CLASS NOTES 1: FUNDAMENTALS OF ERROR CORRECTION CODING

Introduction to Information and Coding Theory

Suayb S. Arslan

This short note introduces some of the basic concepts of Information and coding theory. Once the concepts are defined, some asymptotical bounds on the straightforward code constructions are presented without worrying about the encoding and decoding procedures and their implementation complexity.

1 TRANSMISSION SYSTEMS

Transmission is the act of the process of sending out electronic signals such as radio signals, or messages in some form so that a receiver can receive information and reconstruct the message in an appropriate location or time. In its most conventional form, for example, radio signals sent out over mobile networks is an example of transmission process in space. Storing information in an appropriate storage media is an example of transmission in time. In either way, a way of communication is realized. One of the fundamental problems of such communication idea is the required reliability. Because of this stringent requirement, modern communication and storage systems rely heavily on powerful channel coding methodologies.

In order to realize channel codes that have got good error detection and correction capabilities, we need to go through the theoretical development stages of coding. Such construction methodologies should also pave the way for efficient implementation strategies. For this, a brand new era of mathematical discipline has been commenced with Shannon's seminal work on what is now called the "Information theory". This theory establishes the limits of communication that allow originating a message and its lossless reconstruction in another point in time or space. This class is therefore concerned with information theory fundamentals that shed interesting light to the theory of channel coding.

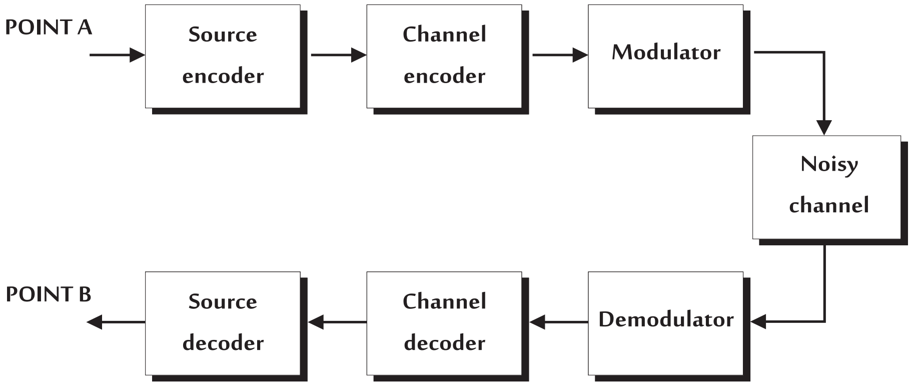

Figure 1.1: Basic system diagram for a communication or a storage system.

1.1 System Diagram

A general system block diagram is shown in Fig. 1.1. A communication system consists of three major building blocks that process the information generated at point A. Raw data first goes through source coding, i.e., the unstructured redundancy is thrown out for saving space. This process is also known as “compression". The compressed content then goes through channel coding where a structured redundancy is added to the message symbols to create some robustness against channel errors. Introduction of structure in to the “redundancy" lends itself to an easier interpretation of the use of it at the receiver with respect to detect and correct as many errors as possible. Based on the channel frequency response characteristics, the modulator block generates the appropriate signal that is most suited for transmission. The symbols transmitted across the channel are received and processed exactly in reverse order in order to reconstruct a copy of the message at point B that is originally generated at point A.

In his seminal work, Claude E. Shannon showed that it is theoretically possible to construct an information transmission system that allow exponentially decreasing error probability in terms of message reconstruction (Shannon, 1948). The prerequisite for this was that the information rate of the information source is smaller than the so-called channel capacity. The term “capacity" will be clear after we introduce some prerequisites of information theory. Given the statistical nature of the channel, capacity can be thought as a fixed number beyond which the information communication is no longer reliable no matter how much clever or much redundancy is applied to the transmitted raw data. In order to reduce the information rate, source coding is used which are implemented by the source encoder in the transmitter and the source decoder at the receiver.

