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

Lecture 1 (**draft;** 9/2/03). Review: Notions of coding theory, asymptotic notation. Goals of coding theory

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

The goals of this course are: to develop in-depth understanding of the problems of modern coding theory and prepare students for research work in the area of error-correcting codes. The course will offer a unified view of the goals, methods and results of coding theory in a variety of communication scenarios in classical and quantum communication channels. The course will stress the trade-off between performance of code families and complexity of their implementation. We will discuss several topics in the forefront of the present-day research including LDPC codes and iterative decoding, list decoding of algebraic codes, expander codes, and (time permitting) quantum codes.

We assume familiarity with basics of error-correcting codes (ENEE 722), although many concepts will be reviewed as needed. For the first few lectures as a motivation we assume transmission of information over a binary symmetric channel.

This lecture contains probably much more than we will need for quite a while.

Codes and their parameters

We give an assortment of definitions related to binary codes. The concepts defined play a fundamental role in many important coding theory problems discussed later in this lecture and in the course.

The binary *Hamming space* $\mathscr{H}_2^n = \{0,1\}^n$ with the metric $\mathrm{d}(\cdot,\cdot)$.

Code $\mathcal{C} \subset \mathscr{H}_2^n$ of length n and size M. Code distance

$$d = d(C) := \min_{\mathbf{x}\neq\mathbf{y}\in C} \mathrm{d}(\mathbf{x},\mathbf{y}),$$

rate $R = R(\mathcal{C}) := (1/n)\log_2 M$. Given the length, size and distance, we write $\mathcal{C}(n, M, d)$ to denote a code with these parameters.

The *distance distribution* of a code with respect to a vector $\mathbf{x} \in \mathcal{C}$ is the vector $B(\mathbf{x}) = (B_0(\mathbf{x}), B_1(\mathbf{x}), \ldots, B_n(\mathbf{x}))$ where $B_i(\mathbf{x})$ is the number of neighbors of $\mathbf{x}$ in $\mathcal{C}$ at distance i:

$$B_i(\mathbf{x}) = |\{\mathbf{y} \in \mathcal{C} : \mathrm{d}(\mathbf{x},\mathbf{y} = i)\}|.$$

The (average) distance distribution of the code is defined as the vector $B = (B_0, \ldots, B_n)$ such that

$$B_i(\mathcal{C}) = \frac{1}{M}\sum_{\mathbf{x}\in\mathcal{C}} B_i(\mathbf{x}).$$

The distance distribution has two main uses: characterizing combinatorial properties of the code and estimating its error probability of decoding.

The polynomial

$$B(x,y) = \sum_{i=0}^{n} B_i x^{n-i} y^i$$

is called the *distance enumerator* of the code $\mathcal{C}$.

If $\mathcal{C}$ forms a linear subspace of $\mathbb{F}_2^n$ it is called a *linear code*. The quantity $k = \log_2 M$ is called the dimension of the code (or the number of information symbols). We write $\mathcal{C}[n,k,d]$ to denote a linear code of length n, dimension k, and distance d The distance distribution of $\mathcal{C}$ (both average and local, for any codevector $\mathbf{x}$) coincides with the weight distribution of $\mathcal{C}$.

Does the distance (or weight) distribution determine the code up to a natural equivalence relation? No, not at all; in fact, even much more information than that is often not enough to fix the code uniquely. For instance, consider two linear $[6,3,2]$ codes given by their generator matrices

$$\mathbf{G}_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \qquad \mathbf{G}_2 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1. \end{bmatrix}$$

The weight distribution of both $\mathcal{C}_1$ and $\mathcal{C}_2$ is given by $(1,0,3,0,3,0,1)$; however the covering radius $r(\mathcal{C}_1) - 2, r(\mathcal{C}_2 = 3)$ is different. Therefore, the codes $\mathcal{C}_1$ and $\mathcal{C}_2$ cannot be equivalent.

