CS 229r Essential Coding Theory Feb 1, 2017

Lecture 3

Instructor: Madhu Sudan *Scribe: Mark Goldstein*

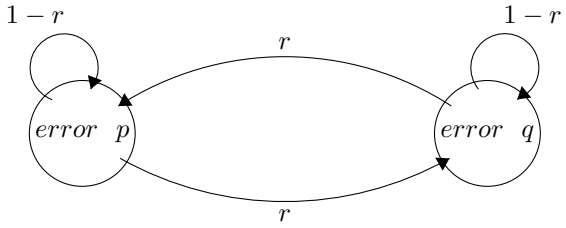
This lecture begins with a brief comparison of the theoretical works of Hamming and Shannon that we have covered so far. Following this, we take a step back and formalize some definitions and terminology that we have used (linear codes, Hamming distance, relative distance). Finally, we take a first step into the asymptotics of rates and distances as our code block length approaches infinity ($n \to \infty$). We explore the Gilbert Greedy Construction of codes and see an example of reduction from code construction to a graph problem, and how using a graphical model instead of a geometric one can allow us to include graph theorems in our toolkit.

1 Hamming versus Shannon

So far we have studied the late 40's and early 50's work of Hamming and Shannon. While Hamming was focused properties of codes under very specific error conditions and gave us some constructive proofs, Shannon was more broad and ambitious in the way that he non-constructively posed his information and coding problems

1. Relationship to codes, encoding (E) functions, and decoding (D) functions.
   - Shannon focused on the existence of good encoding functions.
   - Hamming focused on the code C ($= image(E)$) itself, but not on E nor D.
2. Construction of codes
   - Shannon: abstractly described the communication process as consisting of senders, receivers, compression/decompression, encoding/decoding, a memoryless (acting independently on all bits) communication channel with a certain capacity, and a rate of information flow through that channel.
   - Hamming: Explicit construct of codes with generator and parity matrices, but less of an emphasis of the surrounding context, explicit Hamming Bound on the rate of a code.
3. Error Model
   - Shannon: random errors. Perhaps a bit flips 20% of the time, but we recover with high probability... that's okay!
   - Hamming: "worst case" error model. Code should be robust to all patterns of a bounded # of errors.

1.1 Aside: Shannon + Markov



We don't know the closed form capacity of this channel as a function of p, q, r.

2 Basic Parameters and Terminology

1. **Alphabet** Σ: the set of symbols that make up the message. $q = |\Sigma|$. $\mathbb{F}_q$ might exist. If q is prime $\rightarrow$ $\mathbb{F}_q$ with arithmetic mod q.
2. **Message**: $m \in \Sigma^k$, the information that we want to transmit reliably.
3. **Encoding Function** $E : \Sigma^k \rightarrow \Sigma^n$: Injective mapping from message to codeword.
4. **Code**: Image of E. $\{E(m)|m \in \Sigma^k\}$.

   Which is easier to look at, for the purpose of comparing rates?

   $$\{0,1\}^{1000} \rightarrow (\{0,1\}^8)^{200}$$

   $$(\{0,1\}^8)^{125} \rightarrow (\{0,1\}^8)^{200}$$

5. **Hamming Distance**: $\Delta(\mathbf{x}, \mathbf{y}) = |\{i \in [N]|x_i \neq y_i\}|$ where

   $$\mathbf{x} = (x_1, ..., x_n) \in \Sigma^n$$
   $$\mathbf{y} = (y_1, ..., y_n) \in \Sigma^n$$

   The Hamming distance is only defined over two same-length strings. Formally, it is a metric:

   (a) $\Delta(\mathbf{x}, \mathbf{y}) = 0 \iff \mathbf{x} = \mathbf{y}$
   (b) $\Delta(\mathbf{x}, \mathbf{y}) = \Delta(\mathbf{y}, \mathbf{x})$
   (c) Triangle inequality: $\Delta(\mathbf{x}, \mathbf{z}) \leq \Delta(\mathbf{x}, \mathbf{y}) + \Delta(\mathbf{y}, \mathbf{z})$

   This allows us to think geometrically about codes!