1.2 Information Theory

In order talk about concepts like sets and their relative size, we need the idea of “measure". For example from classical probability theory, a measure on a set is a systematic way to assign a number to each suitable subset of that set. This can be intuitively interpreted as its size. A measure therefore is a generalization of the concepts of length, area, or volume. In this sense, the idea of information is also subject to some form of “measure" to talk about its content. According to Shannon, messages are emitted from information sources that are of random nature. The measure of information is thus interpreted based on the statistics of the symbols emitted by that information source.

1.2.1 Entropy

Entropy is the uncertainty within a given information source that emits infinite or finite number of information symbols according to a probability distribution $p_X(x)$. More formally, it is a measure of randomness (uncertainty) that is defined as

$$H(X) = -\int_x \left(\sum_x\right) p_X(x) \log p_X(x) \tag{1.1}$$

with the minor convention that $0 \log 0 = 0$. If the information source emits binary symbols $x_i \in \{0,1\}$ for $i = 1,2$ with probabilities p_0 and $1 - p_0$, respectively, is given by the binary entropy function

$$h(x) = -\sum_x p_X(x) \log p_X(x) = -p_0 \log_2 p_0 - (1-p_0)\log_2(1-p_0) \tag{1.2}$$

where it attains its maximum value of 1 when $p_0 = 1/2$. The entropy is a concave function of the distribution and equals zero or one as the measure of entropy must be 0 when the source is deterministic. The significance of entropy is also realized in Shannon's source coding theorem. It turns out that the entropy of a random source is a lower bound on the the average number of symbols required to represent it without incurring any information loss.

The definition of entropy for a single random variable (single source) can easily be extended to multiple random variables $X_1, X_2, \dots, X_M$ and hence we have the following joint entropy,

$$H(X_1,\dots,H_M) = -\int_{x_1} \dots \int_{x_M} \left(\sum_{x_1} \cdots \sum_{x_M}\right) p_{X_1,\dots,X_M}(x_1,\dots,x_M) \log p_{X_1,\dots,X_M}(x_1,\dots,x_M) \tag{1.3}$$

For a given two information sources X and Y with the joint probability distribution $p_{X,Y}(x,y)$, the *conditional entropy* is the entropy of a random variable Y conditioned on the availability of another random variable's (X) knowledge and is defined as

$$H(Y|X) = \int_x \left(\sum_x\right) p_X(x) H(Y|X=x) \tag{1.4}$$

$$= -\int_x \left(\sum_x\right) p_X(x) \int_y \left(\sum_y\right) p_{Y|X}(y|x) \log p_{Y|X}(y|x) \tag{1.5}$$

$$= -\int_x \int_y \left(\sum_x \sum_y\right) p_{X,Y}(x,y) \log p_{Y|X}(y|x) \tag{1.6}$$

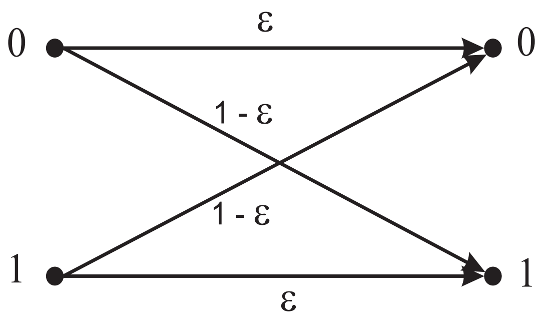

Figure 1.2: A binary symmetric channel with error (binary flip) probability ϵ.

The amount of reduction in the uncertainty about a random variable X due to the information presence of another random variable Y is called *mutual information.* It is conventionally represented by $I(.)$ notation and given by

$$I(X;Y) = H(X) - H(X|Y) = \int_x \int_y \left(\sum_x \sum_y \right) p_{X,Y}(x,y) \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \tag{1.7}$$

Mutual information is a measure of the dependence between the two random sources and equals zero when X and Y are mutually independent.

