vector $\mathbf{y}$ equals $\mathbf{x}$ on E. It is clear that if E contains an information set of $\mathcal{C}$, it is possible to decode correctly. Otherwise, if $\mathrm{rk}(\mathbf{G}(E)) = k - i$, then $\dim(C^{\bar{E}}) = i$.

There is a coset of $\mathscr{H}_2^n/C^{\bar{E}}$ in which every vector coincides with $\mathbf{y}$ on its nonerased positions, so any decoding is very likely to miscorrect.

To summarize, *max-likelihood decoding* on the erasure channel is performed as follows: find $\mathbf{c} \in \mathcal{C}$ which is the unique closest codeword $\mathbf{c} \in \mathcal{C}$ to $\mathbf{y}$ on the nonerased positions. If it does not exist, declare an error, otherwise, output $\mathbf{c}$.

Let $P_e(\mathcal{C}, p)$ be the error probability of decoding of $\mathcal{C}$ on $\mathscr{W}$. As usual, define

$$E(n, R, p) = \max_{\mathcal{C}[n,nR]} -\frac{1}{n} \log_2 P_e(\mathcal{C}, p)$$

$$E(R, p) = \lim_{n\to\infty} E(n, R, p).$$

Theorem 6.

$$E(R, p) \geq D(1 - R\|p).$$

Proof : Let $\mathcal{C}$ be an $[n, k]$ random linear code and $\kappa = k + O(\sqrt{k})$. Let $f_s = \sum_{i=0}^{k-1} U_{n-s}^i$ be the number of subsets E of size $n - s$ and rank $\mathrm{rk}(\mathbf{G}(E)) \leq k - 1$. The error probability

$$P_e(\mathcal{C}, p) = \sum_{e=1}^{n} f_e p^e (1 - p)^{n-e}.$$

By Theorem 2 with probability $\to 1$ over the choice of codes, $f_e = 0$ for $e < n - \kappa$ and $f_e = \binom{n}{e}$ for $e \geq n - \kappa$. Hence there exists a linear $[n, nR]$ code $\mathcal{C}$ such that

$$P_e(\mathcal{C}, p) = \sum_{e=n-\kappa}^{n} \binom{n}{e} p^e (1 - p)^{n-e}.$$

As long as $np \leq n - \kappa$ or $p \leq 1 - R + o(1)$, the exponent of $P_e(\mathcal{C}, p)$ for $n \to \infty$ is $D(1 - R\|p)$. ∎

Note that this argument can also be viewed as an application of Lemma 3.2.

Thus, the capacity $\mathscr{C} \geq 1 - p$. It should also be clear that $\mathscr{C} = 1 - p$ because for $k = Rn > n(1 - p)$, with probability bounded away from zero the received vector $\mathbf{y}$ will contain more than $n - k$ erased positions, so there always will be an ambiguity in decoding. An even stronger fact follows.

Fact to remember:

The critical rate of the channel is $R_{\text{crit}} = 0$, and the entire error exponent is the "sphere packing bound" which corresponds to the third case of the general theorems. This means that the reliability function of the erasure channel is known for all rates $0 \leq R \leq \mathscr{C}$.

Complexity issues in coding theory. In this part we will look at the complexity of max-likelihood decoding and related questions. Our source here is [1]. It is known that the problem of decoding of linear codes in general is NP-hard (meaning that it is unlikely that a polynomial-time algorithm for the general case of this problem exists). Let us take a look at a few basic results related to decoding complexity.

There are a few trivial algorithms: search over all codewords, finding the codeword $\mathbf{c}$ which minimizes the distance $\mathrm{d}(\mathbf{y}, \mathbf{c}')$ where $\mathbf{y}$ is the received vector and $\mathbf{c}' \in \mathcal{C}$. The complexity of this decoding is $O(2^{Rn})$. Or we can decode by the standard array (listing coset leaders and their syndromes), the complexity is $2^{n(1-R)}$.

(Give an example of the standard array: coset leaders, most probable vectors).

