Remark 2.8 If one knows the distance $d(C)$ of the code, then $t = d(C) - 1$ is the best choice for the number of errors that can be detected.

$e = \left\lfloor \frac{d(C)-1}{2} \right\rfloor$ is the largest number of errors, C can correct.

$$\boxed{\begin{aligned} & d(C) \geq 2e+1 \\ \Updownarrow \quad & \frac{d(C)-1}{2} \geq e \\ \Updownarrow \quad & \left\lfloor \frac{d(C)-1}{2} \right\rfloor \geq e \end{aligned}}$$

Def 2.9 An (n, M, d) code is a code $C \subseteq K^n$ of code length n, size $|C| = M$ and the minimum distance $d(C) = d$.

The ratio $\frac{d}{n}$ is called the relative distance.

Example 2.10 The repetition code

$$C = \{ (\underbrace{a, \dots, a}_{n}) \in K^n : a \in K \}$$

Block length: n

The size: $|C| = |K| =: q$ the alphabet size.

The distance: $d(C) = n$.

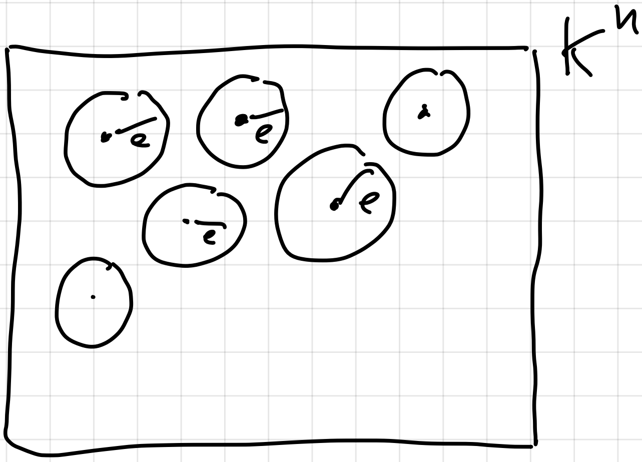
This C can detect $n-1$ errors and correct $\left\lfloor \frac{n-1}{2} \right\rfloor$ errors.

2.2. Bounds on codes

Theorem 2.11 (The sphere packing bound, Hamming bound)

Let $C \subseteq K^n$ be a code over a q-ary alphabet K ($n \in \mathbb{N}$, $q \in \mathbb{N}$, $q \geq 2$) and let $e \in \mathbb{N}_0$ be satisfying $2e+1 \leq d(C)$. Then one has

$$|C| \leq \frac{q^n}{V_q^n(e)}$$



Furthermore, the equality

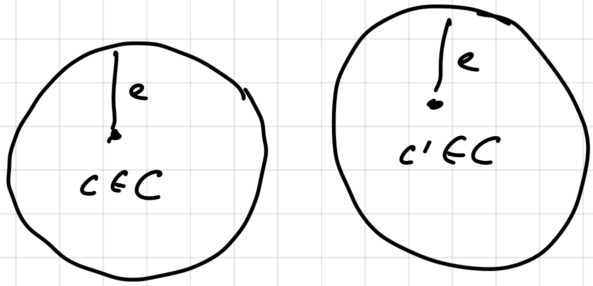
$$|C| = \frac{q^n}{V_q^n(e)}$$ is attained

if and only if each word $x \in K^n$ is contained in exactly one ball $B(c, e)$ with $c \in C$ (the ball w.r.t. the Hamming distance).

Proof: By assumption, $d(c, c') > 2e$ for all $c, c' \in C$, $c \neq c'$

$\Rightarrow B(c,e) \cap B(c',e) = \emptyset$

(see $\boxed{LQ3}$).



So, the balls $B(c,e)$ with $c \in C$ are pairwise disjoint. $\Longrightarrow$

$$\underbrace{\left| \bigcup_{c \in C} B(c,e) \right|}_{} = \sum_{c \in C} |B(c,e)| = |C| \cdot V_q^n(e)$$

On the other hand, this size is at most the size of K^n, which q^n. $\Longrightarrow$

$q^n \geq |C| \cdot V_q^n(e) \Rightarrow |C| \leq \frac{q^n}{V_q^n(e)}$.

In particular, having equality $|C| = \frac{q^n}{V_q^n(e)}$

is equivalent to K^n being the disjoint union of the balls $B(c,e)$ with centers $c \in C$. $\square$

Def. 2.12 Codes satisfying the equality $q^n = |C| \cdot V_q^n(e)$ are called <u>perfect</u>.

One can show that some number of perfect codes exist, but they're "quite rare".

Even when we cannot or don't know how to get to the equality in $|C| \leq \frac{q^n}{V_q^n(e)}$, we at least can try to get close to the equality case.

