More on Coset Assignment

Ziquan Yang

July 8, 2014

[1]

1 Motivation

The type of questions we have cared so far is: given a repair group and minimum distance, what is the optimal dimension of the code? Now we essentially have 3 stategies to bound dimension in our toolbox. I call them "Punctured Code Argument", "$C^\perp$ Rank Argument", and "Subcode Argument". It will be nice if we could infer the existence of a code that actually meets the bound, since then there is no need to improve the bound.

2 Subcode Argument: A First Example

Theorem 2.1. *Let C be a code of length $n = 2^m - 1$, where $4 \mid m$, with locality $r = 4$ and minimum distance $d \geq 14$. Assuming disjointness of repair group, the rank k of C is bounded by:*

$$k \leq \frac{4}{5}(2^m - 1) - 3m + 3$$

for $m \geq 8$.

Proof. Let $n' = n/5$. Up the permutation of columns, $C^\perp$ contains the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix} \otimes I_{n'}$$

Now consider the subcode C' of C with $C'^\perp$ containing the matrix:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix} \otimes I_{n'}$$

C' is formed by adding at most 2 parity checks to each 5-block, some of which might have been in $C^\perp$ already, to $C^\perp$, so the rank k' of C' has bound:

$$k' \geq k - \frac{2}{5}(2^m - 1)$$

Note that generator matrix of C' is contained in the following matrix:

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{bmatrix} \otimes I_{n'}$$

[1] Original title of this note was “Bounds and Achievability”.

and C' inherited minimum distance $d' \geq 14$. Now apply sphere packing to C' and we obtain:

$$2^{k'} \leq \frac{4^{n'}}{1 + 9n'(n'-1)/2 + 9n'(n'-1)(n'-2)/2}$$

which yields

$$\begin{aligned} k' &\leq \log_2(\frac{4^{n'}}{1 + 3n' + 9n'(n'-1)/2 + 9n'(n'-1)(n'-2)/2}) \\ &= \frac{2}{5}(2^m - 1) - \lceil \log_2(1 + 9n'(n'-1)/2 + 9n'(n'-1)(n'-2)/2) \rceil \\ &= \frac{2}{5}(2^m - 1) - 3m + 3 \end{aligned}$$

for $m \geq 8$. The conclusion follows. □

Remark 2.2. Sphere packing at the last step could be replaced by the $C^\perp$ rank argument. We want to add additional parity checks to assign codewords of weight 2 to different cosets of kernel other than kernel itself. We have bound on the number of parity checks we need to add:

$$2^b \geq 1 + 3n' + 9\binom{n'}{2} + 27\binom{n'}{3}$$

which gives exactly the same bound. The coset assignment sheds more light on whether the b additional parity checks exist. However it is still very hard, as shown later. A similar case for $d \geq 6$ is shown with induction, while even $d \geq 10$ is much harder. Moreover, even when we could construct additional parity checks to rule off codewords of weight smaller than 14 in the subcode, it is hard to reconstruct a code, from which it inherits, since it is easy for a subcode to inherit minimum distance, but the process cannot be reversed.

3 Infer More Information from Optimality

The cyclic code constructed for $r = 2$ and $d \geq 6$ has minimum distance exactly 6. The defining set of this cyclic code is all α^{3i} and the cyclotomic cosets of α, α^{-1}. We divided $C^\perp$ into two parts: P and M, where P are the codewords of weight 3 providing the locality and M imposes the minimum distance. In particular, M maps vectors of weight smaller than 2 in $P^\perp$ to distince images. Thus we bounded the rank b of M as follows:

$$2^b \geq 1 + 3 \times \frac{2^m - 1}{3}$$

Thus $b \geq m$. Let S be the set of vectors of weight 2 in $P^\perp$. When $b = m$, this means that S has been mapped bijectively to $\mathbb{F}_2^m - \{0\}$. Thus we label each coset with a vector in S. Now consider a disjoint pair $v_1, v_2 \in S$. Suppose $v_1 + v_2$ lies in the coset labelled by v_3, where $v_3 \in S$. Note that v_3 must be disjoint from $v_1 + v_2$, otherwise $v_1 + v_2 + v_3$ is a vector in the kernel of weight 4, a contradiction. Thus we have $\mathrm{wt}(v_1 + v_2 + v_3) = 6$.