Can we do better in terms of complexity? Let $\mathcal{C}$ be a q-ary linear $[n, k, d]$ code and let W be an information set, i.e. a k-subset $W \subset [n]$ such that $\mathrm{rk}(\mathbf{G}(W)) = k$. A codevector $\mathbf{x}$ can be uniquely restored from its projection on W, for instance, by solving the system of linear equations

$$\mathbf{H}\mathbf{x}^T = 0,$$

where the coordinates of $\mathbf{x}$ outside W are unknown. Denoting $\mathbf{s} = \mathbf{H}(W)\mathbf{x}(W)^T$, we have

$$\mathbf{H}(\bar{W})\mathbf{x}(\bar{W})^T = \mathbf{s}.$$

Since $0 = k - \mathrm{rk}(\mathbf{G}(W)) = n - k - \mathrm{rk}(\mathbf{H}(\bar{W}))$, the rank $\mathrm{rk}(\mathbf{H}(\bar{W})) = n - k$, so the system has a unique solution. This motivates the following decoding algorithm

Information set decoding $(\mathcal{C}, \mathbf{y})$

- Set $\mathbf{c} = 0$.
- Choose an information set W. Compute the codeword $\mathbf{c}'$ such that $\mathbf{c}'(W) = \mathbf{y}(W)$. If $\mathrm{d}(\mathbf{c}', \mathbf{y}) < \mathrm{d}(\mathbf{c}, \mathbf{y})$, assign $\mathbf{c} \leftarrow \mathbf{c}'$.
- Repeat for all information windows. Output $\mathbf{c}$.

Theorem 7. *The information set decoding algorithm performs ML decoding.*

Proof : (see Homework 2)

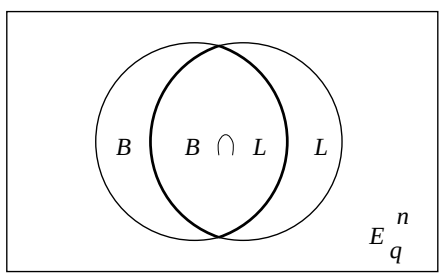
It is possible to use information set decoding to perform bounded distance decoding (that is to correct up to a certain multiplicity of errors). Choosing a small collection of information sets which will solve this problem for a given code turns into a separate task. (Discuss the example of the $[24, 12, 8]$ Golay code).

The following lemma shows that by giving up a little on the code performance we can indeed decode with an exponentially smaller complexity. Given n, k, define (again) the GV distance

$$d_0 = \max\left\{d : \sum_{i=0}^{d-1} \binom{n}{i} (q-1)^i \leq q^{n-k}\right\}.$$

Lemma 8. [4] *Let $\mathcal{C}$ be an $[n, k]$ linear code and let d_0 be the GV distance. Let $p_c = P_e(\mathcal{C}, p)$ be the error probability of ML decoding of $\mathcal{C}$ on a given BSC(p). Let p_b be the error probability of bounded distance decoding of $\mathcal{C}$ in the sphere of radius d_0. Then $p_b \leq 2p_c$.*

Proof :



Let L be the set of q^{n-k} coset leaders. Every error pattern $\mathbf{e}$ outside L contributes to the error probability p_c:

$$p_c = \Pr\{\mathbf{e} \in \mathscr{H}_q^n \setminus L\} \tag{1}$$

For the error probability of bounded distance decoding we have

$$\begin{aligned} p_b = \Pr\{\mathbf{e} \in \mathscr{H}_q^n \setminus (B \cap L)\} &= \Pr\{\mathbf{e} \in \mathscr{H}_q^n \setminus L\} + \Pr\{\mathbf{e} \in L \setminus (B \cap L)\} \\ &\leq p_c + \Pr\{\mathbf{e} \in B \setminus (B \cap L)\} \end{aligned}$$

The last inequality follows because $|B| = |L|$ and B is formed by the most probable vectors. Finally, observe that the last term describes a part of the event in (1), so its probability does not exceed p_c. ∎

Fact to remember:

Let $\mathcal{C}$ be an $[n, k]$ linear code used on a BSC. Performing bounded distance decoding restricted to the sphere of radius equal to the GV distance for the parameters n, k at most doubles the error probability of complete ML decoding.

Therefore, consider the following decoding algorithm. Let $\kappa = k + \lceil 2\log_q n \rceil$.

Sliding window decoding $(\mathcal{C}, \mathbf{y})$

- Set $\mathbf{c} = 0, i = 1$
- For a subset $W = \{i, i+1, i+\kappa-1\} \subset [n]$ of (cyclically) consecutive indices repeat:
  - For every error pattern $\mathbf{e} \in \mathscr{H}_q^\kappa$ of weight $\mathrm{wt}(\mathbf{e}) \le \lfloor d_0\kappa/n \rfloor$ reencode the vector $\mathbf{y}(W) - \mathbf{e}$ to obtain a codeword $\mathbf{c}'$. If $\mathrm{d}(\mathbf{c}', \mathbf{y}) < \mathrm{d}(\mathbf{c}, \mathbf{y})$, assign $\mathbf{c} \leftarrow \mathbf{c}'$.
- Repeat the previous step for all $i = 1, 2, \ldots, n$. Output $\mathbf{c}$.