Thm 2.13 (Gilbert-Varshamov bound = GV bound)

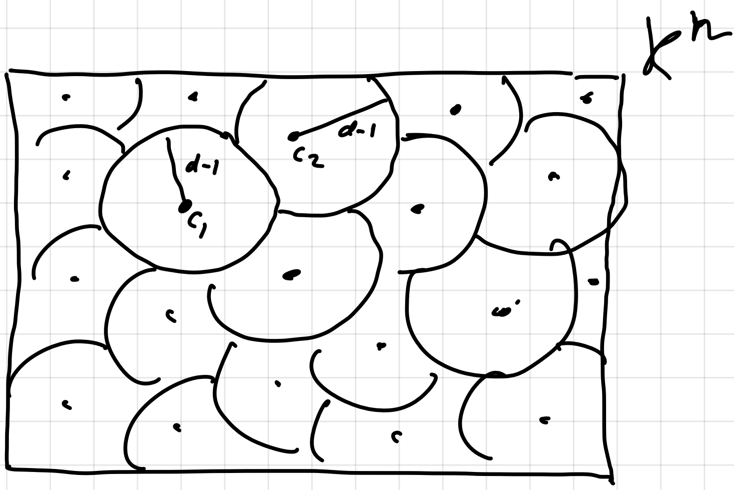
Let $n, q, d \in \mathbb{N}$ with $1 \leq d \leq n$ and $q \geq 2$ be given, and let K be an alphabet of size q.

Then there exists a code $C \subseteq K^n$ of minimum distance $d(C) \geq d$ and size

$$|C| \geq \frac{q^n}{V_q^n(d-1)}.$$

<u>Proof:</u> Let's use a greedy strategy to construct such C. Start with the empty set $C = \emptyset$.

Iteratively, add another element $c \in K^n$ to C such that



$d(c,c') \geq d$ for all $c' \in C$

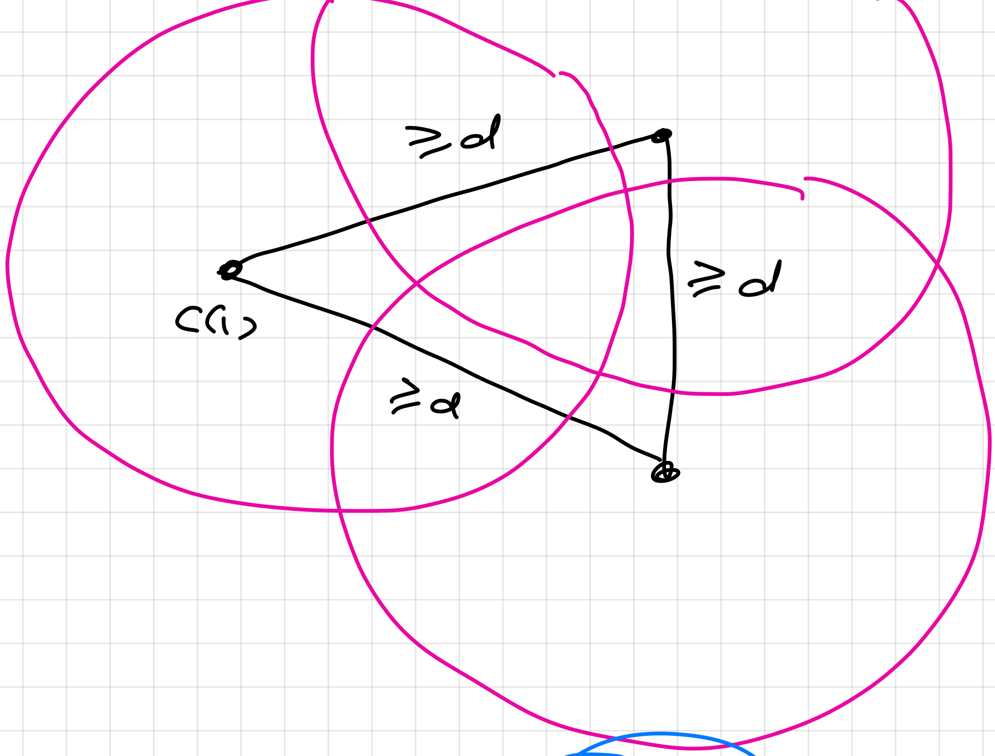
as long as such an element exists). In terms of the balls, the condition on the choice of $c \in K^n$ is
$$c \in K^n \setminus \bigcup_{c' \in C} B(c', d-1).$$

