recall: • $V(q,n,r) := \sum_{i=0}^{r} \binom{n}{i} (q-1)^i$ Volume of Hamming ball $B_r \subseteq \mathbb{Z}_q^n$

• Gilbert-Varshamov bound: $\forall (q,n,d) \in \mathbb{N}^3$ $\exists$ code $C \subseteq \mathbb{Z}_q^n$ s.t.

$$d(C) = d \quad \wedge \quad |C| \geq q^n / V(q,n,d-1)$$

• $f(x) = o(g(x)) \iff \lim_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| = 0$, e.g. $f(x) = o(1)$ means $f(x) \xrightarrow[x\to\infty]{} 0$.

• $f(x) = \Omega(g(x)) \iff \liminf_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| > 0$ i.e. g is asympt. lower bound.

Lemma: For $p \in [0, \frac{1}{2}]$ and increasing $n \in \mathbb{N}$, we have

$$2^{(h(p) - o(1))n} \leq V(2,n,pn) \leq 2^{h(p)n}$$

(where $h(p) := -p\log p + (1-p)\log(1-p)$ is the binary entropy).

Corollary: (i) For $p \in [0, \frac{1}{2}]$ there is a sequence of binary codes $(C_n)_{n\in\mathbb{N}}$ with relative distance $\frac{d(C_n)}{n} \geq p$ such that

$$R(C_n) \geq 1 - h(p).$$

(ii) Conversely, for $p \in [0,1]$ every sequence of binary codes with $\frac{d(C_n)}{n} \xrightarrow{n\to\infty} p$ satisfies

$$R(C_n) \leq 1 - h(\tfrac{p}{2}) + o(1)$$

proof: (i) by definition $R(C_n) := \frac{\log|C_n|}{n}$.

$$R(C_n) \geq 1 - \tfrac{1}{n}\log V(2,n,d-1) \quad \text{by Gilbert-Varshamov}$$

$$\geq 1 - h(p) \quad \text{using the Lemma for } pn = d-1$$

(ii) $R(C_n) \leq 1 - \frac{1}{n}\log V(2,n,\lfloor\frac{d-1}{2}\rfloor)$ Hamming bound

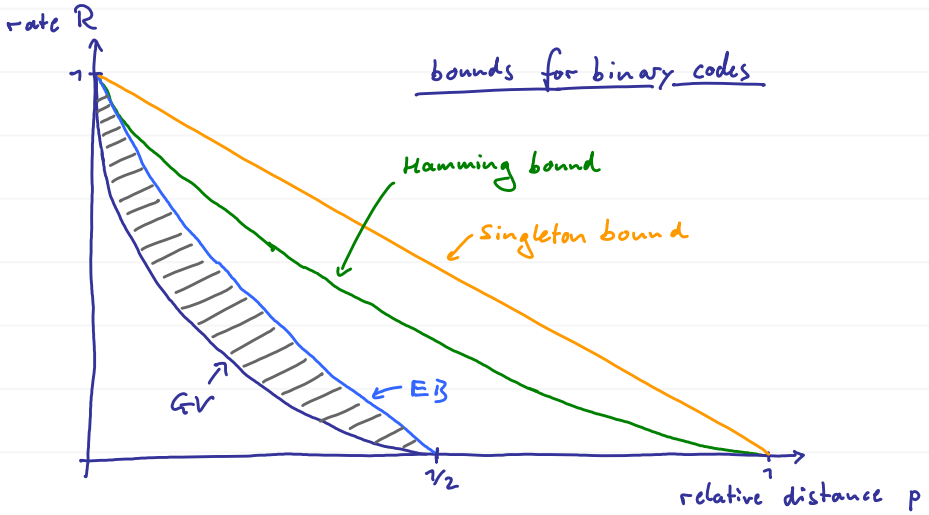
$$\leq 1 - h(\tfrac{p}{2}) + o(1) \quad \text{Lemma}$$

□

Consequence: Asymptotically good binary codes exist !

remark: an improved upper bound is the "Elias-Bassalygo" bound:

$$R(C_n) \leq 1 - h\left(\frac{1-\sqrt{1-2p}}{2}\right) + o(1) \quad \text{for } n \to \infty$$



Prop.: (Singleton bound)

For every q-ary code $C \subseteq \mathbb{Z}_q^n$ with block length $n \in \mathbb{N}$ and distance d, we have

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

proof:

- take all $|C|$ codewords and erase the first $(d-1)$ symbols
- we are left with $|C|$ strings which are distinct (since the distance was d) and of length $n-(d-1)$

$$\Rightarrow \quad |C| \leq q^{n-d+1} \quad = \text{max \# of } q\text{-ary strings of length } n-(d-1)$$

□

Corollary: Any linear $[n,k,d]$ code satisfies $k \leq n-d+1$

proof: $|C| = q^k$. □

Def.: Linear $[n,k]$ codes with distance $d = n-k+1$ are called "maximum distance separable" (MDS) codes.

remarks: • MDS codes require large alphabets:

a sequence of asymptotically good codes can be MDS only if $\boxed{q = \Omega\left(\frac{n}{\log n}\right)}$ with $n \to \infty$

• a sequence of MDS codes whose rel. distance is bounded away from 0 & 1 is asymptotically good, since

$$R = \frac{k}{n} \overset{\text{MDS}}{=} 1 - \frac{d}{n} + \frac{1}{n}.$$

note that $R + \frac{d}{n} \to 1$ for MDS codes !