1.2.2 Channel Capacity

The channel block in a typical communication system shown in Fig. 1.1 is characterized by a conditional probability distribution $p_{Y|X}(y|x)$ of the output Y given the input X. The so-called capacity of the channel C is given by the maximization of the mutual information with respect to input probability distribution $p_X(x)$. In other words,

$$C = \max_{p_X(x)} I(X;Y) \tag{1.8}$$

where the maximum is taken over all the distributions on X. If the rate of transmission is lower than or equal to this number, the it shown (at least mathematically) that the reliable transmission of information is possible. In other words, this significant result advocates that as long as the rate of transmission is below a threshold (the channel capacity), a vanishingly low probability of error is guaranteed.

Exercise 1: Consider a source X with entropy $H(X)$. Is a reliable communication possible if $H(X) > C$? Comment on the implications of this inequality.

Exercise 2: Consider a source X transmitted over a binary symmetric channel (BSC) with parameter ϵ, given in Fig. . The output of the channel is Y. Show that the capacity of this channel is given by $1 - h(\epsilon)$.

PROOF: First observation is that BSC is one of the important discrete memoryless channels. In addition it is symmetric with respect to the input symbols. Therefore, it is relatively easy to see the following using Equation (1.4),

$$H(Y|X) = p_0 h(\epsilon) + (1 - p_0) h(\epsilon) = h(\epsilon) \tag{1.9}$$

where the symmetry leads to the same binary entropy function $h(\epsilon)$ whether we condition on $X = 0$ or $X = 1$. The mutual information between X and Y is defined to be

$$I(Y;X) = H(Y) - H(Y|X) = H(Y) - h(\epsilon). \tag{1.10}$$

Since this expression attains its maximum whenever $H(Y)$ is maximum, the capacity is $1-h(\epsilon)$. Note that $H(Y)$ is maximum whenever the probability distribution $p_Y(y)$ is uniformly random. It is left to the reader to check X is uniformly distributed if and only if Y is uniformly distributed. Thus, this completes the proof. □

Although information theory let us know about something quite brilliant that it is theoretically possible to find a channel codes that for a given channel leads to vanishing error probabilities as required, the design of good and efficient channel codes is generally difficult problem to tackle.

1.2.3 A simple channel code: Repetition code

A binary message stream is transmitted through a binary symmetric channel that causes seldom bit flips (errors). The fundamental operation of coding is to add extra check symbols to the message stream so that the errors can be efficiently found and corrected at the receiver.

Let us assume that a message symbol z consists of 30 bits. Before transmitting that information over the channel, we duplicate each bit an odd number z_d times. This operation is known as "encoding" the message symbol stream. After the channel a subset of $30z_d$ bits are in error where the size of the subset is a function of the channel characteristics. At the decoder, if the majority of duplicated bits is $i \in \{0,1\}$, the transmitted bit is decoded to be $i \in \{0,1\}$. Such an encoding/decoding pair with the associated code is known as the "repetition" code. Note that this simple code can correct up to $(z_d - 1)/2$ bits.

1.3 Basics of Channel Codes

A binary code C of size $|C|$ and blocklength n is a set of $|C|$ binary words, $\{c_1, \ldots, c_{|C|}\}$ of length n each called *codewords*. The set of all the codewords is known as the *codebook*. In most practical cases, $|C| = 2^k$ and the coded is referred to as a (n,k) binary code. The first job is to create a codebook to choose the codewords from. the next operation is the map all the binary message sequences with appropriate codewords. This mapping in fact defines the encoding operation. The following is an example of a codebook and the mapping function of message symbols.

$$\left\{\begin{matrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{matrix}\right\} \begin{matrix} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{matrix} \left\{\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{matrix}\right\} \begin{matrix} \rightarrow \\ \rightarrow \\ \rightarrow \\ \rightarrow \end{matrix} \left\{\begin{matrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{matrix}\right\} = C \tag{1.11}$$

Finally, a decoding procedure is to undo the encoding operation after the channel after deciding on one of the codewords which we believe transmitted. We already have seen one example in the previous section that majority of bit labeling determines the decoding decision on the transmitted data bits. If the received codeword is none of the codewords in the codebook, then one of the codewords should be picked before undoing the encoding operation. One of the obvious choice is to choose codeword from the codebook that is closest to the received word in some distance measure such as Euclidian distance. Here from implementation point of view, one can find simpler definitions of distance measure such as *Hamming distance* we define next. Although we stick to binary codes throughout for simplicity, we try to keep our definitions general.