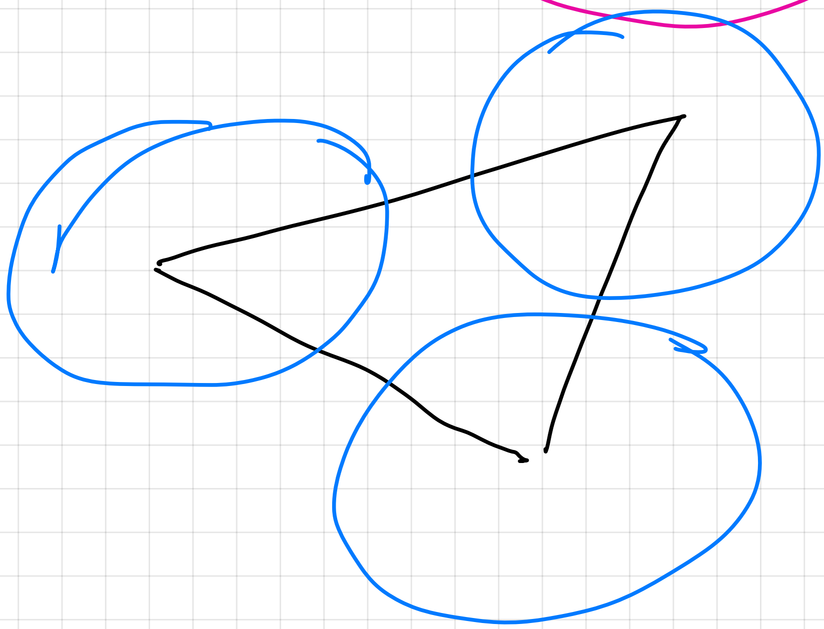
This process terminates after finitely many iterations, because we pick elements from the finite set K^n.

Upon termination, we have
$$\bigcup_{c' \in C} B(c', d-1) = K^n. \quad \text{So,}$$

$$q^n = |K^n| = \left| \bigcup_{c' \in C} B(c', d-1) \right| \leq \sum_{c' \in C} |B(c', d-1)|$$
$$= |C| \cdot V_q^n(d-1)$$

$$\Rightarrow \quad |C| \geq \frac{q^n}{V_q^n(d-1)}. \qquad \square$$





$$\boxed{\begin{aligned} d &= d(C) \\ e &= \left\lfloor \frac{d-1}{2} \right\rfloor \end{aligned}}$$

For every C

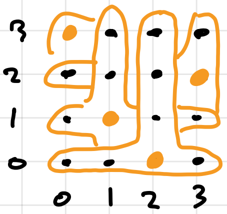
$$|C| \leq \frac{q^n}{V_q^n\left(\lfloor \frac{d-1}{2} \rfloor\right)}$$

there exists C with $d(C) \geq d$

$$|C| \geq \frac{q^n}{V_q^n(d-1)}.$$

Example 2.14 $K=\{0,1,2,3\}$, $n=2$

$d=2$



$C=\{20,\ 11,\ 32,\ 03\}$

Code over the alphabet of size with block length $n=2$, the minimum distance 2 and size 4.

Another code like this:

$\{00, 11, 22, 33\}$

LQ5 $n=6$, $d=3$, $q=2$:

Try to find a large code with this parameters: $C\subseteq\{0,1\}^6$,

$d(C)\geq 3$

Try using greedy strategy as above?

Can you do better?

Thm 2.16 (Singleton bound) Let $C\subseteq K^n$ be a q-ary code with the minimum distance $d=d(C)$, where $q,d,n\in\mathbb{N}$ and $q\geq 2$. Then

$$|C|\leq q^{n-d+1}$$

or, equivalently

$$d\leq n-\log_q|C|+1.$$

Proof:

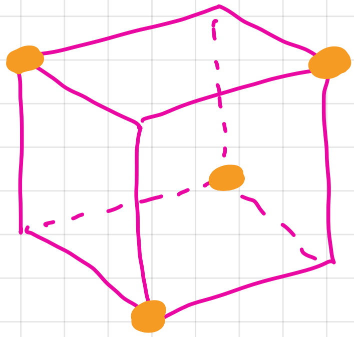
$c=\cancel{c_1c_2c_3\ \cdots\ c_{d-1}}\ c_d\ \cdots\cdots\ c_n\in C$

$c'=\cancel{c_1'c_2'c_3'\ \cdots\ c_{d-1}'}\ c_d'\ \cdots\cdots\ c_n'\in C$

if the distance is d and $c\neq c'$, after the erasure of the first $d-1$ symbols there is still a difference.

Consider the map $T: C \to K^{n-d+1}$ given by $T(c_1 \dots c_n) := c_d \dots c_n$. Since $d(C) = d$ any two distinct codewords $c, c' \in C$ would satisfy $T(c) \neq T(c')$, because c and c' differ in at least d positions and one of this positions $i = 1, \dots, n$ fall into the range $i \in \{d, \dots, n\}$. So T is an injective map.