V.4. Reed-Solomon codes

Def.: For integers $1 \le k < n \le q$ and $\alpha \in GF(q)^n$ with distinct components

$$C := \left\{ p(\alpha) \in GF(q)^n \;\middle|\; p \text{ is polynomial over } GF(q) \text{ of degree} < k \right\}$$

is called "Reed-Solomon code" and we will write $[n,k]$-RS code.

Encoding: • we identify a messag $m \in GF(q)^k$ with a polynomial

$$p_m(x) := \sum_{i=0}^{k-1} m_i x^i$$

the codeword is then $p_m(\alpha) = (p_m(\alpha_1), \dots, p_m(\alpha_n))$

• $p_m(\alpha) = Gm$ where $G \in GF(q)^{n \times k}$ is a "Vandermonde matrix" with $G_{xy} := \alpha_x^{y-1}$

$\Rightarrow$ RS codes are linear

remarks: • RS codes have large alphabet (since $q \ge n$)

• typical choices for α:

(i) $\{\alpha\} = \{GF(q)\}$ i.e. $q = n$

(ii) $\{\alpha\} = \{GF(q)\} \setminus \{0\}$ i.e. $q = n+1$

$\alpha = (\beta^0, \beta^1, \dots, \beta^{n-1})$, β "primitive element" of $GF(q)$

Def.: $\mathbb{K}_d[X]$:= space of polynomials over the field $\mathbb{K}$ with degree $\leq d$.

Lemma: If $\alpha \in \mathbb{K}^{d+1}$ has distinct components, then $\hat{\alpha}: \mathbb{K}_d[X] \to \mathbb{K}^{d+1}$,

$\hat{\alpha}: p \mapsto (p(\alpha_1), \dots, p(\alpha_{d+1}))$ is bijective.

proof: "Lagrange interpolation" define $L_i \in \mathbb{K}_d[X]$,

$$L_i(x) := \frac{\prod_{k \neq i} (x - \alpha_k)}{\prod_{l \neq i} (\alpha_i - \alpha_l)}, \quad i, k, l \in \{1, \dots, d+1\}$$

Then $L_i(\alpha_j) = \delta_{ij}$. For any $\beta \in \mathbb{K}^{d+1}$ define $p(x) := \sum_{i=1}^{d+1} \beta_i L_i(x)$.

Then $p \in \mathbb{K}_d$ and $p(\alpha_i) = \beta_i$. Hence $\hat{\alpha}$ is surjective.

Conversely, if $p, \tilde{p} \in \mathbb{K}_d[X]$, then $(p - \tilde{p}) \in \mathbb{K}_d[X]$ has $(d+1)$ roots $\{\alpha_i\}_{i=1}^{d+1}$

$\Rightarrow p - \tilde{p} = 0$, so $\hat{\alpha}$ is also injective. □

Thm.: For an $[n, k]$-RS code over $GF(q)$ we have

(i) $|C| = q^k$

(ii) $d(C) = n - k + 1$

proof: (i) follows from injectivity of G

(ii) Linearity $\Rightarrow d(C) = \min_{c \in C \setminus \{0\}} d(c)$

for $m \in GF(q)^k \setminus \{0\}$ $p_m(x)$ has at most $k-1$ roots as $p_m \in \mathbb{K}_{k-1}[X]$.

$\Rightarrow$ codeword $c = p_m(\alpha)$ has at most $k-1$ zeros

$\Rightarrow d(C) \geq n - k + 1$

Singleton bound: $d(C) \leq n - k + 1$. □

Corollary: RS-codes are MDS codes (i.e. they achieve the Singleton bound)

V.5. Error bursts & interleaving

sources for errors are often not memoryless / uncorrelated, e.g.:

- scratches on CD
- disturbance / loss of signal for time intervals

→ "bursts" of errors

simple ways to deal with this:

(i) use codes with large alphabet (e.g. RS) & represent symbols using smaller alphabets. E.g. code over $GF(2^m)$ with distance d corrects bursts of length $\left(\left\lfloor \frac{d-1}{2} \right\rfloor - 1\right)m + 1$ when information is stored using contiguous bits.

(ii) interleaving = rearranging symbols in concatenated codewords.

Consider $[n,k]$-code & let $c^{(i)} = (c_1^{(i)}, \dots, c_n^{(i)}) \in C$, $i \in \{1, \dots, t\}$.

Define new $[nt, kt]$-code $\tilde{C}$ from all codewords of the form

$$\tilde{c} = \left(c_1^{(1)} c_1^{(2)} \cdots c_1^{(t)} \; c_2^{(1)} \cdots c_2^{(t)} \cdots c_n^{(t)} \right)$$

C corrects bursts of length $b \Rightarrow \tilde{C}$ corrects bursts of length $\tilde{b} = t \cdot b$

Example: $[256, 229]$ – RS code: rate $\sim 90\%$, corrects 13 byte errors

→ corrects bursts of $12 \cdot 8 + 1 = 97$ bit errors

→ $t = 37$ interleaving corrects burst up to 3 kbits

(essentially this happens on a CD. 3 kbits $\hat{=}$ 2.5 mm on surface)