Week 2

Parameters. Bounds

Version 2023-10-01. To accessible online version of this chapter

Synopsis. *Basic properties of a code C can be expressed by numbers called* **parameters.** *We learn why such parameters as the* **rate**, *R, and the* **minimum distance**, *$d(C)$, are important when C is used for channel coding. We also learn to use the notation $(n, M, d)_q$ and $[n, k, d]_q$. It turns out that there is a trade-off between the rate and the minimum distance: both cannot be high (good) at the same time. This trade-off is expressed by inequalities known as* **bounds.** *We only prove the Hamming bound and the Singleton bound in this course, although other bounds have been obtained in coding theory research.*

Parameters of a code

Parameters are numerical characteristics of a code. The most important parameters are:

Definition: parameters of a code

Let F be an alphabet and $C \subseteq F^n$ be a code. Then:

- q denotes the **size of the alphabet,** i.e., $q = \#F$;
- n is called the **length** of the code — each codeword consists of n symbols;
- M denotes the **number of codewords** in the code, i.e., $M = \#C$;
- $k = \log_q M$ is the **information dimension** of C;
- $d(C) = \min\{d(\underline{v}, \underline{w}) : \underline{v}, \underline{w} \in C, \underline{v} \neq \underline{w}\}$ is the **minimum distance** of C;
- $R = k/n$ is the **rate** of C;
- $\delta = d/n$ is the **relative distance** of C.

We say that C is **an** $(n, M, d)_q$**-code** or **an** $[n, k, d]_q$**-code**.

The importance of the minimum distance for error detection and correction

We will now see that the higher $d(C)$, the more symbol errors per codeword is the code C guaranteed to detect and correct.

Notation: $[a]$ denotes the integer part of a real a; e.g., $[3] = [3.5] = [\pi] = 3$, $[7.99] = 7$.

Theorem 2.1: the number of errors detected/corrected by a code

Let C be a code with $d(C) = d$. Throughout the course, t will denote $[(d-1)/2]$. Let $\underline{v} \in C$ and $\underline{y} \in F^n$.

1. If $1 \leq d(\underline{v}, \underline{y}) \leq d-1$, then $\underline{y} \notin C$. Thus, if at most $d-1$ errors occur in a transmitted codeword, they will be *detected*.
2. If $d(\underline{v}, \underline{y}) \leq t$, then $\underline{y}$ has a unique nearest neighbour in C, which is $\underline{v}$. So if at most t errors occur in a codeword, a decoder will *correct* them by decoding $\underline{y}$ back to $\underline{c}$.

Proof. 1. If $\underline{y} \in C$ then by definition of minimum distance, either $d(\underline{v}, \underline{y}) = 0$ or $d(\underline{v}, \underline{y}) \geq d$. So the statement follows by contrapositive.

2. We use proof by contradiction, so we must assume for contradiction that $\underline{w}$ is a nearest neighbour of $\underline{y}$ in C such that $\underline{w} \neq \underline{v}$. Then $d(\underline{y}, \underline{w}) \leq d(\underline{y}, \underline{v}) \leq t$ so by the triangle inequality

$$0 < d(\underline{v}, \underline{w}) \leq d(\underline{v}, \underline{y}) + d(\underline{y}, \underline{w}) \leq t + t = 2t \leq d - 1.$$

Hence $\underline{v}, \underline{w}$ are distinct codewords at distance less than d. This contradicts d being the minimum distance of C. □

Remark

The Theorem is expressed by saying that a code of minimal distance d **detects up to** $d-1$ **errors** and **corrects up to** $[(d-1)/2]$ **errors** in a codeword.

