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

Lecture 2 (**draft;** 9/5/03). Average properties of codes.

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

In brief, we do not know very well how to construct good codes. This claim comes with a number of qualifiers (yes we know some very nice algebraic code families and some good randomized code ensembles), and yet in general it is true. How do we know what is achievable in principle in terms of code construction with infinite computational power? The answer is: by computing average properties of some appropriately chosen code ensemble we can then claim that there exist codes that are very close to average. For the main asymptotic problem, that of finding the highest achievable rate $R(\delta)$ of a sequence of binary codes with relative distance δ it is conjectured that "the average is the best possible" (caveat: in a number of other situations this claim is not true).

What can be expected of M vectors collected randomly in $\mathscr{H}_q^n$? The first result is given by

Theorem 1. *Let M be such that*

$$M(M-1) < \frac{q^n}{B_{d-1}},$$

where

$$B_{d-1} = \mathrm{vol}(\mathcal{B}_{d-1}) = \sum_{i=0}^{d-1} \binom{n}{i} (q-1)^i$$

is the volume of the ball in $\mathscr{H}_q^n$. Then there exists an (n, M, d) code $\mathcal{C} \subset \mathscr{H}_q^n$.

Proof : Let $\mathcal{C} = \{x_1, \dots, x_M\}$ be an ordered collection of points. Call $\mathcal{C}$ bad if $d(\mathcal{C}) \le d-1$ and call a point $x_i \in \mathcal{C}$ bad if it has neighbors in $\mathcal{C}$ at distance $\le d-1$. If the points $x_2, \dots, x_M$ are fixed, then x_1 is bad in at most $(M-1)B_{d-1}$ codes. The points $x_2, \dots, x_M$ can be chosen in $q^{n(M-1)}$ ways, so there are no more than $(M-1)B_{d-1}q^{n(M-1)}$ codes in which point x_1 is bad. This is true for any point $x_i, 1 \le i \le M$; thus, there are no more than $M(M-1)B_{d-1}q^{n(M-1)}$ bad codes. If this number is less than the total number of codes q^{nM}, i.e., if

$$M(M-1)B_{d-1} < q^n,$$

then there exists a good code. ∎

This theorem (a coding theory folk lore) exemplifies the random choice method of coding theory. Substantially more refined results along these lines are given in [2]. It turns out that for large n, Theorem 1 accurately describes the parameters of typical codes in $\mathscr{H}_q^n$. More formally, we have the following result.

Theorem 2. [3, 1] *Let $X = \mathscr{H}_2^n$ and $n \to \infty$. For all codes in X of rate R except for a fraction of codes that decreases exponentially with n, the relative distance approaches the bound*

$$2R = 1 - h_2(\delta).$$

Proof : Consider the Shannon ensemble $\mathcal{A}$ of 2^{nM} binary codes, $M = 2^{Rn}$, where every code has probability 2^{-nM}. Or, what is the same, consider a random code formed of M independently chosen vectors, where all the coordinates of every vector are i.i.d. Bernoulli r.v.'s with $P(0) = P(1) = 1/2$.

Let us assume that δ is chosen to satisfy $2R = 1 - h_2(\delta) + \epsilon$, where $\epsilon > 0$. We will prove that with probability approaching 1 a random (n, M) code $\mathcal{C} \in \mathcal{A}$ contains a pair of vectors at distance δn or less. Let $x_1, x_2, \dots, x_M$ be an ordered (multi)set of independent random vectors such that $\Pr[x_i = y] = 2^{-n}$ for any $y \in \{0,1\}^n$. Let $\nu_{i,j}, 1 < j < i < M$, be the indicator random variable of the event $\mathrm{d}(x_i, x_j) = \delta n$. The $\nu_{i,j}$ are pairwise-independent random variables, each with mean

$$\mathsf{E}\nu_{i,j} = \Pr[\nu_{i,j} = 1]$$

and variance

$$\mathrm{Var}[\nu_{i,j}] = \mathsf{E}\nu_{i,j}^2 - (\mathsf{E}\nu_{i,j})^2 = \mathsf{E}\nu_{i,j} - (\mathsf{E}\nu_{i,j})^2 < \mathsf{E}\nu_{i,j}.$$

Consider the number $N_{\mathcal{C}}(d) = \sum_{j<i} \nu_{i,j}$ of unordered pairs of codewords (x_i, x_j) with $i \neq j$ in $\mathcal{C}$ at distance $d = \delta n$ apart. We have

$$\begin{aligned} \mathsf{E}\, N_{\mathcal{C}}(d) &= \binom{M}{2} \mathsf{E}\, \nu_{i,j} \cong 2^{n(2R-1+h_2(\delta))} \\ \mathrm{Var}\,[N_{\mathcal{C}}(d)] &= \binom{M}{2} \mathrm{Var}\,[\nu_{i,j}] < \mathsf{E}\, N_{\mathcal{C}}(d). \end{aligned}$$