6. **Distance of a code**:

   $$\Delta(C) = \min_{\mathbf{x} \neq \mathbf{y},\ \mathbf{x}, \mathbf{y} \in C} \Big(\Delta(\mathbf{x}, \mathbf{y}) \Big)$$

   The distance of a code captures its "worst case aspects". Imagine that an adversary chooses a message and corrupts it with an undesirable error pattern. How robust can we be? We would like:

   - to correct lots of errors
   - long messages
   - large distance
   - smaller block length is better

7. **Linear Code**: a linear code is a code for which any linear combination of codewords is also a codeword.

   - $\forall \mathbf{x}, \mathbf{y} \in C, (\mathbf{x} + \mathbf{y}) \in C$
   - $\exists \mathbf{G} \in \Sigma^{k \times n}$ such that $C = \{\mathbf{xG}|\mathbf{x} \in \Sigma^k\}$
   - $\exists \mathbf{H} \in \Sigma^{n \times (n-k)}$ such that $C = \{\mathbf{y} \in \Sigma^n | \mathbf{yH} = 0\}$

   where $\mathbf{G}$ is a generator matrix and $\mathbf{H}$ is a parity check matrix.

8. **Code notation**: We specify codes in shorthand with $(n, k, d)_q$ meaning a code with block length n, message length k, distance d, and $q = |\Sigma|$. $C \subseteq \Sigma^n, |C| \geq q^k, \Delta(C) \geq d$. We use square brackets for linear codes: $[n, k, d]_q$

It's obvious that we would like to push n (block length) down, push k (message length) up, and push $\Delta(C)$ (distance of code) up. How about $q = |\Sigma|$? It's not clear! Empirical observations show that it should be small.

3 Brief Reminder: Hamming Codes last time

We saw in the first two classes the following code: $[n, n - \log_2 n, 3]_2$. More generally, Hamming gave to us $[n, n - (q-1)\log_q n, 3]_q$. Hamming is optimal by packing bound.

Aside: consider instead $q = 6$, the first non-prime. How can one work with this? We no longer have arithmetic mod q.

We will now move into asymptotics of rates and relative distances.

4 Asymptotics for fixed q as $n \to \infty$

We move on to a brief preview to the kinds of bounds-oriented work we will explore this semester. We define the **rate** and **relative distance** of codes, go through the Gilbert Greedy Code Construction (exponential time in n) and the Gilbert (lower) Bound for the size of a code, and consider a reduction to a graph problem that tightens the bound.

4.1 Some definitions

Rate of a Code: The rate of a code, $R(C)$, is defined as $\frac{k}{n}$.

Relative Distance: Normalized by the codeword block length n, the relative distance, $\delta(C)$ allows us to compare the distances of various block-lengthed codes. It is defined as $\frac{\Delta(C)}{n}$.

$$0 \leq R(C), \delta(C) \leq 1$$

Hamming Ball: $Ball(\mathbf{v}, r) = \{\mathbf{x} \in \Sigma^n | \Delta(\mathbf{x}, \mathbf{v} \leq r\}$

Volume of a Hamming Ball: $Vol(n, r) = |Ball(\mathbf{v}, r)|$

A **Constructable** code is one for which there is a known polynomial-time encoding procedure. **Non-constructive** codes are shown to exist and may have known exponential time encoding algorithms. Consider the example of non-deterministically guessing an encoding matrix, much like the operation of non-determinstic Turing Machines on **NP** problems. Before Shannon, there was no evidence that one could find a code with both $R(C)$ and $\delta(C) > 0$.

4.2 Gilbert Greedy Construction

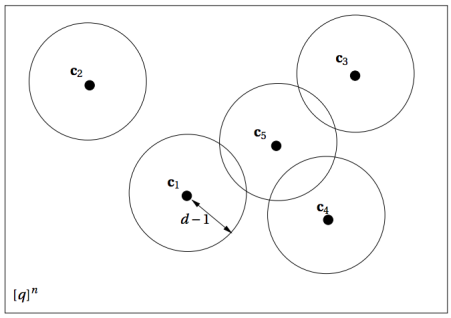
For this construction and for the Gilbert Bound below, fix δ, interpret n as large, $d = \delta n$

