TC10 / **1. Block Codes**

S. Xambó

Index

The Hamming space

Codes and coders

Decoders

Correcting capacity

The basic scheme for the correction of errors

Minimum distance decoding

The Singleton bound. MDS codes

The Hamming upper bound

Perfect codes

Optimal codes. The function $A_q(n,d)$

The Gilbert bound

The Hamming space

If T is a finite alphabet and n is a positive integer, the elements of the set T^n are the *words* of length *n* formed with the elements of *T*. Often the elements of T^n are also called *vectors* of length *n.*

The *Hamming distance,* $hd(x, y)$, between two words is the number of indices i in the range $1..n$ with $x_i \neq y_i$.

The Hamming distance is a distance in the set T^n, which means that it satisfies the following properties:

1. $hd(x, y) = hd(y, x)$ for all $x, y \in T^n$.
2. $hd(x, y) > 0$ if $x \neq y$ and $hd(x, y) = 0$ if $x = y$.
3. $hd(x, z) \leq hd(x, y) + hd(y, z)$ for all $x, y, z \in T^n$.

Codes

A (block) *T-code of length n* is a subset C of T^n. If M is the cardinal of C, we say that C is of type (n, M).

The *minimum distance* of C, $d = d_C$, is the minimum of the values $hd(x, x')$ for all pairs $x, x' \in C$ such that $x \neq x'$.

In block error-correcting, the stream of symbols output by the compressor is broken into blocks of length k. These blocks are called *information vectors* or *messages*.

If the symbols belong to the alphabet T, then the set of messages is

$$\mathcal{M} = T^k.$$

Coders

A *coder* of type $[n,k]$ is a **one-to-one** map $f\colon T^k \to T^n$. The image of this map is a subset C of T^n which will be called the *code* of f.

If $d = d_C$, we say that C has type $[n,k,d]$, and often we will write

$$C \sim [n,k,d]$$

to denote this.

For example, in the case of the repetition code of order 3 we have $k = 1$, $\mathcal{M} = \{0,1\}, n = 3$ and $C = \{000,111\}$. Therefore $C \sim [3,1,3]$.

Similarly, for the Hamming encoder $f\colon B^4 \to B^7$ we have $k = 4$, $n = 7$ and $d = 3$, and so the corresponding code has type $[7,4,3]$.

Decoders

A *decoder* for the coder f is a map

$$g\colon T^n \to C \cup \{?\} \text{ such that } g(x) = x \text{ for all } x \in C.$$

The vectors in the set

$$D = g^{-1}(C)$$

are said to be *decodable* with respect to g, while the vectors in the set

$$E = g^{-1}(?)$$

are said to be *error vectors* for g.

We say that g *corrects s errors* (s a non-negative integer) if any vector y for which there is an $x \in C$ such that $hd(x, y) \leq s$ is decodable and $g(y) = x$. The highest s for which g corrects s errors is called the *correcting capacity* of g and usually will be denoted by t.

Examples. For the binary repetition code of length 3, any vector of $\{0,1\}^3$ is decodable and $t = 1$. Similarly, for the binary Hamming code of length 7, any vector of $\{0,1\}^7$ is decodable and $t = 1$.

Basic scheme for using a coder/decoder

A coder f of type $[n, k]$ and an f-decoder g with correcting capacity t can be used in a communications system as follows:

1. The information stream output by the compressor is broken into blocks u of length k (messages).
2. For each information block u, calculate the code vector $x = f(u)$. This vector is sent through the channel.
3. If y is the received vector, find $x' = g(y)$.
4. If $x' = ?$, return an error message; otherwise return x', which is a vector in C.

It is clear, directly from the definitions, that if $hd(x,y) \leq t$ (this condition means that the number of errors produced by the channel is not greater than t), then $x\acute{} = x.$ Consequently, the coding/decoding scheme guarantees that any error pattern with at most t errors is corrected.

Note that if the number of errors is $> t$, then it can happen that y is decodable but with $x' \neq x$. In this case we speak of an *undetectable error.* For example, if 000 is the vector sent in the binary repetition code and the received vector is 011, then the decoded symbol is 1, which does not agree with the information symbol 0.