Definition 1: *The Hamming distance between two sequences x and y of length n, whose elements are chosen from a finite set Ω_q of size q, is denoted as $D(x,y)$ and is defined to be the number of positions in which x and y differ. The fractional Hamming distance is the ratio of $D(x,y)$ to the block length n i.e., $\gamma(x,y) = D(x,y)/n$.*

One can verify that Hamming distance is nonnegative and symmetric. Moreover, it satisfies the triangular inequality and therefore it defines a metric on Ω_q^n.

Definition 2: *For a code $C = \{c_1, \dots, c_{|C|}\}$., minimum Hamming distance $D_{\min}$ of C is defined to be the smallest Hamming distance between all possible codeword pairs, each of length n symbols. In other words,*

$$D_{\min}(C) = \min_{c_i, c_j \in C, i \neq j} D(c_i, c_j) \tag{1.12}$$

Similarly, the fractional minimum Hamming distance of the code C is given by $D_{\min}(C)/n$.

Exercise 3: Note that Hamming distance is defined between two q-ary sequences. In this exercise we will take the first step to generalize this definition. Suppose that instead of considering two codewords, we consider three codewords at the same time i.e., x_1, x_2, y and assume that each codeword being mapped equally likely. Let us define $D_2(\bar{x}, y) = [D(x_1, y); D(x_2, y)]$ and S to be the covariance matrix that defines the correlation between $\bar{x} = [x_1, x_2]$ and $\bar{y} = [y, y]$ vectors. We define *3-Hamming distance* as

$$D^3(\bar{x}, y) = \sqrt{D_2(\bar{x}, y)^T S^{-1} D_2(\bar{x}, y)} \tag{1.13}$$

Show that *3-Hamming distance* is a metric and it reduces down to conventional definition of Hamming distance when we consider only two codewords at a time.

Hint: Since codewords are independently generated and equally likely, $S = [1/2,\ 0; 0,\ 1/2]$.

Exercise 4: What is the minimum distance of the code shown in Eqn. (1.11).

Definition 3: *The code rate of a code $C \subseteq \Omega_q^n$ is defined by*

$$R_C = \frac{\log_q |C|}{n \log_q |\Omega_q|} = \frac{\log_q |C|}{n} \tag{1.14}$$

Note that for a $(n, k, D_{\min}(C))$ block code we $|C| = q^k$, then the rate of the code is given by $R_C = k/n$.

As can be obvious by now that the minimum distance of the code C should have something to do with its error correction capability. Before explaining the relationship, we need to define the concept of *Hamming sphere* to be able to define precisely how the decoder establishes the rule for deciding which codeword is transmitted.

Definition 4: *Given a codeword $x \in C$, the Hamming sphere of radius r around x is defined to be the set of all codewords that are in Hamming distance of r or less to x i.e.,* $\{\forall y \in C \subseteq \Omega_q^n | D(x,y) \leq r\}$.

Let us think of such spheres geometrically filling the space of all possible q-tuples i.e., Ω_q^n. Our objective is to draw spheres around each codeword such that no sphere has intersection with another sphere and each sphere has the largest radius. For a given code C with minimum distance $D_{\min}(C)$ or $D_{\min}$ for short, in order to make sure no sphere intersects, we must have equal size spheres of radius $\lfloor \frac{D_{\min}-1}{2} \rfloor$. This is because if we allow radius of $D_{\min}/2$, there will be at least two spheres touching eachother. In order to make sure they do not touch, we subtract 1 from $D_{\min}$ and divide by two to find the greatest radius that will allow no intersection between Hamming spheres. This discussion leads to the following *Lemma*.