Codes over alphabets of greater size $q \geq 2$ and their parameters are defined analogously. We define the rate of a q-ary code as

$$R = \frac{1}{n} \log_2 M$$

(this is nonstandard but enables us to use binary logs throughout[1]). A *sphere* $\mathcal{S}_w(\mathbf{x}) \in \mathscr{H}_2^n$ is defined as

$$\mathcal{S}_w(\mathbf{x}) = \{\mathbf{y} \in \mathscr{H}_2^n : \mathrm{d}(\mathbf{x}, \mathbf{y}) = w\}$$

A *ball* $\mathcal{B}_w(\mathbf{x})$ is defined accordingly. The volume of the sphere (denoted $\mathrm{vol}(\mathcal{S}_w)$) is independent of the center. Consider next the following function

$$p_{i,j}^k(\mathscr{H}_2^n) = \sharp\{\mathbf{z} \in \mathscr{H}_q^n : \mathrm{d}(\mathbf{z}, \mathbf{x}) = i, \mathrm{d}(\mathbf{z}, \mathbf{y}) = j; \mathrm{d}(\mathbf{x}, \mathbf{y}) = k\}$$

which is the number of triangles with fixed vertices $\mathbf{x}$ and $\mathbf{y}$, distance k apart, and a floating vertex $\mathbf{z}$ that obeys the distance conditions. This is called the *intersection number* of the Hamming space. Convince yourselves that given i, j, k, the intersection number is independent of the points $\mathbf{x}, \mathbf{y}, \mathbf{z}$. Next calculate

$$p_{i,j}^k = \binom{k}{\frac{1}{2}(j-i+k)} \binom{n-k}{\frac{1}{2}(j+i-k)}$$

assuming that $j - i + k$ is even, and $p_{i,j}^k = 0$ if it is odd.

A code defines a packing of spheres in $\mathscr{H}_2^n$ (i.e., an arrangement of pairwise disjoint spheres). The number $t = \lfloor (d-1)/2 \rfloor$ is called the *packing radius* of the code $\mathcal{C}$. The number

$$r(\mathcal{C}) = \max_{\mathbf{x} \in \mathscr{H}_2^n} \min_{\mathbf{c} \in \mathcal{C}} \mathrm{d}(\mathbf{x}, \mathbf{c})$$

is called the *covering radius* of $\mathcal{C}$.

Remark on binary vs. q-ary codes. In coding theory there is a large group of problems for which the binary results can be formulated in a closed form while the pursuit of generality leads to cumbersome calculations (as an example, compute $p_{i,j}^k(\mathscr{H}_q^n)$; note that this is still a closed-form expression). We take a mixed approach, stating the results for arbitrary q in cases where this does not lead to excessive computations and sticking to the binary case otherwise. Note also that traditionally coding theory problems are first studied for binary codes and then "generalized" to the nonbinary case. Admittedly, doing everything in full generality without special mention of the binary case is also a justifiable approach, taken in several coding theory texts.

Important functions

Binary entropy

$$h_2(x) = -x \log_2 x - (1-x) \log_2 (1-x)$$

(entropy of the binomial distribution with two outcomes $0, 1$ taken with probabilities x and $1-x$ respectively). Intuition about $h_2(x)$: it grows very fast for x in the neighborhood of zero, in particular, faster than any polynomial $p(x)$. $h_2(x)$ is symmetric w.r.t. $1/2$. In the neighborhood of $1/2$ the function $h_2(x)$ is essentially a parabola (is approximated well by a quadratic function in x). Indeed, let $\theta = 1/2 - x$, then

$$h_2(\theta) = 1 - \frac{2}{\ln 2} \theta^2 + O(\theta^4).$$

The function $h_2^{-1}(x)$ is a continuous monotone increasing function of $x \in (0,1)$.

[1] In principle, the base of the logarithms does not matter as long as it is the same as the base of the exponent in asymptotic expressions

q-ary entropy, $q \geq 2$

$$h_q(x) = -x \log_2 \frac{x}{q-1} - (1-x)\log_2(1-x).$$

Relative entropy (a.k.a. information divergence or Kullback-Leibler distance between two binomial distributions)

$$D(x\|y) = x \log_2 \frac{x}{y} + (1-x)\log_2 \frac{1-x}{1-y}$$

The role of these functions in coding theory is determined by the asymptotic identities under ASYMPTOTICS.

ASYMPTOTICS

Many if not most general results in coding theory have asymptotic flavor.