This code construction procedure, demonstrated by **Gilbert**, achieves $R, \delta > 0$ in exponential time by building $C \subseteq \{0,1\}^n$ greedily. **Algorithm 5 from Chapter 4 in the textbook:**

Algorithm 5 Gilbert's Greedy Code Construction

INPUT: n, q, d
OUTPUT: A code $C \subseteq [q]^n$ of distance d

1: $C \leftarrow \emptyset$
2: WHILE there exists a $\mathbf{v} \in [q]^n$ such that $\Delta(\mathbf{v}, \mathbf{c}) \geq d$ for every $\mathbf{c} \in C$ DO
3: Add $\mathbf{v}$ to C
4: RETURN C



Our Implementation from Class:

$C \leftarrow \emptyset$, $S \leftarrow \{0,1\}^n$
while $S \neq \emptyset$ **do**
let $\mathbf{v} \in S$
$C \leftarrow C \cup \mathbf{v}$
$S \leftarrow S \setminus Ball(\mathbf{v}, d-1)$
end.

Notice that you can add codewords $\mathbf{v}$ still in S at any point during the procedure, even though some codewords in $Ball(\mathbf{v}, d-1)$ may have already been discard. This means that we can have overlapping balls, as long as no codeword itself is in another codeword's ball.

claim $\Delta(C) \geq d$ (radius of the balls in the algorithm $+1$)

claim $|C| \geq \frac{2^n}{|Ball(\mathbf{x} \in C, d-1)|} \approx 2^{n(1-H(\delta))}$

theorem $\exists$ codes with $R \approx 1 - H(\delta)$

4.3 Gilbert Bound

$$\bigcup_{\mathbf{v}\in C} Ball(\mathbf{v}, d-1) = \{0,1\}^n$$

$$|C|Vol(n, d-1) \geq 2^n$$

$$|C| \geq \frac{2^n}{Vol(n, d-1)}$$

$$(Vol(n, \delta n) \approx 2^{H(\delta)n})$$

$$|C| \geq 2^{n(1-H(\delta))}$$

This is a good lower bound on $|C|$. But, can we do better?

4.4 Code as Independent Set of a Graph

We note that codes of distance d correspond to "independent sets" (a set of vertices, no pair of which are adjacent), in the following graph $G_{n,d} = (V, E)$:

$$V = \{0,1\}^n$$
$$E = \{(\mathbf{u}, \mathbf{v}) | \Delta(\mathbf{u}, \mathbf{v}) \leq d-1\}$$

An old result due to Turan says that if a graph has N vertices with maximum degree $\leq D$, there it has an independent set of size at least $\frac{N}{D+1}$. Applied to our problem this gives the Gilbert bound since in our graph $N = 2^n$ and $D = Vol(n, d-1) - 1$ (and indeed Gilbert's construction is identical to the Turan construction).