Since T is injective



?10

$$|C| = |T(C)| \le |K^{n-d+1}| = q^{n-d+1}$$

$$\Rightarrow |C| \le q^{n-d+1}$$

$$\Leftrightarrow \log_q |C| \le n-d+1$$

$$\Leftrightarrow d \le n - \log_q |C| + 1. \qquad \square$$

Rem 2.17.

Let's discuss what it means.

Assume one encodes K^k (that is, one send k symbols) via some code C.

That means $|C| = q^k$. The bound is

$$d \le \underbrace{n-k}_{} + 1.$$

the overhead; we have k symbols of information and $n-k$ symbols for redundancy.

Assume we decide to have $n-k=5$.

Then, Singleton bound says $d \le 5+1=6$.

So, the maximum distance we can hope for is 6.

If we attain it, we can correct at most $\left\lfloor \frac{6-1}{2} \right\rfloor = 2$ errors.

Definition 2.18 A code attaining the Singleton bound with equality is called *maximum distance separable* or an *MDS code*.

Remark 2.19 With MDS codes one can restore unreadable parts: $0\,1\,0\,1\,1\,0 \rightsquigarrow 0\,?\,0\,1\,1\,?$

$\boxed{d=3}$ $\quad c_1' c_2' c_3' c_4' c_5' c_6'$

3 Decoding for stochastic channels

3.1 Probability over finite sample spaces in a nutshell

Def 3.1 A finite probability space is a pair (Ω, P) where Ω is a nonempty finite set and $P : \Omega \to \mathbb{R}$ is a function with $P(\omega) \ge 0$ for all $\omega \in \Omega$ and

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

$\omega \in \Omega$ is called the elementary event, $P(\omega)$ is called its probability, $A \subseteq \Omega$ is called an event and $P(A) := \sum_{\omega \in A} P(\omega)$ is called the probability of A.

Example 3.2. Outcomes of 2 tosses of a fair coin.

$$\Omega = \{00, 10, 01, 11\} = \{0,1\}^2$$

$P(\omega) = \frac{1}{4}$ for all $\omega \in \Omega$

P(outcomes were different)

$= P(\{01, 10\}) = P(01) + P(10) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$.

Example 3.3. 4 Tosses of a biased coin with the probability of tails equal to ε.

$0 =$ ☺ Probability $1-\varepsilon$

$1 =$ (1) Probability ε.

$$P(1001) = \varepsilon^2 \cdot (1-\varepsilon)^2$$
$$P(1011) = \varepsilon^3 \cdot (1-\varepsilon)$$
$$P(1111) = \varepsilon^4$$
$$P(\underbrace{x_1 x_2 x_3 x_4}_{x}) = \varepsilon^{wt(x)} \cdot (1-\varepsilon)^{4-wt(x)}$$

$wt(x) =$ # of non-zero entries in x

$$P(0000) = (1-\varepsilon)^4$$

$P(\text{exactly one tail}) = \binom{4}{1} \varepsilon \cdot (1-\varepsilon)^3$

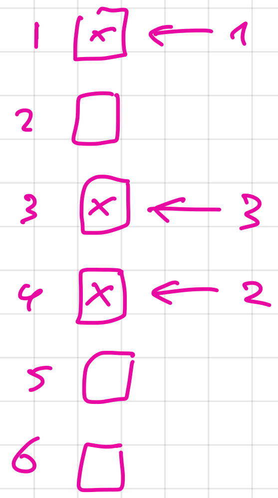
$P(\text{exactly two tails}) = \binom{4}{2} \varepsilon^2 (1-\varepsilon)^2$

$P(\text{exactly three tails}) = \binom{4}{3} \varepsilon^3 (1-\varepsilon)^1$

$P(\text{all four being tails}) = \binom{4}{4} \varepsilon^4 \cdot (1-\varepsilon)^0$

Catching up on counting: Binomial coefficients.

There are n courses offered in the Master's program: let's number them from $1, \dots, n$. A student has decided to take i of these courses. $i = 0, \dots, n$. What is the number of possibilities one can take i out of n courses.



$$\frac{6 \cdot 5 \cdot 4}{3 \cdot 2 \cdot 1} = 20$$

$$\frac{n \cdot (n-1) \cdot \dots \cdot (n-i+1)}{i \cdot (i-1) \cdot \dots \cdot 1} = \binom{n}{i}$$

$$= \frac{n!}{i!\,(n-i)!}$$ ← the binomial coefficient.

Get some experience with counting (combinatorics).

LQ6