Lemma 1: *If C can correct up to t errors, $D_{\min}(C) = 2t+1$. This code can also detect $2t$ errors or erasures. Here the "erasure" means that the location of errors are known and error values must be computed. Any subset of codeword symbols of size $2t$ can be corrected.*

If the received codeword contains more than t symbol errors, one of the two things can happen: The received codeword will either be closer to some other codeword and decoded wrong (this is also called decoding error or miscorrection) or it will be in none of the hamming spheres we draw although it might be closer to only one of the codewords. Since hamming spheres do not necessarily fill up the whole space, the received word may well fall in the interstitial space between hamming spheres. In this case, a possibility could be to declare "decoding failure" and the decoder outputs a flag indicating that the received codeword has so many errors that the decoding cannot be completed. This type of decoding is known as *Bounded Distance Decoder* and are easier to implement compared to a generic *Complete Decoder* which never flags any decoding failure. The latter type might be preferable in terms of performance at the expense of increased complexity.

The ultimate error correction capability of a code is given by the covering radius. The covering radius is related to the maximum weight of a random error vector that can be corrected by the code. This definition of code's power allows correction beyond half the minimum distance of the code of interest. The covering radius of a block code C of length n is defined as the smallest integer r such that all vectors in the containing space are within Hamming distance r of some codeword i.e.,

$$Co(C) = \max_{x \in \Omega_q^n} \left\{ \min_{c \in C} \{D(x,c)\} \right\} \tag{1.15}$$

Unfortunately, finding the covering radius of a code is NP-hard. There are efficient upper and lower bounds on the covering radius however, exact formulations are still open problems of the classical coding theory.

1.4 Bounds for Code Constructions

As we have defined Hamming spheres for code $C \subseteq \Omega_q^n$, it is natural to wonder about the size of the code for a given error correction capability t.In other words, one desirable property of code would be to have high rate i.e., low redundant information added to the data, on the other hand high rate means low error correction capability because it will be harder to pack so many spheres in Ω_q^n. Thus, the natural upper bound on the size of the code C arises and is given by

Lemma 2 (Hamming bound): *Let $C \subseteq \Omega_q^n$ be a code with minimum Hamming distance $D_{\min}$. Then,*

$$|C| \leq \frac{q^n}{\sum_{j=0}^{\lfloor \frac{D_{\min}-1}{2} \rfloor} \binom{n}{j}(q-1)^j} \tag{1.16}$$

PROOF: First note that the Hamming spheres do not overlap if the sphere radius $r \leq \lfloor \frac{D_{\min}-1}{2} \rfloor$ as we argued before. It is easy to compute the volume of a Hamming sphere with radius r to be of the form

$$Vol_q(n,r) \triangleq \sum_{j=0}^{r} \binom{n}{j}(q-1)^j \tag{1.17}$$

Therefore, since the spheres do not have anything in common, $|C| \leq q^n / Vol_q(n, D_{\min})$. □

Any code construction that achieves this bound with equality are called *perfect codes*. We will give few examples of perfect codes in this class, but they are not abundant. Let us give a greedy code construction approach that allows us to have a code with a minimum distance $D_{\min}$. Let us start with any sequence $x_0 \in \Omega_q^n$ and put it in C. Next, choose another sequence x_1, if any, which has a distance at least $D_{\min}$ from x_0. Similarly choose x_2 which has a distance at least $D_{\min}$ from both x_0 and x_1. This approach ceases if we are no longer able to find sequences having a distance at least $D_{\min}$ from all the valid sequences already chosen to be in C. Suppose we draw hamming spheres of radius $D_{\min} - 1$ centered around the codewords of C so that no sphere contains more than one codeword. It can be verified that the union of all these spheres contains the set Ω_q^n. It that was not the case there would be some interstitial space between such spheres. In other words, there would be sequences in that space that are at least distance $D_{\min}$ from all the codewords of C. In that case, our greedy approach would have not terminated. Since we the greedy construction method halts, there remains no interstitial space between hamming spheres of radius $D_{\min} - 1$ centered around the codewords of C.

