Optimal possibly nonlinear 3-PIR codes of small size

Presentation: Urmas Luhaäär
Work: Urmas Luhaäär Henk Hollmann

May 2022

Definitions and notations

- A binary k-to-n encoder ϵ is a one-to-one mapping $\epsilon\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$.
- We call the image of ϵ the associated code of ϵ the elements of the code we call code words.
- A decoder δ of a code is the inverse of the encoder.

Definitions and notations

- Let $I = \{i_1, \ldots, i_s\} \subseteq [n]$ with $i_1 \leq \cdots \leq i_s$.
- Given some code word $c \in \mathbb{F}_2^n$ let the restriction c_I of c to I be $c_I = (c_{i_1}, \ldots, c_{i_s})$.
- We say that I is the recovery set of the j-th data symbol if for all $c \in \mathbb{F}_2^n$ and $a \in \mathbb{F}_2^k$, whenever $c = \epsilon(a)$, the restriction c_I uniquely determines a_j.
- Such a set is called minimal if no proper subset has this property.

Definitions and notations

- A query of ϵ is a sequence $i_1, \ldots, i_t$ of elements of $[k]$.
- Given a code word c a query $i_1, \ldots, i_t$ is a request for the data symbols $a_{i_1}, \ldots a_{i_t}$.
- We say that the sets $I_1, \ldots, I_t$ serve the query $i_1, \ldots, i_t$, if for every $j \in [t]$, I_j is a recovery set for the i_j-th data symbol.
- We say that $I_1, \ldots, I_t$ serve the query with multiplicity μ if every position $i \in [n]$ occurs in at most μ sets.

Definitions and notations

- We say that an encoder $\epsilon\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ is a (t, ∞, μ)-PIR code if any query of the form $i, \ldots, i$ (t times) can be served with multiplicity μ.
- (The ∞ denotes the maximum width of the sets I_{i_j}, we will allow them to be of any size up to n)

Definitions and notations

- Let t and k be positive integers. We define $P(k, t)$ as the smallest length n $t-$PIR code(possibly nonlinear).

A generalization of a known bound

Theorem

Let C be a code over an alphabet Σ with length n and minimal distance d, and suppose that C has an encoder that is a (t, ∞, μ)PIR-code. Then $\lceil t/\mu \rceil \leq d$.

Hamming codes

- For an integer $r \geq 2$, the binary r-th order Hamming code is a linear code of length $n = 2^r - 1$ and dimension $k = 2^r - 1 - r$. with the parity check matrix H_k, whose columns are the $2^r - 1$ nonzero binary words of length r.
- Hamming codes have a minimum Hamming distance of three.

Hamming codes as PIR codes

- We label the positions of the code words with sets $S \subseteq \mathbb{F}_2^r$ and identify these sets with their characteristic vectors χ_S of length $2^r - 1$.
- If we take a set $S = \{u, v, u + v\}$, where $u \neq v$, then this set will correspond to a weigth 3 code word in the Hamming code.
- These weight 3 vectors correspond to lines in the projective geometry $\mathrm{PG}(r - 1, 2)$.
- In $\mathrm{PG}(r - 1, 2)$, every point is on $(2^r - 2)/2 = 2^{r-1} - 1$ lines. Since the Hamming code is linear, then we can add together all the lines to see that for $r \geq 2$ the all-one word is in the Hamming code.
- The all zero code word is also in the Hamming code because it is linear.

Hamming codes as PIR codes

Theorem

A Hamming code of order $r \geq 3$ cannot have more than two disjoint recovery sets.

Proof

- We first choose some data symbol a_1.
- Suppose that there are at least three minimal recovery sets I_1, I_2, I_3 for a_1.
- First we prove that if a line L in $\mathrm{PG}(r-1,2)$ intersects two of the sets I_1, I_2, I_3 then it must also intersect the third.
- Then we show that none of the sets contain a line.
- If we take two points P and Q from one of the sets I_i, then they form a line with three points. From the second claim it follows that the third point is not in I_i and from the first it follows that the third point is not in the other sets either.
- Suppose that we have some code word c and ℓ is the weight three code word associated with the line L. Then $c+\ell$ is also a code word.
- The restrictions on the sets other than I_i remain unaffected by adding ℓ to c. Therefore c and $c+\ell$ have the same data symbol a_1. Since a_1 was chosen arbitrarily, it holds for all data symbols.

Proof

- Since the set I_i and the points P, Q were chosen arbitrarily and adding a line constructed in the manner described doesn't change the data symbol, code words with even weighted restrictions on a set I_i will have equal data symbols a_1 and code words with odd restrictions will have equal data symbols a_1.
- To prove the theorem, it suffices to prove that all of the sets I_1, I_2, I_3 are of even length because then the all one code word and all zero code word would become indistinguishable.
- Consider the set H consisting of the zero vector and all the vectors outside $I_1 \cup I_2 \cup I_3$. It turns out that this set is a subspace of $\mathbb{F}_2^r$ and I_1, I_2, I_3 are it's cosets. Therefore the sets are of size $2^r/4 = 2^{r-2}$.

Optimal 3-PIR code lengths of for small k

k	1	2	3	4	5	6	7	8
n	3	5	6	8	9	10	12	13

Table 1 Smallest length n of k-dimensional linear 3-PIR codes

n	3	4	5	6	7	8	9	10	11	12
M	2	2	4	8	16	20	40	72	144	256

Table 2 Maximum size M of a binary code of length n and minimum distance 3

Open problem Does there exist a (nonlinear) binary 3-PIR code of length 11 and size 2^k?

Thank you!