Let $n \to \infty$. We use standard notation $f(n) = o(g(n))$ if $\lim_{n\to\infty} f(n)/g(n) = 0$; $f(n) = O(g(n))$ if there exists a number A independent of n such that $f(n) \leq Ag(n)$ at least for all n greater than some n_0. If both $f(n) = O(g(n))$ and $g(n) = O(f(n))$ we write $f(n) = \Theta(g(n))$. Finally for finite or infinite a

$$f(x) \sim g(x) \quad \Leftrightarrow \quad \lim_{x\to a} \frac{f(x)}{g(x)} = 1$$

Examples: $\frac{1}{n} = o(1)$, $\log n = o(n)$, $n^k = o(e^n)$ for constant k, etc.

Exercises. 1. What is $n!$ to n^n? 2. Let k be a constant. Is it true that $\binom{n}{k} \sim n^k$? $\binom{n}{k} = O(\frac{n^k}{k!})$?

We also use nonstandard notation $f(n) \cong g(n)$ if

$$\lim_{n\to\infty} \frac{1}{n} \log \frac{f(n)}{g(n)} = 0.$$

We write $\lesssim$ if this limit is less than zero and $\gtrsim$ if it is greater than zero.

Example: $p(n)2^n \cong q(n)2^n$ where $p(n)$ and $q(n)$ are arbitrary nonzero polynomials.

Sometimes we use an informal notation $f(n) \approx g(n)$ (reads "$f(x)$ is essentially the same as $g(x)$") if the functions $f(n)$ and $g(n)$ behave essentially in a similar way as n grows, and we do not want to specify the nonessential terms. For instance, $p(x)2^{ax} \approx 2^{ax}$ and so on.

Important asymptotic equalities.

$$\binom{n}{\lambda n} \cong 2^{nh_2(\lambda)}$$

$$\Big[\text{more precisely} \qquad \frac{1}{\sqrt{8n\lambda(1-\lambda)}} 2^{nh_2(\lambda)} \leq \binom{n}{\lambda n} \leq \frac{1}{\sqrt{2\pi n\lambda(1-\lambda)}} 2^{nh_2(\lambda)} \qquad \Big]$$

Proof : The proof is based on the Stirling formula for $\Gamma(x)$:

$$\sqrt{2\pi} x^{x-1/2} e^{-x+\frac{1}{12x}-\frac{1}{360x^3}} < \Gamma(x) < \sqrt{2\pi} x^{x-1/2} e^{-x+\frac{1}{12x}}$$

(proved in most calculus books). For instance, let us prove the upper bound. Recalling that for a positive integer n, $\Gamma(n+1) = n!$ and letting $\mu = 1-\lambda$, we compute

$$\binom{n}{\lambda n} = \frac{n!}{(\lambda n)!(\mu n)!} < \frac{1}{\sqrt{2\pi n\lambda\mu}} \frac{1}{\lambda^{\lambda n}\mu^{\mu n}} e^{1/12n - 1/12\lambda\mu n - (\lambda\mu - 1)/360(\lambda\mu n)^3}$$

We get in the exponent $\big(1 - \frac{1}{20(\lambda\mu n)^2}\big)(\lambda\mu - 1)/12\lambda\mu n$, which is negative for all n, λ (since $1/4 \geq \lambda\mu \geq (n-1)/n^2$). Thus the exponential term is less than 1 and the required inequality follows.

The lower bound is analogous. ∎

What to remember about the Stirling formula: $\boxed{n! \cong (n/e)^n}$.

Let $0 < \lambda < p < 1$.

$$\sum_{i=0}^{\lambda n} \binom{n}{i} p^i (1-p)^{n-i} \cong 2^{-nD(\lambda\|p)}$$

$$\text{(more precisely} \qquad \frac{1}{\sqrt{8n\lambda(1-\lambda)}} 2^{-nD(\lambda\|p)} \le \sum_{i=0}^{\lambda n} \binom{n}{i} p^i (1-p)^{n-i} \le 2^{-nD(\lambda\|p)} \qquad)$$

In particular, for $\omega < p = 1/2$ we obtain for the volume of the ball in the Hamming space (irrespective of its center)

$$\frac{1}{\sqrt{8n\lambda(1-\lambda)}} 2^{nh_2(\omega)} \le \mathrm{vol}(\mathcal{B}_{\omega n}) = \sum_{i=0}^{\omega n} \binom{n}{i} \le 2^{nh_2(\omega)}$$

Note also that $\mathrm{vol}(\mathcal{S}_w) \cong \mathrm{vol}(\mathcal{B}_w)$.