Lemma 3 (Gilbert-Varshamov bound): *Let $C \subseteq \Omega_q^n$ be a code consisting of $q-ary$ sequences of length n with minimum Hamming distance $D_{\min}$. Then, the largest size of the code C is lower bounded by*

$$|C| \geq \frac{|\Omega_q^n|}{Vol_q(n, D_{\min}-1)} = \frac{q^n}{\sum_{j=0}^{D_{\min}-1} \binom{n}{j}(q-1)^j} \tag{1.18}$$

PROOF: The proof follows from our previous discussion.

Exercise 5: For a prime power q, show that there exists (n,k) code C with minimum distance d with

$$k \geq n - \left\lfloor \log_q \left(\sum_{j=0}^{d-2} \binom{n-1}{j} (q-1)^j \right) \right\rfloor - 1. \tag{1.19}$$

Let us suppose that $|C| = q^k$ for some prime power q and $k \le n$. Using Gilbert-Varshamov bound, we can find a realizable range of values for code rate R_C. By replacing C with q^k in Eqn. (1.18) and taking $\log_q$ of both sides yields the bound $k \ge n - \log_q\left(\lfloor Vol_q(n, D_{\min} - 1)\rfloor\right)$. On the other hand, using Hamming bound, we have $k \le n - \log_q\left(\lfloor Vol_q(n, \lfloor (D_{\min} - 1)/2\rfloor)\rfloor\right)$. Therefore using the definition of rate R_C, we have

$$1 - \frac{1}{n}\log_q\left(\lfloor Vol_q(n, \lfloor (D_{\min} - 1)/2\rfloor)\rfloor\right) \ge R_C \ge 1 - \frac{1}{n}\log_q\left(\lfloor Vol_q(n, D_{\min} - 1)\rfloor\right) \tag{1.20}$$

Next, we give a result from combinatorial probability about the relationship between entropy function and hamming balls that will be useful for finding asymptotical bounds $n \to \infty$.

Lemma 4: *For positive integer n and $q \ge 2$, and a real number $\gamma \in [0, 1 - 1/q]$*

$$q^{nh_q(\gamma) - o(n)} \le Vol_q(n, \gamma n) \le q^{nh_q(\gamma)} \tag{1.21}$$

where the entropy function is given by

$$h_q(x) = x\log_q(q-1) - x\log_q x - (1-x)\log_q(1-x). \tag{1.22}$$

As $n \to \infty$, we will have $Vol_q(n, \gamma n) \to q^{nh_q(\gamma)}$ where $\gamma = D_{\min}/n$ is simply the fractional minimum distance we introduced earlier. This result also implies the following asymptotical bound on rate of the code:

$$R_C \ge 1 - h_q(\gamma) - o(1) \tag{1.23}$$

Using a similar approach we can compute the asymptotical Hamming upper bound on rate as follows. The details are left as an exercise.

$$R_C \le 1 - h_q(D_{\min}(C)/2n) + o(1) = 1 - h_q(\gamma/2) + o(1) \tag{1.24}$$

Before making an interesting observation about achieving the asymptotic Hamming bound, we remember the famous Chernoff bound from classical probability theory and consider binary codes for simplicity i.e., $q = 2$,

Lemma 5: (Chernoff bound for i.i.d. Bernoulli random variables) *Let $X_1, X_2, X_3, \ldots, X_m$ are independent identically distributed binary random variables with $Pr(X_i = 1) = p$, then for all ϵ and large n, we have the following relationships*

$$Pr\left(\sum_{i=1}^{n} X_i \ge (p+\epsilon)n\right) \le 2^{\frac{-\epsilon^2 n}{3}} \tag{1.25}$$

$$Pr\left(\sum_{i=1}^{n} X_i \le (p-\epsilon)n\right) \le 2^{\frac{-\epsilon^2 n}{3}} \tag{1.26}$$