4 Coset Assignment and Existence

4.1 A Simple Example

Using the same notation as in Section 3, consider the following question: Does there exists a injective map $L : S \mapsto \mathbb{F}_2^m - \{0\}$ that respects all the linear dependence relations within

S? Clearly, such L exists if and only if the bound we have shown is achievable. With an explicit construction of an optimal code (i.e. the cyclic code we constructed), we know that this question has a positive answer. However, I think it is more of a theoretical interest to prove the existence of an optimal code without an explicit construction. S is a subset of a vector space over $\mathbb{F}_2$ and let $\{v_1, v_2, \cdots, v_n\}$ be its elements, where $n = 2^m - 1$. Consider a linear dependence relation:

$$\sum_{j=1}^{k} v_{i_j} = 0$$

we represent it as a vector $w \in \mathbb{F}_2^n$ with $\text{supp}(w) = \{i_1, i_2, \cdots, i_k\}$. Clearly, the linear dependence relations form a subspace in $\mathbb{F}_2^n$. With inspection we see that a basis for this subspace gives a parition of S into 3-subsets. The basis vectors are disjoint and are of the form: $v_{i_1} + v_{i_2} + v_{i_3} = 0$. This means that we only need to show the following:

Proposition 4.1. *There exists a partition of $\mathbb{F}_2^m - \{0\}$ into 3-subsets and the vectors in each subsets add up to 0. Now we show this by induction.*

Take $m = 2$, then we have $\mathbb{F}_2^m = \{00, 01, 10, 11\}$. The trivial partition will do. Now suppose $\{v_1, v_2, v_3\}, \{v_4, v_5, v_6\}, \cdots, \{v_{n-2}, v_{n-1}, v_n\}$ is a partition satisfying the requirement for $n = |\mathbb{F}_2^m - \{0\}|$. We get $\mathbb{F}_2^{m+2}$ from $\mathbb{F}_2^m$ by appending two bits. For all vectors starting with $\{v_1, v_2, v_3\}$ we give an explicit construction:

$$\begin{array}{lll} v_1 00 & v_2 00 & v_3 00 \\ v_1 01 & v_2 10 & v_3 11 \\ v_1 10 & v_2 11 & v_3 01 \\ v_1 11 & v_2 01 & v_3 10 \end{array}$$

Each row add up to zero, and actually there are 12 such assignments. Clearly we can do this for all 3-subsets. Let v_0 be the all-zero vector in $\mathbb{F}_2^m$, then clearly we have $v_0 01 + v_0 10 + v_0 11 = 0$. Thus we have constructed a partition for $\mathbb{F}_2^{m+2} - \{0\}$. The induction process allows us to count the number of L that satisfies the given condition, but I assume this number is not of theoretical interest.

4.2 A More Complicated Example

Now let us try to push up minimum distance and construct an optimal code that satisfies the following conditions:

$$r = 2, d \geq 10, k = \frac{2}{3}(2^m - 1) - 2m + 1$$

We have already shown that all vectors in $P^\perp$ that have weight smaller than 10 could be written as a sum of two vectors of weight 4, and each vector of weight 2 is a sum of two disjoint vectors in S. Let D be the set of all vectors of weight 4. Suppose we have a injective map $L : D \mapsto \mathbb{F}_2^{2m-1} - \{0\}$ that respects all linear dependence relations within D, then we can extend L linearly to a linear map $\widetilde{L} : P^\perp \mapsto \mathbb{F}_2^{2m-1}$. If L exists, then we can obtain an optimal code.

4.2.1 A Failed Construction

It is very tempting to try to construct an optimal code for $d \geq 10$ by modifying the optimal code for $d = 6$ by appending additional bits to differentiate vectors in D, however, this construction is know to fail. Let M be the linear transformation in the optimal code for

$d = 6$. Let $v_0 \in S$, then consider the set of unordered pairs $B = \{(v_1, v_2) \mid M(v_1 + v_2) = Mv_0\}$. Since each coset contains the same number of pairs, we have

$$|B| = \frac{1}{2^m - 1}\binom{2^m - 1}{2} = 2^{m-1} - 1$$

We need the rest $m - 1$ coordinates to differentiate the all $v_1 + v_2$ in B. To describe the structure of B, we label the 3-blocks as $b_1, b_2, b_3, \cdots, b_{n'}$, where $n' = n/3$ then we label vectors in S as $b_i - 1, b_i - \omega, b_i - \overline{\omega}$, where $\{0, 1, \omega, \overline{\omega}\}$ is the additive group of $\mathbb{F}_4$. If a vector v lies in block b_i, we denote it as $v \in b_i$. Now let $v_0 \in b_i$. It is easy to realize that B partitions $S - \{v_0\}$ into 2-subsets, with one of them (v_0^1, v_2^2) where $v_0^1, v_0^2 \in b_i$. Thus we have

$$\sum_{(v_1,v_2)\in B-\{(v_0^1,v_2^2)\}} (v_1 + v_2) = (\sum_{v\in S} v) - (v_0 + v_0^1 + v_2^2) = 0$$

and also note that $\{v_1 + v_2 \mid (v_1, v_2) \in B - \{(v_0^1, v_2^2)\}\}$ is the set of all vectors of weight 4 in the preimage of v_0 under M. However, there does not exists a set of $2^{m-1} - 2$ elements that sum up to 0 in $\mathbb{F}_2^{m-1}$.