It is clear that if the weight of error $\mathrm{wt}(\mathbf{e}) \le d_0$, one of the projections $\mathbf{e}(W)$ will be of weight at most $\lfloor d_0\kappa/n \rfloor$. Furthermore, for almost all codes every submatrix $\mathbf{G}(W)$ will have rank k. Moreover, for each W we have to search over

$$\sum_{i=0}^{\lfloor d_0\kappa/n \rfloor} \binom{\kappa}{i}(q-1)^i \le 2^{\kappa h_q(\delta_{\mathrm{GV}})} \lesssim 2^{\kappa(\log_2 q - R)}$$

error patterns.

Hence we have

Theorem 9. [4] *Almost all linear codes can be decoded with error probability $p \le 2p_c$ and complexity of order $O(n^4 2^{nR(1-R\log_q 2)})$.*

The complexity estimate in this theorem can be improved. We know that every $k \times k$ submatrix $\mathbf{G}(E)$ of a random generator matrix $\mathbf{G}$ is of rank $\ge k - O(\sqrt{k})$. In other words, for $E \subset [n], |E| = k$ we have $\dim(\mathcal{C}_E) \ge k - O(\sqrt{k})$. Thus the number of codewords project to the all-zero vector (or in general project identically) on E does not exceed $2^{O(\sqrt{k})}$ (recall Thm. 6.1(ii): $\mathcal{C}_E \cong \mathcal{C}/\mathcal{C}^{\bar{E}}$).

Let

$$T_n(k) = (n\log n)\binom{n}{d_0}\Big/\binom{n-k}{d_0}.$$

consider the following (probabilistic) algorithm:

Covering set decoding $(\mathcal{C}, \mathbf{y})$

- Set $\mathbf{c} = 0$.
- Choose randomly a k-subset W. Form a list of codewords $L(W) = \{\mathbf{c} \in \mathcal{C} \,|\, \mathbf{c}(W) = \mathbf{y}(W)\}$.
- If there is a $\mathbf{c}' \in L(W)$ such that $\mathrm{d}(\mathbf{c}', \mathbf{y}) < \mathrm{d}(\mathbf{c}, \mathbf{y})$, assign $\mathbf{c} \leftarrow \mathbf{c}'$.
- Repeat the last two steps $T_n(k)$ times. Output $\mathbf{c}$.

The properties of this algorithm are summarized as follows.

Theorem 10. [2, 7] *For almost every choice of a linear code $\mathcal{C}$, covering set decoding for almost all codes has error probability $\le 2p_c(\mathcal{C})(1+o(1))$. Its implementation complexity has exponential order*

$$\exp_2[n(h_2(\delta_{\mathrm{GV}}) - (1-R)h_2\big(\frac{\delta_{\mathrm{GV}}}{1-R}\big)(1+o(1))].$$

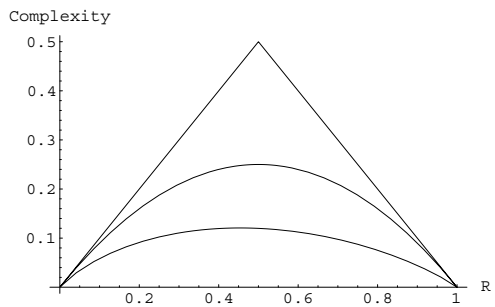
Proof : A decoding error can occur if the transmitted codeword is not the nearest one in the code to the received word $\mathbf{y}$, which happens with probability p_c, or if the repeated choice fails to find an error-free k-set. The probability that a randomly chosen k-set W is not error-free equals

$$1 - \binom{n-k}{d_0}\Big/\binom{n}{d_0}.$$

Performing the choice independently $T_n(k)$ times we observe that this probability falls as $e^{-n\log n}$. This term declines faster than the error probability of minimum distance decoding p_c; hence, its contribution to the overall error rate is negligible. This proves the first part of our claim.