Channel coding. The importance of rate

To understand why the information dimension k and hence the rate R are defined via logarithm, we recall **channel coding**, the most common use case for codes discussed in the previous chapter. Here is a diagram which shows channel coding with error correction:

$$\begin{tikzcd}
\boxed{\text{sender}} \arrow[r, "m", "\text{message}"'] & {} \arrow[r, "\mathtt{ENCODE}(m)", "\text{codeword}"'] & \boxed{\text{channel}} \arrow[r, "\text{received}", "\text{word}"'] & {} \arrow[r, "\mathtt{DECODE}()"] & {} \arrow[r, "\mathtt{ENCODE}^{-1}()"] & \boxed{\text{receiver}}
\end{tikzcd}$$

Messages are arbitrary words of length k, that is, elements of F^k. The cardinality of F^k is q^k, because a word $(u_1, u_2, \dots, u_k) \in F^k$ can be chosen in q^k ways: q choices for u_1, q independent choices for u_2 and so on. Therefore,

$$M = \#C = \#(F^k) = q^k \qquad \Longrightarrow \qquad k = \log_q M.$$

For each k symbols of information, the sender will transmit a codeword of n symbols. Recall that the rate is $R = \frac{k}{n}$. One has $R \leq 1$ (see the trivial bound below):

Remark: high R, close to 1, is good

Encoding increases transmission costs by a factor of R^{-1}. The higher the rate R, the more economical and efficient the code is.

Although the increase in transmission costs is a proce to pay for error detection or correction, we want this increase to be as small as possible. One can try to construct codes with higher rate, without degrading the error detection or correction performance, by using more sophisticated mathematics. This is one of the main themes in Coding Theory.

There are obstacles to increasing the rate. Mathematically, they are expressed by **bounds.**

Bounds

Proposition 2.2: the trivial bound

If $[n, k, d]_q$-codes exist, then $k \leq n$ and $d \leq n$.

Proof. Let C be an $[n, k, d]_q$-code. Then, by definition, C is a non-empty subset of F^n with $\#F = q$. The cardinality of a set is greater than or equal to the cardinality of its subset. In particular, $M = \#C \leq \#F^n = q^n$. Applying the monotone function $\log_q$ to both sides of the inequality, we obtain $k = \log_q M \leq n$.

Furthermore, the Hamming distance between any two words of length n is an integer between 0 and n. Therefore, $0 < d(C) \leq n$ for any code of length n. □

It is easy to describe the codes which attain $k = n$. All of them are given in the following

Example: F^n, the trivial code of length n

The **trivial code** of length n over the alphabet F is the code $C = F^n$.

Exercise: prove that a code has $k = n$ **if and only if** it is a trivial code. Show that trivial codes have $d = 1$. Show that some codes are not trivial but still have $d = 1$.

We will now give a simple example of codes which attain $d = n$.

Example: $Rep(n, F)$, the repetition code of length n

$Rep(n, F) = \{aaa \dots a \mid a \in F\} \subset F^n$ is the **repetition code** of length n over the alphabet F. All codewords are formed by repeating a symbol n times.

Exercise: prove that $Rep(n, F)$ has $d = n$. Show that some codes are not repetition codes but still have $d = n$.

The Hamming bound

To state the next bound, we recall that $\binom{n}{i}$ is the number of ways to choose i positions out of n. This integer is called the binomial coefficient. It is given by the formula $\binom{n}{i} = \dfrac{n!}{(n-i)!\,i!} = \dfrac{n(n-1)\dots(n-i+1)}{1 \cdot 2 \cdot \dots \cdot i}$.

Theorem 2.3: the Hamming bound

Denote $t = [(d-1)/2]$. If $(n, M, d)_q$-codes exist, $M \leq \dfrac{q^n}{\sum\limits_{i=0}^{t} \binom{n}{i} (q-1)^i}$.