This simply means that for large n, probability of having pn is close to one. In order to correct so many errors, we need at least a code with minimum distance $D_{\min} = 2pn$. Thus, asymptotically we have the upper bound on rate $R_C \le 1 - h(2pn/2n) = 1 - h(p)$. This result simply says that if a code achieves the hamming bound, then it also achieves the capacity of the BSC with error probability p. This simply proves the existence of codes that can achieve the capacity for BSC.

Lemma 6 (Singleton bound): *Let $C \subseteq \Omega_q^n$ be a code consisting of $q-ary$ sequences of length n with minimum distance $D_{\min}$, then, $|C| \leq q^{n-D_{\min}+1}$.*

PROOF: Observe that we already have an upper bound by Hamming. Singleton bound is another upper bound and is tight for codes with large alphabet sizes. We will later see that for binary codes, Hamming bound is a much more improved upper bound. Suppose $|C| > q^{n-D_{\min}+1}$, since each codeword is unique sequence there must be at least two codewords which agree in some $n-d+1$ locations. Therefore, the Hamming distance between these codewords is $n-(n-d+1) = d-1 < d$. This contradicts the hypothesis that code C has a minimum distance d. □

This result also implies that codes must asymptotically satisfy $R_C \leq 1 - D_{\min}/n + o(1)$. Singleton bound does not assume that q and n are chosen independently. It is valid for any q, n and d. Another observation is that in the derivation of asymptotical result for GV bound, we assumed $0 < \gamma < 1 - 1/q$. This was something needed for the upper bound on the volume $Vol_q(n,\gamma n) \leq q^{h_q(\gamma n}$. So one can wonder is it possible to have positive rate and $\gamma > 1 - 1/q$ at the same time. This question is answered by the Plotkin bound.

Lemma 6 (Plotkin Bound): *Let $C \subseteq \Omega_q^n$ be a code consisting of $q - ary$ sequences of length n with fractional minimum distance γ, then*

$$if \gamma = (1-1/q), \quad |C| \leq 2qn \tag{1.27}$$

$$if \gamma > (1-1/q), \quad |C| \leq \frac{q\gamma}{q\gamma - q + 1} \tag{1.28}$$

PROOF: The proof of Plotkin bound can be found in every coding theory book and therefore it will be skipped here.

The implications of this bound are significant. Using the definition of R_C for $\gamma \geq (1-1/q)$, we have

$$R_C = \frac{\log_q |C|}{n} \leq \begin{cases} \frac{\log_q 2qn}{n} & \text{if } \gamma = (1-1/q) \\ \frac{\log_q\left(\frac{q\gamma}{q\gamma-q+1}\right)}{n} & \text{if } \gamma > (1-1/q) \end{cases} \tag{1.29}$$

In both cases as n tends to infinity, the upper bound converge to 0 making us conclude $R \leq 0$. Since rate is non-negative quantity we conclude $R = 0$ for $\gamma \geq 1 - 1/q$.

Let partition the code space defined by the code C by grouping the codewords which have the first $n-m$ symbol locations agree. For example for $x-th$ partition, the codewords within that partition can be named as $C_x \subseteq C$ and $\bigcup_x C_x = C$. Since codewords agree on the first $n-m$ symbol locations, the minimum distance of C_x must be equal to the minimum distance of code C, $D_{\min}$. The block length of codewords of C_x is selected to be $m < \frac{q}{q-1} D_{\min}$, so that every possible $x \in \Omega_q^{n-m}$ will be used and the number of partitions are q^{n-m}.