The complexity of each independent decoding attempt is formed by the time needed to diagonalize the matrix G with respect to W, which takes at most n^3 operations, and a number of linear-time subroutines. Hence the overall complexity is at most $O(n^3 T_n(k)|L(W)|)$. ∎

This result is an improvement over the previous theorem (and both improve substantially over the trivial complexity estimates) as seen from the following figure.



This is not the end of the story: further improvements in complexity are possible (see [1]), although all the known maximum likelihood decoding algorithms have exponentially growing implementation complexity.

Gradient-like decoding. Consider a general decoding procedure motivated by the concept of minimal vectors from the previous lecture. Recall that for a given linear code $\mathcal{C}$, $D(0,\mathcal{C})$ denotes the Voronoi region of zero. Call $T \subset \mathscr{H}_q^n$ a *test set* if for every $\mathbf{y} \in \mathscr{H}_q^n$ either $\mathbf{y} \in D(0,\mathcal{C})$ or there is a codeword $\mathbf{z} \in T$ such that

$$\mathrm{wt}(\mathbf{y}-\mathbf{z}) < \mathrm{wt}(\mathbf{z}).$$

Gradient-like decoding $(\mathcal{C}, T, \mathbf{y})$

- Set $\mathbf{c} = 0$.
- Find $\mathbf{z} \in T$ such that $\mathrm{wt}(\mathbf{y}-\mathbf{z}) < \mathrm{wt}\,\mathbf{y}$. Let $\mathbf{c} \leftarrow \mathbf{c}+\mathbf{z},\ \mathbf{y} \leftarrow \mathbf{y}-\mathbf{z}$.
- Repeat until no such $\mathbf{z}$ is found. Output $\mathbf{c}$.

Theorem 11. *The gradient-like decoding algorithm performs complete ML decoding. Its implementation complexity is* $O(n^2|T|)$.

Proof : Let $\mathbf{y} \notin D(0)$. Then there exists $\mathbf{z} \in T$ such that $\mathrm{d}(\mathbf{y},0) < \mathrm{d}(\mathbf{y},\mathbf{z}_1)$ for some $\mathbf{z}_1 \in T$. Suppose that $\mathbf{y}-\mathbf{z}_1 \notin D(0)$, then there is $\mathbf{z}_2$ such that subtracting it from $\mathbf{y}-\mathbf{z}_1$ reduces its weight. After a certain number of steps, say m, we will obtain a vector $\mathbf{y}-\sum_{i=1}^m \mathbf{z}_u \in D(0)$. This means that $\mathbf{y} \in D(\sum \mathbf{z}_u, \mathcal{C})$, which completes the decoding task. ∎

Theorem 12. *Minimal vectors in a* binary *linear code form a test set.*

Proof : Suppose $\mathbf{y} \notin D(0,\mathcal{C})$, then there exists a codevector $\mathbf{c}$ such that $\mathrm{d}(\mathbf{c},\mathbf{y}) < \mathrm{d}(\mathbf{c},0)$ or $\mathrm{wt}(\mathbf{y}+\mathbf{c}) < \mathrm{wt}(\mathbf{y})$. Next write $\mathbf{c} = \sum_i \mathbf{m}_i$ where the $\mathbf{m}_i$ are minimal vectors with disjoint supports. Clearly for at least one of these vectors, say $\mathbf{m}_1$, we must have $\mathrm{wt}(\mathbf{y}+\mathbf{m}_1) < \mathrm{wt}(\mathbf{y})$. ∎

This theorem implies that if $\mathrm{d}(\mathbf{y},\mathbf{m}) \geq \mathrm{d}(\mathbf{y},0)$ for every $\mathbf{m} \in \mathcal{M}(\mathcal{C})$ then $\mathbf{y}$ is a correctable error. Therefore, we can improve our estimate of the error probability of max-likelihood decoding from Eq. (3.2)

as follows;

$$P_e(\mathcal{C},p) \le \sum_{w\ge d} A_w^M \sum_{e=\lceil w/2\rceil}^{n} \pi(e) \sum_{s=0}^{e} p_{e,s}^w,$$

where A_w^M is the number of minimal vectors of weight w. As we said in the previous lecture, for a random code and large n the number $A_w^M \cong A_w$ so there is no hope for improving our exponential estimates of the error probability[1]. In examples the number of minimal vectors is also large. For instance, let $\mathcal{C}$ be a $[2^m, \binom{m}{2}+m+1, 2^{m-2}]$ second order Reed-Muller code, $m=6$. Then $|\mathcal{C}| = 4194304, |\mathcal{M}| = 3821804$.