For any $\alpha > 0$ by the Chebyshev inequality we have

$$\begin{aligned} \Pr[|N_{\mathcal{C}}(d) - \mathsf{E}\, N_{\mathcal{C}}(d)| \geq \mathsf{E}\, N_{\mathcal{C}}(d)^{(1+\alpha)/2}] &\leq (\mathsf{E}\, N_{\mathcal{C}}(d))^{\alpha} \\ &\cong 2^{\alpha n(1-2R-h_2(\delta))} = 2^{-\alpha n \epsilon} \to 0. \end{aligned}$$

Thus, in particular, with probability tending to 1 we have $N_{\mathcal{C}}(d) > 0$, or, in other words, a random code contains a pair of vectors at distance $d = \delta n$. Since $\epsilon > 0$ can be taken arbitrarily small, this proves an upper bound $\delta \leq h_2^{-1}(1-2R)$ on the relative distance of almost all codes in $\mathcal{A}$.

On the other hand, for any δ such that $2R = 1 - h_2(\delta) - \epsilon$ the average number of codeword pairs with relative distance $d = \delta n$ decreases exponentially with n. Then

$$\Pr[N_{\mathcal{C}}(d) > 1] \leq \mathsf{E}\, N_{\mathcal{C}}(d) \to 0;$$

hence with probability tending to 1 a random code $\mathcal{C}$ has distance $\geq \delta n$. ∎

This theorem implies that for $R > 1/2$ the relative distance of almost all long codes converges to zero. Let us show that there exist codes with much better parameters, which can be constructed by the Gilbert algorithm [4]. This is a greedy algorithm which proves the following theorem.

Theorem 3. [4] *Let M be any number such that*

$$M \,\mathrm{vol}(\mathcal{B}_{d-1}) < q^n.$$

Then there exists an $(n, M+1, \geq d)$ code $\mathcal{C}$ in $\mathscr{H}_q^n$ that can be constructed in at most $O(nq^n M)$ q-ary operations.

Proof : Consider the following greedy algorithm:

(1) Pick any point in $\mathscr{H}_q^n$;
(2) Given a subset $\mathcal{C}_i$ of $i \leq M$ points with minimum pairwise distance at least d, take an arbitrary point $\mathbf{c}_{i+1}$ in
$$\mathscr{H}_q^n \setminus \bigcup_{\mathbf{x} \in \mathcal{C}_i} \mathcal{B}_{d-1}(\mathbf{x}).$$
If such a point does not exist, stop. Output $\mathcal{C} = \mathcal{C}_i$.
(3) Otherwise form $\mathcal{C}_{i+1} = \mathcal{C}_i \cup \{\mathbf{c}_{i+1}\}$, augment i by one, and return to the previous step.

By the volume argument it is clear that if

$$i \cdot \mathrm{vol}(\mathcal{B}_{d-1}) < q^n$$

the algorithm will find a point $\mathbf{c}_{i+1}$. This proves the existence of a code $\mathcal{C}$ with the claimed properties. The complexity of this algorithm can be crudely estimated by the complexity of going over the entire space $\mathscr{H}_q^n$ in each of the M steps. ∎

Of course, complexity considerations were not a part of [4]—this is a later accretion (recall that back then the word "computer" was not very much in use). Back in the 1950s any code construction was a substantial advance.

The *Varshamov procedure* [6] has a somewhat lower complexity that the Gilbert algorithm and produces a code with somewhat better parameters which is also linear!

Theorem 4. [6] *Let n, k, d be a triple of numbers that satisfy*

$$\mathrm{vol}(\mathcal{B}_{d-2}) q^k < q^n.$$

Then there exists a linear $[n+1, \geq k+1, \geq d]$ code $\mathcal{C} \in \mathscr{H}_q^n$ that can be constructed in $O(n^3 q^{n-k})$ operations.

Proof : Let us say that a q-ary matrix has strict column rank m if any m or fewer sets of its columns are linearly independent, and there exists a collection of $m+1$ linearly dependent columns. An $(n-k)\times n$ matrix $\mathbf{H}$ of strict column rank $d-1$ defines an $[n,\geq k,d]$ linear code $\mathcal{C}$. Consider the following procedure.