The failed trial to construct the optimal code this way gives the following corollary:

Corollary 4.2. *There does not exist a set of m coordinates in $\mathbb{F}_2^{2m-1}$ onto which an optimal $\widetilde{L}(S)$ projects injectively.*

Remark 4.3. I have no idea if this conclusion is useful for the existence or nonexistence of an optimal code whatsoever, but is proven to be a huge waste of time.

5 Theory of Linear Embedding

Before venturing more to the more complicated example, we can develop some theory. In this section, vector spaces are assumed to be over $\mathbb{F}_2$ and of finite dimension unless indicated otherwise.

5.1 Concepts and Definitions

Definition 5.1. Let S be a subset of a vector space V. Suppose $S = \{v_0, v_1, \cdots, v_{n-1}\}$ and $0 \notin S$. Then we define the *space of linear dependence relations*, denoted as Λ_S as the subspace in $\mathbb{F}_2^n$, where

$$\lambda = (\lambda_0, \lambda_1, \cdots, \lambda_n) \in \Lambda_S \Leftrightarrow \sum_{i=0}^{n-1} \lambda_i v_i = 0 \text{ as a vector in } V$$

We call S an *independent* subset of V if and only if Λ_S is trivial. Additionally, we define the kth power of S as:

$$S^k = \{\sum_{i=0}^{k-1} v_i \mid v_i \in S \text{ for } 0 \le i \le k-1 \text{ and } v_i \ne v_j \text{ for } i \ne j\}$$

Remark 5.2. It is easy to verify that Λ_S is a vector space, so the above definition is successful. Also note that the minimum distance of Λ_S is at least 3 if S contains distinct nonzero vectors. Actually we do not care about the base vector space V of S once Λ_S is known, so we can talk abstractly about a pair (S, Λ_S), which I call *linear substrate.* This pair could be easily used to define a matroid as well, but I feel a representation of a matroid is quite different from a linear embedding, which is really of interest here.

Definition 5.3. Suppose we have map $L : S \subset V \mapsto \mathbb{F}_{2^m}^{\times}$ where $0 \notin S, |S| = n$ and $2^m \geq n$. Then L is called a *linear embedding* if L is injective and respects the linear dependence relations in S, which is equivalent to $\Lambda_S \subset \Lambda_{L(S)}$.

Remark 5.4. Note that $\Lambda_{L(S)}$ could be seen as the shortened code of $\Lambda_{\mathbb{F}_{2^m}^{\times}}$ on $L(S)$.

5.2 Basic Results on the Space $\Lambda_{\mathbb{F}_{2^m}^{\times}}$

Lemma 5.5. *Let V be a subspace in $\mathbb{F}_2^{2^m-1}$ with $d \geq 3$, then*

$$\dim V \leq (2^m - 1) - m$$

.

Proof. $V^{\perp}$ has to map all vectors of weight 1 and the zero vector into different cosets, i.e.

$$2^{\dim V^{\perp}} \geq (2^m - 1) + 1$$

thus $\text{rank} V^{\perp} \geq m$. □

Remark 5.6. Note that we have $\dim \Lambda_{\mathbb{F}_{2^m}^{\times}} = (2^m - 1) - m$, achieving the bound. It should be expected that $\Lambda_{\mathbb{F}_{2^m}^{\times}}$ is the maximum space that represents linear dependence relations among distinct symbols, since it is the set of linear dependence relations from a complete vector space. Heuristically, it should be able to contain all linear dependence relations if those relations do not mess up with injectivity. This intuitive idea is captured in the following corollary:

Corollary 5.7. *A vector $w \in \Lambda_{\mathbb{F}_{2^m}^{\times}}$ if and only if the code with generator matrix $w \bigcup \Lambda_{\mathbb{F}_{2^m}^{\times}}$ still has $d > 2$.*

Remark 5.8. I once thought this corollary would be useful in giving a second proof to the simple example of $d = 6$, however, it is much harder to determine if a vector messes up with minimum weight than to determine if it lies in the span directly. Yet another observation from the proof of optimality is very important. Again consider the columns of a generator matrix of $V^{\perp}$. Each one of the $2^m - 1$ columns is distinct but there are only $2^m - 1$ possibilities, so each nonzero vector in $\mathbb{F}_2^m$ appears exactly once. Thus a code (a vector space) achieving the bound in the lemma is unique up to permutation of columns. Now note that there is another well-know code achieving this bound and we have the following corollary:

Corollary 5.9. *$\Lambda_{\mathbb{F}_{2^m}^{\times}}$ is permutation equivalent to the binary cyclic code of length $2^m - 1$ with the conjugacy class of the primitive root α of $\mathbb{F}_{2^m}$ as the defining set.*

Remark 5.10. This corollary completely determines what the space $\Lambda_{\mathbb{F}_{2^m}^{\times}}$ looks like. With this observation at hand, we can give a second proof of Proposition 4.1: there obviously exist disjoint vectors of weight 3 that cover the coordinates in $\Lambda_{\mathbb{F}_{2^m}^{\times}}$ for m even since the cyclic code with defining set $\{\alpha^i : 3 \mid i\}$ is a subcode of it.

5.3 A First Non-trivial Minor Improvement

Replace $d \geq 14$ condition by $d \geq 6$ in Theorem 5.11, then we replace the inequality on number of additional parity checks by

$$2^b \geq 1 + 3n'$$

which gives $b \geq m$ and hence

$$k \leq \frac{4}{5}(2^m - 1) - m$$

Now I show how a more direct approach and considerations of linear embedding improves this bound given by subcode argument slightly:

Theorem 5.11. *Let C be a code of length $n = 2^m - 1$, where $4 \mid m$, with locality $r = 4$ and minimum distance $d \geq 6$. Assuming disjointness of repair group, the rank k of C is bounded by:*

$$k \leq \frac{4}{5}(2^m - 1) - m - 2$$

Proof. Let P be the matrix formed by disjoint repair groups of weight 5. Let $S = \{v \in P^\perp \mid \mathrm{wt}(v) = 2\}$. Then it is easy to get

$$|S| = \frac{1}{5}(2^m - 1) \times \binom{5}{2} = 2^{m+1} - 2$$

$d \geq 6$ if and only if all vectors in S must be assigned to different cosets by additional parity checks. Injectivity enforces that the target space is at least $m + 1$ dimensional. However, note that the all-one vector is contained in Λ_S, thus by Corollary 5.7 there does not exist a linear embedding $L : S \mapsto \mathbb{F}_2^{m+1} - \{0\}$. Hence S must be embedded to a vector space of at least $m + 2$ dimensions and the conclusion follows. □

5.4 Directions of Future Research

I think of two natural ways to get new lineat substrate from old ones: one is to raise it to a power, the other is to quotient it out by a subset of itself. I also really want to get something useful to determine whether there exists an optimal code for $r = 2, d \geq 10$ case. I also have the following wild guess:

Conjecture 5.12. *For each $d = 2(2l + 1)$, there exists a linear code C of length $2^m - 1$, not necessarily cyclic, that has disjoint repair groups with locality $r = 2$ and dimension*

$$k = \frac{2}{3}(2^m - 1) - lm$$

6 Subcode Argument on Codes over $\mathbb{F}_4$

Subcode argument is useful in giving bounds when each block could contain multiple weights but just as in the first example, it is hard to derive implication on existence of optimal code and in the following example we do not have an easy construction for optimal code.

Theorem 6.1. *Let C be a locally repairable code with length $n = 4^m - 1$, locality $r = 2$, and minimum distance $d \geq 6$. Assuming disjoint repair groups, we have*

$$k \leq \frac{2}{3}(4^m - 1) - m$$

where k is the dimension of C.

Proof. Let $n' = n/3$. Up the permutation of columns, $C^\perp$ contains the matrix:

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \otimes I_{n'}$$

Now consider the subcode C' of C with ${C'}^{\perp}$ containing the matrix:

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix} \otimes I_{n'}$$

Again we have

$$k' \geq k - \frac{1}{3}(4^m - 1)$$

The additional parity checks kill codewords of weight 3 contained in each block, leaving only one binary degree of freedom in each block, so sphere packing gives:

$$4^{k'} \leq \frac{4^{n'}}{1 + 3n'}$$

which yields

$$k' \leq n' - \log_4(1 + 3n') = n' - m$$

and the conclusion follows. □

Remark 6.2. The argument could be easily applied to $d \geq 10$ case. For $d' \geq 10$, we replace the last inequality by:

$$k' \leq n' - \lceil \log_4(1 + 3n' + 9n'(n' - 1)/2) \rceil = n' - 2m$$

for $m \geq 3$ and hence

$$k \leq \frac{2}{3}(4^m - 1) - 2m$$

In both cases, we do not have a construction of optimal code at hand. The closest cyclic codes we have, similar the those in the binary case, have m and $2m$ fewer dimensions than the proven bound for $d \geq 6$ and $d \geq 10$ respectively.