Another example of a test set is given by zero neighbors. Consider the set B of (error) vectors $\mathbf{y}$ such that $\mathrm{d}(\mathbf{y}, D(0,\mathcal{C})) = 1$. Consider the set of codewords

$$\mathcal{N} = \{\mathbf{c} : \exists_{\mathbf{y}\in B}\, \mathrm{d}(\mathbf{c},\mathbf{y}) \le \mathrm{wt}(\mathbf{y})\}.$$

(Intuition: the Voronoi regions of the codewords in $\mathcal{N}$ share a common boundary with $D(0,\mathcal{C})$.) Codewords in this set are called *zero neighbors*. It is possible to prove

Theorem 13. *Zero neighbors in a linear code form a test set.*

This result is due to [8].

In communication practice a more important problem related to maximum likelihood decoding is that of decoding of codes on the Gaussian channel. The above results still apply although they are substantially more complicated[2].

Let $\mathcal{C}$ be a code used on a memoryless channel $\mathscr{W} : \mathcal{X} \to \mathcal{Y}$, where we assume that $\mathcal{Y}$ is an additive group (possibly infinite) and that $\mathcal{X} \subset \mathcal{Y}$ is its subgroup. Our goal is, as usual, having received $\mathbf{y}$ from the channel to find the max-likelihood decision $\mathbf{x} \in \mathcal{C}$:

$$\mathbf{x} = \arg\max W(\mathbf{y}|\mathbf{x}).$$

First we note that minimal-vectors decoding of binary linear codes can be used to solve this problem. We proceed as follows. Let

$$v_{\mathbf{y}}(\mathbf{c}) = -\sum_{i=1}^{n} \log W(y_i|c_i).$$

Minimal-vectors ML decoding $(\mathscr{W}, \mathcal{C}, \mathbf{y})$

- Set $\mathbf{c} = 0$.
- Find $\mathbf{m} \in \mathcal{M}$ such that $v_{\mathbf{y}}(\mathbf{c}+\mathbf{m}) < v_{\mathbf{y}}(\mathbf{c})$. Let $\mathbf{c} \leftarrow \mathbf{c}+\mathbf{m}$.
- Repeat until no such $\mathbf{m}$ is found. Output $\mathbf{c}$.

Proposition 14. *For any binary linear code C this algorithm performs complete maximum likelihood decoding.*

Proof : Let $\mathbf{c}$ be the current approximation to the decoding result. Suppose there is a $\mathbf{c}' \in C$ such that $v_{\mathbf{y}}(\mathbf{c}+\mathbf{c}') < v_{\mathbf{y}}(c)$. By Lemma 6a.11 it is possible to write $\mathbf{c}'$ as a sum of minimal vectors. Therefore, let

$$\mathbf{c}' = \sum \mathbf{m}_u, \quad \mathbf{m}_u \in \mathcal{M}.$$

Note that since the code is binary, the supports of different vectors in this expansion are disjoint. Since $\mathbf{c}'$ improves the current decision, so does at least one of the minimal vectors in its expansion. To prove that this process eventually converges, note that if $\epsilon = \min_{\mathbf{c}\ne\mathbf{c}'} |v_{\mathbf{y}}(\mathbf{c}) - v_{\mathbf{y}}(\mathbf{c}')|$, then every decoding iteration reduces the weight by at least ϵ. ∎

[1] An even stronger statement holds true: let $w = (n-k+1)-\ell$. Then if $\ell \to \infty$, no matter how slowly, $A_w^M \sim A_w$.

[2] In general, passing from hard decision decoding to soft decision, or from error multiplicities to probabilities (reliabilities) of symbols usually poses serious problems.

Almost ML decoding on a general channel Let us call a memoryless channel $\mathscr{W} : \mathcal{X} \to \mathcal{Y}$ *symmetric* if $\mathcal{Y}$ can be written as a disjoint union of finite sets $\mathcal{Y} = \cup_\alpha Y_\alpha$ such that every matrix $W_\alpha = \|W(y|x)|, y \in Y_\alpha$ satisfies the usual definition of a symmetric channel (every row (column) is a permutation of a fixed set of numbers).

For instance, the AWGN channel is symmetric: write $Y = \mathbb{R}$ as $\cup_x \{\pm x\}$.