Since $\frac{D_{\min}}{m} > \frac{q-1}{q}$, we have

$$|C_x| \leq \frac{q(D_{\min}/m)}{q(D_{\min}/m) - q + 1} \leq qD_{\min} \tag{1.30}$$

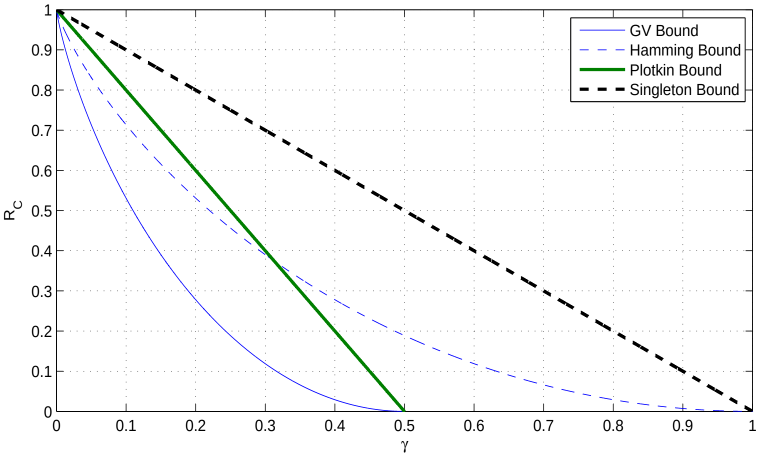

Figure 1.3: For $q = 2$, bounds on the rate R_C as a function of fractional minimum distance γ.

where the first inequality is due to Plotkin upper bound and the second inequality is because $qD_{\min} - (q-1)m$ is an integer. Since partitions are disjoint, we finally have

$$|C| = \sum_{x\in\Omega_q^{n-m}} |C_x| \leq \sum_{x\in\Omega_q^{n-m}} qD_{\min} = q^{n-m}.qD_{\min} \leq q^{n-\frac{q}{q-1}D_{\min}+o(n)} \tag{1.31}$$

Considering R_C and dividing the both sides of the inequality by n will yield

$$R_C \leq 1 - \left(\frac{q}{q-1}\right)\gamma + o(1) \tag{1.32}$$

Note that this is a better bound than singleton bound particularly for small values of q. Another way of stating this comparison is that if q is a fixed number and is independent of n, then such class of codes cannot achieve the singleton bound. Fig. 1.3 depicts all the bounds we discussed in this class for binary codes i.e., $q = 2$. similarly, the same bounds are drawn for $q = 256$ in Fig. 1.4 . As can be noticed that with large q, Singleton and plotkin bounds converge and Hamming bound become loose compared to singleton and Plotkin bounds.

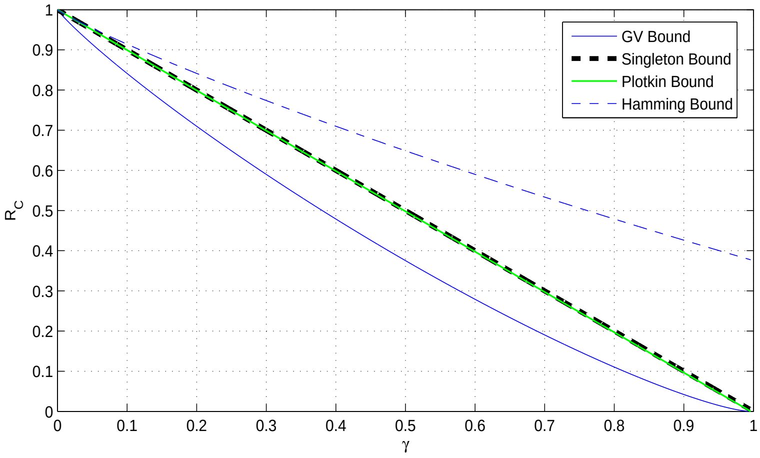

Figure 1.4: For $q = 256$, bounds on the rate R_C as a function of fractional minimum distance γ.