(1) Let $\mathbf{h}_1$ be any nonzero $(n-k)$-vector.
(2) Given $i\leq n$ column vectors $\mathbf{h}_1,\mathbf{h}_2,\ldots,\mathbf{h}_i$ which form an $(n-k)\times i$ matrix $\mathbf{H}_i$ of strict column rank $\geq d-1$ find a column $\mathbf{h}_{i+1}$ appending which to $\mathbf{H}_i$ does not violate the rank condition. If this is impossible, stop and output $\mathbf{H}=\mathbf{H}_i$.
(3) Form $\mathbf{H}_{i+1}=[\mathbf{H}_i|\mathbf{h}_{i+1}]$. Augment i by one and return to the previous step.

Clearly if the inequality

$$\sum_{j=0}^{i}\binom{i}{j}(q-1)^j\leq \mathrm{vol}(\mathcal{B}_{d-2})<q^{n-k}$$

holds true, it is possible to find $\mathbf{h}_{i+1}$. This proves the existence claim.

The complexity claim follows by observing that for every i we have to find a column vector outside the union of all linear combinations of $d-2$ or fewer columns. Suppose that for some i we have constructed the matrix $\mathbf{H}_i$ with these properties. Let us construct a list of all 2^i linear combinations of column vectors for each of the $\binom{i}{d-2}$ choices of $d-2$ columns. This can be done in at most

$$n^2\sum_{j=0}^{d-2}\binom{i}{j}(q-1)^j\leq n^2\sum_{j=0}^{d-2}\binom{n}{j}(q-1)^j\leq n^2 2^{nh_q(\frac{d-2}{n})}\leq n^2 2^{n(\log_2 q-R)}$$

operations. Since this complexity is required for every $i=1,\ldots,n$ we finally obtain the result. ∎

The situation with the value of the distance in the above theorem is typical for coding theory: by some constructive procedure we claim that the code distance is at least d (but may be greater). In this situation d is called the *designed* distance of the code as opposed to the *true* distance (remember the designed distance of BCH and other cyclic codes?).

As n grows, both Gilbert and Varshamov procedures lead to one and the same asymptotic parameters.

Definition 1. *The* relative Gilbert-Varshamov distance $\delta_{\mathrm{GV}}(R)$ *is defined by the equation*

$$R=\log_2 q-h_q(\delta)\qquad(0\leq\delta\leq 1-1/q).$$

This relation between the code rate R and relative distance δ is called the Gilbert-Varshamov (GV) *bound.*

Let us prove that typical linear codes approach the GV bound.

Theorem 5. *Let $n\to\infty$. For all linear codes in $\mathscr{H}_q^n$ of rate R except for a fraction of codes that decreases exponentially with n, the relative distance approaches $\delta_{\mathrm{GV}}(R)$.*

Proof : (outline) Consider the ensemble $\mathcal{L}$ of random $[n,k=Rn]$ linear binary codes defined by $(n-k)\times n$ parity-check matrices whose elements are chosen independently with $P(0)=P(1)=1/2$. If N_w is the random variable equal to the number of vectors of weight $w>0$ in a code $\mathcal{C}\in\mathcal{L}$, then $\mathsf{E}N_w=\binom{n}{w}(q-1)^w/q^{n-k}$ and $\mathrm{Var}\,N_w\leq\mathsf{E}N_w$. Thus, $\mathsf{E}N_w$ grows exponentially in n for $\omega:=\frac{w}{n}>\delta_{\mathrm{GV}}(R)$. Thus, the relative distance of a random linear code approaches $\delta_{\mathrm{GV}}(R)$ as n grows, and the fraction of codes whose relative distance deviates from δ_{GV} by ϵ tends to 0 exponentially in n for any $\epsilon>0$. ∎

Thus, unrestricted codes (i.e., codes that are not necessarily linear) on the average are much worse than linear codes.

In the nest theorem we compute the average weight distribution of a linear code.