What to remember:

$$\boxed{\mathrm{vol}(\mathcal{B}_{\omega n}) \cong 2^{nh_2(\omega)}}$$

A similar equality holds for the q-ary Hamming space: $\mathrm{vol}(\mathcal{B}_{\omega n}(\mathscr{H}_q^n)) \cong 2^{nh_q(\omega)}$.

Some coding theory problems

Let

$$A(n,d) = \max |\mathcal{C}|$$

be the maximum size of a code $\mathcal{C} \in \mathscr{H}_2^n$ with distance $d = \delta n$. Define

$$R(\delta) = \lim_{n\to\infty} \frac{1}{n} \log_2 A(n,d)$$

Does this limit exist? We do not know (although many believe it does). For the moment think of lim sup or lim inf instead of lim as appropriate. For instance,

$$\overline{R}(\delta) := \limsup_{n\to\infty} \frac{1}{n} \log_2 A(n,d).$$

An elegant observation: $\overline{R}(\delta)$ is a continuous function. We may prove this later; if not see M. Aaltonen, *Notes on the asymptotic behavior of the information rate of block codes*, IEEE Trans. Inform. Theory **30** (1984), no. 1, 84–85.

Problem: Find or bound $R(\delta)$.

By analogy, define $\delta(R)$ as the largest achievable relative distance of a sequence of codes of rate R. The above problem can be reformulated as finding $\delta(R)$ for a given R.

Extremal problems like this one are also studied for all the other parameters of codes introduced above and for many other functions defined on codes. We will see some of these problems later on.

Some of the main goals of coding theory are to construct good sequences of codes: for instance, codes with a high minimum distance, or codes with a low error probability of decoding on a Gaussian channel, binary symmetric channel and other more complicated communication channels while maintaining low complexity of construction and decoding.

Definition 1. *A sequence of codes of asymptotic rate R and relative distance δ is called* asymptotically good *if $R\delta > 0$.*

One of the important goals of coding theory can be formulated as constructing low-complexity asymptotically good codes with simple (e.g., polynomial-time) construction and decoding complexity.

General references for coding theory are [1]-[15]. It is hard to single out one book that will have it all: each of them is good for its own range of topics. Ask me if you want to read about a particular question, I will probably be able to give a reference.

References

1. R. E. Blahut, *Algebraic codes for data transmission*, Cambridge University Press, 2002.
2. T. M. Cover and J. A. Thomas, *Elements of information theory*, John Wiley & Sons, New York e.a., 1991.
3. I. Csiszár and J. Körner, *Information theory. Coding theorems for discrete memoryless channels*, Akadémiai Kiadó, Budapest, 1981.
4. W. C. Huffman and V. Pless, *Fundamentals of error-correcting codes*, New York: Cambridge University Press, 2003.
5. R Johannesson and K. Sh. Zigangirov, *Fundamentals of convolutional coding*, IEEE Press, New York, 1999.
6. F. J. MacWilliams and N. J. A. Sloane, *The theory of error-correcting codes*, 3 ed., North-Holland, Amsterdam, 1991.
7. H. Niederreiter and C. Xing, *Rational points on curves over finite fields*, Cambdridge University Press, 2001.
8. V. Pless, *Introduction to the theory of error-correcting codes*, John Wiley, New York, NY, 1988.
9. V. Pless and W. C. Huffman (eds.), *Handbook of coding theory, vol. 1,2*, Elsevier Science, Amsterdam, 1998, 2169pp.
10. O. Pretzel, *Codes and algebraic curves*, The Clarendon Press, Oxford University Press, New York, 1998. MR **2000c:**11098
11. M. Sudan, *Algorithmic introduction to coding theory*, Lecture notes, MIT Course 6.897, available from http://theory.lcs.mit.edu/~madhu, 2002.
12. M. Tsfasman and S. Vlăduţ, *Algebraic-geometric codes*, Kluwer, Dordrecht, 1991.
13. J. H. van Lint, *Introduction to coding theory*, 3 ed., Springer-Verlag, Berlin e. a., 1999, (1st ed. 1981).
14. A. J. Viterbi and J. K. Omura, *Principles of digital communication and coding*, McGraw-Hill, 1979.
15. S. B. Wicker and S. Kim, *Fundamentals of codes, graphs, and iterative decoding*, Boston : Kluwer Academic Publishers, 2003.