Minimum distance decoding

Notation. Given $w \in T^n$ and a non-negative integer r, we set

$$B(w,r) = \{z \in T^n | hd(w,z) \leq r\}.$$

The set $B(w,r)$ is called the (Hamming) *ball of center* w and *radius* r.

Lemma. $|B(w,r)| = \mathrm{vol}_q(n,r)$, where $\mathrm{vol}_q(n,r) = \sum_{i=0}^{r} \binom{n}{i} (q-1)^i$. For $q = 2$, $\mathrm{vol}_2(n,r) = \sum_{i=0}^{r} \binom{n}{i}$.

Proof. Exercise.

If $C = \{x_1, \dots, x_M\}$, let $D_i = B(x_i, t)$, where $t = \lfloor (d-1)/2 \rfloor$, with d the minimum distance of C. It is clear that $C \cap D_i = \{x_i\}$ and that $D_i \cap D_j = \emptyset$ if $i \neq j$ (by definition of t and the triangular inequality of the Hamming distance). Therefore, if we set $D_C = \coprod_{x_i \in C} D_i$, there is a unique map

$$g: D_C \to C$$

such that $g(y) = x_i$ for all $y \in D_i$. We extend g to T^n by setting $g(y) = ?$ if $y \notin D_C$. By construction, g is a decoder of C and it corrects t errors. This decoder g is called the *minimum distance decoder* of C.

- $g(y)$ is the word $x' \in C$ such that $hd(y, x') \leq t$, if such an x' exists, and otherwise y is non-decodable for g.
- If y is decodable and $g(y) = x'$, then $hd(x, y) > t$ for all $x \in C - \{x'\}$.

Remark. The usefulness of the minimum distance decoder arises from the fact that we can hope, in most ordinary situations, that the transmissions $x \mapsto y$ that lead to a decoder error ($y \notin D$), or to undetectable errors ($y \in D$, but $hd(y,x) > t$) will in general be less likely than the transmissions $x \mapsto y$ for which y is decodable and $g(y) = x$.

To be more precise, the minimum distance decoder maximizes the likelihood of correcting errors if all the transmission symbols have the same probability of being altered by the channel noise and if the $q - 1$ possible errors for a given symbol are equally likely. If these conditions are satisfied, the channel is said to be a (q-ary) *symmetric channel*. Unless otherwise declared, henceforth we will understand that ‘channel’ means ‘symmetric channel’.

From the computational point of view, the minimum distance decoder, as defined above, is inefficient in general, even if d_C is known, for it has to calculate $hd(y,x)$, for $x \in C$, until $hd(y,x) \leq t$, so that the average number of distances that have to be calculated is of the order of $|C| = q^k$. Note also that this requires having generated the q^k elements of C.

But we also have to say that the progress in block coding theory in the last sixty years can be seen, to a considerable extent, as a series of milestones that signal conceptual and algorithmic improvements that make possible to deal with the minimum distance decoder, for large classes of codes, in ever more efficient ways.

The Singleton bound. MDS codes

Proposition (Singleton bound)**. 1**) *For any code of type (n, M, d),*

$$M \leq q^{n-d+1}.$$

2) *For any code of type $[n, k, d]$, $k + d \leq n + 1$.*

Proof. Indeed, if C is any code of type (n, M, d), let us write $C' \subseteq T^{n-d+1}$ to denote the subset obtained by discarding the last $d - 1$ symbols of each vector of C. Then C' has the same cardinal as C, by definition of d, and so $q^k = |C| = |C'| \leq q^{n-d+1}$. Hence $k \leq n - d + 1$, which is equivalent to the stated inequality.

MDS codes. Codes that satisfy the equality in the Singleton inequality are called *maximum distance separable* codes, or *MDS codes* for short. The binary repetition code $\mathrm{Rep}(3)$ is MDS, while the Hamming code $[7,4,3]$ is not.

Remark. For a given length n, the code rate is proportional to k, and the correcting capacity is $\sim d/2$. We seek, therefore, that k and d are as high as possible. The Singleton bound shows that these two requirements cannot be met simultaneously. In practice, a compromise is required.