Theorem 6. *Let $S_w=\mathrm{vol}(\mathcal{S}_w)$. There exists a linear $[n,k]$ code $\mathcal{C}$ with $A_0=1$,*

$$A_w(\mathcal{C})\leq\begin{cases}n^2q^{k-n}S_w, & w\ \textit{such that}\ \log S_w\geq(n-k)\log q-2\log n,\\ 0, & w:\log S_w<(n-k)\log q-2\log n.\end{cases}$$

Proof : Consider linear codes defined in the same way as in the proof of Theorem 5. A vector of weight $w > 0$ lies in the kernel of $q^{(n-1)(n-k)}$ matrices. All the $S_w = \binom{n}{w}(q-1)^w$ vectors of weight w are annihilated by at most $S_w q^{(n-1)(n-k)}$ matrices. Thus, on the average the number of vectors of weight w in the code does not exceed $S_w q^{-(n-k)}$ and the fraction of matrices for which this number is $\geq n^2 S_w q^{-(n-k)}$ (call them bad) is at most n^{-2}. Even if the sets of bad matrices for different $w = 1, 2, \ldots, n$ are disjoint, this leaves us with a fraction of $1 - n^{-1}$ of good matrices; any good matrix defines a code $\mathcal{C}$ of dimension $\dim \mathcal{C} \geq k$ over $\mathbb{F}_q$ with

$$A_w(\mathcal{C}) \leq n^2 q^{k-n} S_w, \quad 1 \leq w \leq n.$$

Writing the right-hand side as $\exp[(k-n)\log q + \log S_w + 2\log n]$, we see that once w is such that the exponent becomes negative, we obtain $A_w < 1$. Since $\mathcal{C}$ is linear, this implies that $A_w = 0$ for these values of w. ∎

By a slight modification of this proof we can take n^2 down to $n^{1+\alpha}$ for any $\alpha > 0$.

What to remember:

Most linear codes achieve the GV bound, most unrestricted codes do not. For one and the same relative distance δ the size of a typical unrestricted code is about the square root of the size of a typical linear code.

Definition 2. *Given a code $\mathcal{C}[n, k, d]$ with weight distribution $A_0, A_1, \ldots, A_n$ define its weight profile as follows*

$$\alpha_{w/n} = \frac{1}{n}\log_2 A_w, \quad w = 0, 1, \ldots, n.$$

Thus, $\alpha_0 = 1$ and $\alpha_{w/n} = -\infty$ for $w = 1, \ldots, d-1$. For large n we think of the weight profile as of a continuous function of $\omega \in [0, 1]$.

Corollary 7. *For any $R < \log q - h_q(\delta)$ there exists a sequence of linear codes of growing length n with weight profile α_0, where $\alpha_{0,0} = 0$,*

$$\begin{aligned} &\alpha_{0,\omega} \leq R - \log q + h_q(\omega) && (\delta_{\mathrm{GV}}(R) < \omega < 1 - \delta_{\mathrm{GV}}(R)), \\ &\alpha_{0,\omega} = -\infty && (0 < \omega < \delta_{\mathrm{GV}}(R)). \end{aligned}$$

So far we used averaging for establishing lower bounds on codes[1]. Averaging can be also useful for upper bounds. Indeed, we know from ENEE 722 (see also [5]) that for an (n, M, d) code $\mathcal{C}$

$$M \leq \frac{d}{d - \frac{q-1}{q}n}$$

whenever the denominator is positive. This inequality, called the *Plotkin bound* is proved by computing the average distance between pairs of codewords.

The reason that we are not fully satisfied with the existence results of this lecture is the only way we know of constructing, say, a linear code that meets the GV bound is of complexity that grows as an exponential function of the code length n. This becomes infeasible very rapidly; and even if we are given such a code we still do not know of a good way of decoding it except a (more or less) exhaustive search.

What to remember:

Constructing code sequences that reach the GV bound by any known procedure involves exponential-complexity algorithms

[1]Note that by lower bounds we mean results of the form "there exist codes with some particular properties." A more conventional use of the term *lower bound* not applicable to this lecture would be: "for every code the value of a particular parameter is above some bound." Therefore sometimes the GV bound is referred to as a lower existence bound or simply existence bound.

References

1. A. Barg and G. D. Forney, Jr., *Random codes: Minimum distances and error exponents*, IEEE Trans. Inform. Theory **48** (2002), no. 9, 2568–2573.
2. L. A. Bassalygo, S. I. Gelfand, and M. S. Pinsker, *Simple methods for deriving lower bounds in the theory of codes*, Problems of Information Transmission **27** (1991), no. 4, 3–8 (in Russian) and 277–281 (English transaltion).
3. J. T. Coffey and R. M. Goodman, *Any code of which we cannot think is good*, IEEE Trans. Inform. Theory **36** (1990), no. 6, 1453–1461.
4. E. N. Gilbert, *A comparison of signalling alphabets*, Bell Syst. Techn. Journ. **31** (1952), 504–522.
5. F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, 3 ed., North-Holland, Amsterdam, 1991.
6. R. R. Varshamov, *Estimate of the number of signals in error correcting codes*, Dokl. Akad. Nauk SSSR Soviet Math. - Doklady **117** (1957), 739–741.