Before proving the Theorem, we introduce

Definition: Hamming sphere

If $\underline{y} \in F^n$ and $r \leq n$, the **Hamming sphere** with centre $\underline{y}$ and radius r is the set

$$S_r(\underline{y}) = \{\underline{v} \in F^n : d(\underline{v}, \underline{y}) \leq r\}.$$

The number of words in the Hamming sphere depends only on the radius r (not on $\underline{y}$):

Lemma 2.4: the cardinality of a Hamming sphere

$$\#S_r(\underline{y}) = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i.$$

Proof. To construct a word $\underline{v}$ at distance i from $\underline{y}$, we need to choose i positions out of n where $\underline{y}$ will differ from $\underline{v}$. Then we need to change the symbol in each of the i chosen positions to one of the other $q-1$ symbols. The total number of choices for $\underline{v}$ which is at distance exactly i from $\underline{y}$ is thus $\binom{n}{i}(q-1)^i$.

The Hamming sphere contains all vectors at distance $0 \le i \le r$ from $\underline{v}$, so we sum over i from 0 up to r. The Lemma is proved. □

Proof of Theorem 2.3. First of all, we prove that spheres of radius t centred at distinct codewords $\underline{c}$ do not overlap. Indeed, by Theorem 2.1(2), each word in $S_t(\underline{c})$ has unique nearest neighbour, which is $\underline{c}$. Hence a word in $S_t(\underline{c})$ cannot lie in another such sphere (a word cannot have two *unique* nearest neighbours!)

Hence the whole set F^n contains M *disjoint* spheres centred at codewords. By Lemma 2.4, each of the M spheres contains $\sum_{i=0}^{t}\binom{n}{i}(q-1)^i$ words. The number of elements in a disjoint union of sets is equal to the sum of cardinalities of the sets, hence the total number of words in the M spheres is $M\sum_{i=0}^{t}\binom{n}{i}(q-1)^i$. Since the union of the M spheres is a subset of F^n, this does not exceed $\#F^n = q^n$. The bound follows. □

Given the length n and the minimum distance d, we may wish to know whether there are codes with the number of codewords *equal* to the Hamming bound. Such a code would be the most economical (highest possible number M of codewords). Such codes have a special name:

Definition: perfect code

A code which attains the Hamming bound is called a **perfect** code.

It turns out that meaningful perfect codes are quite rare. When the number of symbols in the alphabet is a prime power, a complete classification of perfect codes up to parameter equivalence is known; we will see it later in the course.

Remark: what does it mean to attain the bound?

Attains the bound means: the inequality in the bound becomes equality for this code. It is a mistake to say that perfect codes are those that “satisfy” the Hamming bound. Every code *satisfies* the Hamming bound — only perfect codes *attain* it!

The Singleton bound

Another upper bound on the number M of codewords can be conveniently stated for $k = \log_q M$.

Theorem 2.5: the Singleton bound

If $[n,k,d]_q$ codes exist, $k \le n-d+1$.

Proof. Let C be an $[n,k,d]_q$-code. Consider the function $f\colon C \to F^{n-d+1}$ where $f(\underline{v})$ is the word obtained from $\underline{v}$ by deleting the last $d-1$ symbols.

I claim that f is an injective function. Indeed, if $\underline{v}, \underline{w} \in C$, $\underline{v} \neq \underline{w}$, then by definition of the minimum distance, $\underline{v}$ and $\underline{w}$ differ in at least d positions. Since f deletes only $d-1$ symbols, the words $f(\underline{v})$ and $f(\underline{w})$ still differ in at least one position. So $f(\underline{v}) \neq f(\underline{w})$. Injectivity of f is proved.

Now, by the Pigeonhole Principle, injective functions $f\colon C \to F^{n-d+1}$ exist only if $\#C \le \#F^{n-d+1}$. We conclude that $\#C \le q^{n-d+1}$ so that $k = \log_q \#C \le n-d+1$ as claimed. □

Definition: maximum distance separable code, MDS code

A code which attains the Singleton bound is called a **maximum distance separable (MDS)** code.

Remark: the bounds do not work in reverse

It is important to remember that the converses to Theorems 2.3 and 2.5 do not hold. That is, if the numbers n,k,d,q satisfy the Hamming bound and the Singleton bound, it **does not imply that an** $[n,k,d]_q$**-code exists.** For example, n,k,d,q may fail further bounds, not covered in this course.