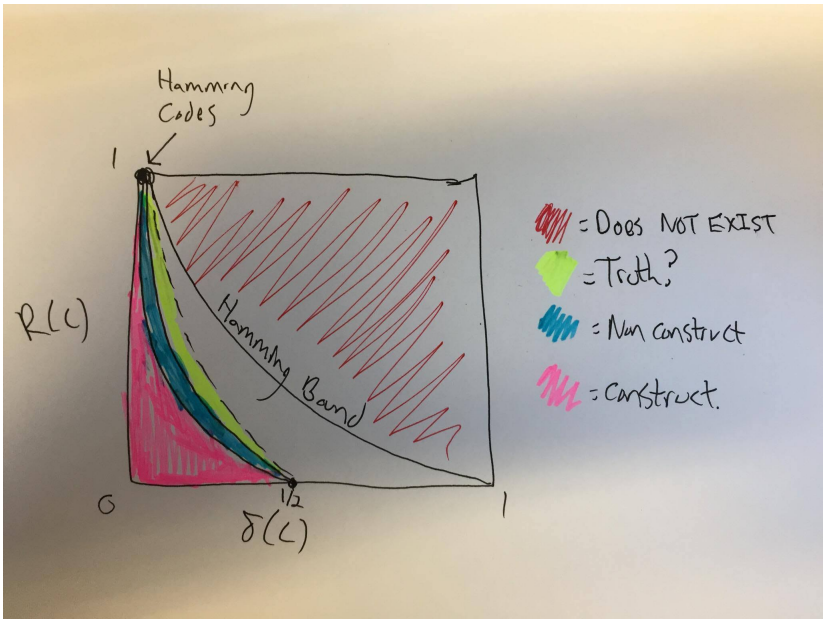
In 2004, Jiang and Vardy [Jiang-Vardy] managed to give an asymptotic improvement over the Gilbert bound, using the structure of $G_{n,d}$. They prove that $G_{n,d}$ has few triangles (cliques of three vertices), and then apply a result from combinatorics (see, for instance, [Bollobas, page 296]) that graphs with "few" triangles have independent sets of size $\Omega\left(N \cdot \frac{\log D}{D}\right)$. This, in turn, is built on a result of Ajtai, Komlos, and Szemerdi which shows roughly the same result for graphs with no triangles. How many trangles is few? The naive upper bound on the number of triangles in a graph with N vertices and degree at most D is $O(ND^2)$ (there are N choices for the first vertex of the triangle, and at most D choices each for the second and third vertex, since they must be adjacent to the first). Turns out any $o(ND^2)$ bound on the number of triangles is a good enough definition of "few"; and indeed Jiang and Vardy do show that the number of triangles is $o(ND^2)$ (see exercise below) and thereby conclude that there is a code $C \subseteq 0, 1^n$ of distance d of size at least $\log(Vol(n, d-1)) \cdot 2^n / Vol(n, d-1) \geq d \cdot 2^n / Vol(n, d-1)$.

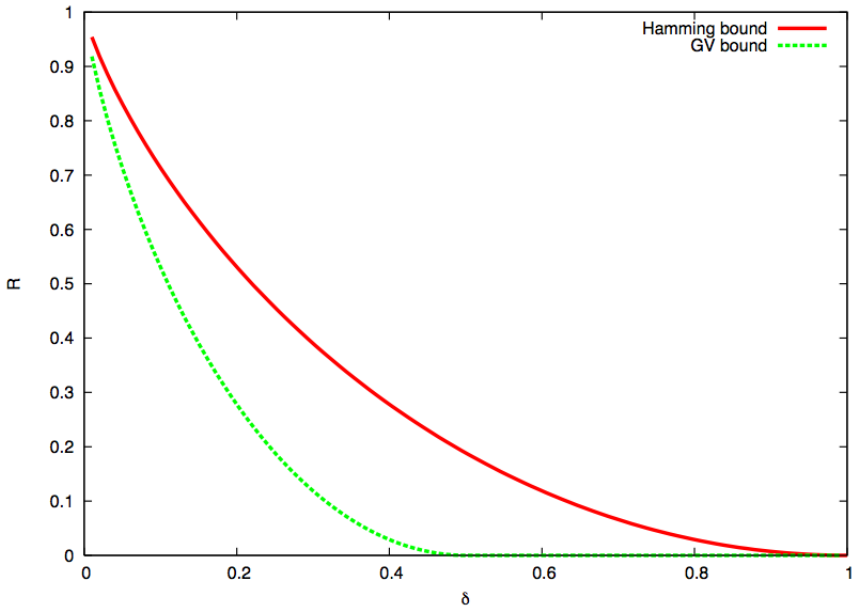
Exercise 1. *1. Fix $\delta \in (0, 1/2)$ and let n be a growing number. Let v, w be two random vectors in $0, 1^n$ drawn independently such each coordinate of v and w is 1 with probability δ and 0 otherwise. Prove that the probability that $\delta(u, w) \leq \delta$ is $o(1)$.*

2. Let δ and n be as above and let $d = \lfloor \delta n \rfloor$. Prove that for a random vertex $u \in G_{n,d}$ and a random pair of neighbors $v, w \in G_{n,d}$ of u, the probability that v is adjacent to u is $o(1)$. Conclude that the number of triangles in $G_{n,d}$ is $o(ND^2)$ where $N = 2^n$ and $D = Vol(n, d-1) - 1$.

Next time: Varshamov bound with $d-2$ ball. Together with Gilbert, these are known as the **GV Bound**.

4.5 Next Time: The Relationship of Rates and Relative Distances





References