Let $\mathbf{y}$ be the received vector. Recall from Lecture 3 that if the messages are equiprobable, ML decoding finds a codevector $\mathbf{c}$ that satisfies

$$\Pr(\mathbf{c}|\mathbf{y}) > \Pr(\mathbf{c}'|\mathbf{y}), \quad \mathbf{c}' \neq \mathbf{c}.$$

Let us establish a total order $\mathbf{x}_1, \mathbf{x}_2, \dots$ on $\mathcal{X}^n$ induced by the order of aposteriori probabilities:

$$\Pr(\mathbf{x}_1|\mathbf{y}) \geq \Pr(\mathbf{x}_2|\mathbf{y}) \geq \dots$$

where vectors with equal aposteriori probabilities are ordered lexicographically. Given a vector $\mathbf{y}$ let us denote the rank of a vector $\mathbf{x} \in \mathcal{X}^n$ (its number in this ordering) by $\sharp_{\mathbf{y}}(\mathbf{x})$.

Let $N \leq |\mathcal{C}|$ be some number. Let $\mathbf{c}_0 = \psi(\mathbf{y})$ be the ML decision for a given $\mathbf{y} \in \mathcal{Y}^n$. Our goal is to parallellize "bounded distance decoding": we will perform search over a subset of N most probable vectors. Namely, for a given $\mathbf{y}$ let ψ_N be a decoding mapping defined by

$$\psi_N(\mathbf{y}) = \begin{cases} \mathbf{c}_0 & \sharp_{\mathbf{y}}(\mathbf{c}_0) \leq N \\ 0 & \text{otherwise} \end{cases}$$

(to maintain the intuition, recall the framed remark after Lemma 8).

Let us state a far-reaching generalization of Lemma 8 due to [3]

Lemma 15. [3] *Let p_N be the error probability of decoding ψ_N. Then*

$$p_N \leq P_e(\mathcal{C}, \mathscr{W})(1 + \frac{T}{N-T}),$$

where P_e is the error probability of ML decoding of $\mathcal{C}$ on the channel $\mathscr{W}$.

The proof is not easy, so we refer to the source (or to an exposition in [1]).

The goal of this lemma is to formulate a decoding algorithm which will have implementation complexity of the same order as Sliding Window Decoding (i.e., $\exp n(R(1-R))$). The result is due to [3]. We will stop here.

What if we are not satisfied with almost max-likelihood decoding and prefer max-likelihood instead? In that case we may rely on trellis decoding which you have seen in ENEE722.

(Define the syndrome trellis, give examples).

More results in this direction are found in [9], see also a recent paper [6].

This concludes our discussion of exponential-complexity decoding algorithms. In the following lectures we will concentrate on code families which afford polynomial-time decoding algorithms. It will be seen that under these complexity restrictions it is still possible to construct codes with very good performance.

References

1. A. Barg, *Complexity issues in coding theory*, Handbook of Coding Theory (V. Pless and W. C. Huffman, eds.), vol. 1, Elsevier Science, Amsterdam, 1998, pp. 649–754.
2. J. T. Coffey and R. M. Goodman, *The complexity of information set decoding*, IEEE Trans. Inform. Theory **36** (1990), no. 5, 1031–1037.
3. I. Dumer, *Suboptimal decoding of linear codes: partition technique*, IEEE Trans. Inform. Theory **42** (1996), no. 6, part 1, 1971–1986, Codes and complexity.
4. G. S. Evseev, *Complexity od decoding for linear codes*, Problems of Information Transmission **19** (1983), no. 1, 3–8.
5. T. Helleseth, T. Kløve, and V. I. Levenshtein, *On the information function of an error-correcting code*, IEEE Trans. Inform. Theory **43** (1997), no. 2, 549–557.
6. R. Koetter and A. Vardy, *The structure of tail-biting trellises: Minimality and basic principles*, IEEE Trans. Inform. Theory **49** (2003), no. 9, 2081–2105.

7. E. A. Kruk, *A bound for the complexity of decoding of linear block codes*, Problemy Peredachi Informatsii **25** (1989), no. 3, 103–107.
8. L. Levitin and C. R. P. Hartmann, *A new approach to the general minimum distance decoding problem: The zero-neighbors algorithm*, IEEE Trans. Inform. Theory **31** (1985), 378–384.
9. A. Vardy, *Trellis structure of codes*, Handbook of coding theory, Vol. I, II, North-Holland, Amsterdam, 1998, pp. 1989–2117.