The Hamming upper bound

Theorem. Let C be a T-code of type (n, M, d). Let $t = \lfloor (d-1)/2 \rfloor$. Then

$$M \leq \frac{q^n}{\mathrm{vol}_q(n,t)}.$$

Proof. The Hamming balls of radius t and center in the elements of C are pair-wise disjoint and hence

$$\sum_{x\in C}|B(x,t)| \leq |T^n| = q^n.$$

On the other hand $|B(x,t)| = \mathrm{vol}_q(n,t)$ and so we have

$$\sum_{x\in C}|B(x,t)| = M\cdot\mathrm{vol}_q(n,t).$$

The conclusion is now obvious.

Remark. The Hamming upper bound is also called *sphere-packing upper bound*, or simply sphere upper bound.

Perfect codes

In general D_C is a proper subset of T^n, which means that there are elements $y\in T^n$ for which there is no $x\in C$ with $hd(y,x)\leq t$.

If $D_C = T^n$, then C is said to be *perfect*. In this case, for every $y\in T^n$ there is a (necessarily unique) $x\in C$ such that $hd(y,x)\leq t$.

Taking into account the reasoning involved in proving the sphere-bound, we see that the necessary and sufficient condition for a code C to be perfect is that

$$\sum_{i=0}^{t} \binom{n}{i} (q-1)^i = q^n/M \ \ (= q^{n-k}),$$

where $M = |C| = q^k$ (this will be called the *sphere* or *perfect* condition).

Examples. The total code T^n and the binary repetition code of **odd** length are examples of perfect codes, with parameters

$(n, q^n, 1)$ and $(2m+1, 2, 2m+1)$,

respectively. Such codes are said to be *trivial perfect codes*. We have also seen that the Hamming code $[7,4,3]$ is perfect (actually this is a direct check).

Optimal codes. The function $A_q(n,d)$. A code $C \sim (n, M, d)$ is said to be *optimal* if $M \geq M'$ for any code (n, M', d). The cardinal M of an optimal code depends only of n and d, and it is denoted $A_q(n,d)$.

Examples. $A_q(n,d) \leq q^{n-d+1}$, by the Singleton bound,[N1] and $A_q(n,d) \leq q^n/\mathrm{vol}_q(n,t)$, by the Hamming bound.[N2]

Theorem (Gilbert lower bound). $A_q(n,d) \geq \frac{q^n}{\mathrm{vol}_q(n,d-1)}$.[N3]

Proof. If $C \sim (n,M,d)$ is optimal, any element of T^n lies at a distance $\leq d-1$ of an element of C, for otherwise there would be a word $y \in T^n$ lying at a distance $\geq d$ from all elements of C and $C \cup \{y\}$ would be a code of length n, minimum distance d and with a greater cardinal than $m = |C|$, contradicting the optimality of C. This means that the union of the balls of radius $d-1$ with center the elements of C is the whole T^n. It follows that $M \cdot \mathrm{vol}_q(n,d-1) \geq q^n$, and this ends the proof as $M = A_q(n,d)$.

Some values of $A_2(n,d)$

n	$d=3$	$d=5$	$d=7$
5	4	2	--
6	8	2	--
7	16	2	2
8	**20**	4	2
9	40	6	2
10	72-79	12	2
11	144-158	24	4
12	256	32	4
13	512	64	8
14	1024	128	16
15	2048	256	32
16	2720-3276	256-340	36-37

This table gives the values of $A_2(n,d)$, or the best known bounds, for $5 \leq n \leq 16$ and $3 \leq d \leq 8$. The values for d even (4 and 6 in this table) are determined by the relation $A_2(n+1,d) = A_2(n,d-1)$.

Notes

N1. # Singleton upper-bound

```
ub_singleton(n,d,q):= q^(n-d+1);
ub_singleton(n,d):= 2^(n-d+1);
ub_singleton(8,3) → 64
```

N2. # Sphere-packing upper-bound

```
ub_sphere(n,d,q):= floor(q^n/volume(n,(d-1)//2,q))
ub_sphere(n,d):= floor(2^n/volume(n,(d-1)//2))
ub_sphere(8,20) → 28
```

N3. # Gilbert lower bound

```
lb_gilbert(n,d,q) := ceil(q^n/volume(n,d-1,q))
lb_gilbert(n,d) := ceil(2^n/volume(n,d-1))
lb_gilbert(8,20) → 7
```