Rank of Doubly Repairable Codes

Ziquan Yang

June 14, 2014

1 Disjoint Repairable Groups

Suppose a code C of length $n = 15m$ has two disjoint repairable groups $P = \{p_i\}_{i=1}^{5m}$ and $Q = \{q_j\}_{j=1}^{3m}$ with $r_1 = 2$ and $r_2 = 5$ in its dual $C^\perp$, then we have the following result on the dimension given by these parity checks:

$$7m \leq \dim \operatorname{span}(P \bigcup Q) \leq 8m - 1$$

Consider the $8m \times 15m$ matrix M with $p_1, p_2, \cdots, p_{5m}$ as its first $5m$ row vectors and $q_1, q_2, \cdots, q_{3m}$ as its last $3m$ vectors. We first show that the left kernel of the matrix $\ker C^T$ is well-structured and give a bound for it. To simply notation, we define a function $\varphi : 2^{[n]} \mapsto 2^P$ by mapping a subset $I \subset [n]$ to the minimal subset of $P' \subset P$ that covers I. Since we assume that p_is are disjoint, this mapping is well defined. Similarly we define $\psi : 2^{[n]} \mapsto 2^Q$ for Q. Support of a set of vectors means the union of supports of all vectors in the set and is also denoted as supp. With the given notation, each vector $w \in \ker M^T$ corresponds to a pair $(P', Q') \in 2^P \times 2^Q$ that satisfy

$$\operatorname{supp}(P') = \operatorname{supp}(Q')$$

This is obvious since if p_i and p_j are disjoint, then addition corresponds nicely to the union of supports: $\operatorname{supp}(p_i + p_j) = \operatorname{supp}(p_i) \bigcup \operatorname{supp}(p_j)$ and similarly for any q_i and q_j. Now given a particular $w \in \ker M^T$ and corresponding $(P', Q') \in 2^P \times 2^Q$. Heuristically we know that to constuct a linear dependence relation $w \in \ker M^T$, once we pick a p_i then we have to pick q_js to cover its support, and again we need to pick vectors in P to cover those q_js' supports, and then we need to pick more vecors in Q until the process terminates. This means that once we pick a vector in P, there is a subset of P and a subset of Q that we have to pick as well. In terms of the structure of $\ker M^T$, this means that each coordinate $k \in [8m]$ (corresponding to either some p_i or some q_j) corresponds to a "minimal" vector v, such that if $k \in \operatorname{supp}(w)$ for some $w \in \ker M^T$, then $\operatorname{supp}(v) \subset \operatorname{supp}(w)$. To formalize the idea, suppose $1 \leq k \leq 5m$ and $p_k \in P'$. We can inductively define a sequence $S_0, S_1, \cdots, S_l$ n $2^{[8m]}$ by: $S_0 = \{p_k\}$ and $S_{l+1} = \psi(\operatorname{supp}(S_l)) \bigcup \varphi(\operatorname{supp}(S_l))$. It is easy to verify that $S_l = S_{l+1}$ if and only if $\operatorname{supp}(P') = \operatorname{supp}(Q') = \operatorname{supp}(S_l)$. The terminal S_l correspond to the minimal vector of k and the case is similar for $5m < k \leq 8m$ giving $q_{k-5m} \in Q$. Clearly the minimal vectors are mutually disjoint, hence linearly independent, and span $\ker M^T$, thus we see that

$$\dim \operatorname{span}(P \bigcup Q) = \operatorname{rank} M = 8m - \#\{\text{minimal vectors}\}$$

Note that $3 \mid |\operatorname{supp}(P')|$ for any $P' \subset P$ and $5 \mid |\operatorname{supp}(Q')|$ for any $Q' \subset Q$, thus $15 \mid |\operatorname{supp}(S_l)|$ for any S_l corresponding to a minimal vector v. Thus the number of minimal vectors is bounded below by 1 and above by m. For each pair of disjoint repair groups, we can associate a matrix A with $a_{ij} = |\text{supp}(p_j) \bigcap \text{supp}(q_i)|$. Note that entries in the same row add up to 5 and entries in the same column add up to 3. Each $w \in \ker M^T$ gives a block decomposition of A after row and column permutation and minimal vectors of a given pair of repair groups together give the finest block decomposition of A up to permutation of rows and columns that preserve blocks. The number of minimal vectors thus equal to the number of blocks in the finest block decomposition. And conversely from any A that satisfies the condition on rows and columns we can reconstruct a pair of disjoint repair groups. Let A_1 denote the matrix associated with any pair of repair groups of length 15, then $A = A_1 \otimes I_m$ gives the lower bound $7m$ dimensions for the repair groups. To show the upper bound is also achievable, consider the following construction of A:

$$\begin{bmatrix} 1 & 3 & 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 3 & 1 & 0 & \cdots & 0 \\ & & & \vdots & & \ddots & & \\ 0 & 0 & 0 & 0 & 0 & 1 & 3 & 1 \end{bmatrix}$$

for odd m (replace the last row by $\cdots 1211$ for even m) and add whatever additional row vectors to balance the sum of each columns. This construction of A can not be further decomposed to blocks and thus achieves maximum rank for the corresponding M. Furthermore, to achieve any rank within the bound we can just use a mix of both constructions for blocks.

2 A Counterexample in the General Case

It might be appealing to make the hypothesis that the $7m$ bound in dimension holds in the general ase, where we do not assume disjoint repair groups. However, this is not true and a concrete counterexample is contructed. Let $m = 8$, $n = 120$, we divide the coordinates into 15 blocks, each of which is of length 8. For each block, consider the following matrix:

$$M = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}$$

We see that $\text{rank} M = 3$ and it covers all 8 coordinates with a parity check of weight 5 and 6 coordinates with a parity check of weight 3, leaving out two in the middle. There are $2 \times 15 = 30$ coordinates uncovered in total and we can cover them with 10 parity checks of weight 3. Together with $M \otimes I_{15}$, we have constructed a $C^\perp$ with two repairable groups with total rank

$$3 \times 15 + 10 = 55 < 56 = 120 \times \frac{7}{15}$$

From the counterexample we also see that the actual lower bound is contingent on m, which makes the general case even harder to handle. With only one repair group, we have seen two theorems giving bounds of the form $c_1 n + c_2 \log_2 n$, where n is the length of the code, c_1 is a fraction and c_2 is a natural number and our tensor product construction reaches optimality by achieving c_1. In the proofs we also see that c_1 is determined by locality and c_2 is mostly determined by minimum distance. The counterexample shows that without assumption on disjoint repair groups, our tensor product construction does not always reach c_1 and it is unlikely that it can "catch up" by minimum distance contraints since $\log_2 n << n$.

3 Initial Ventures to the General Case

In special case we restricted the length to be a multiple of 15 due to divisibility concerns for disjoint repairable groups. Thus if we no longer put the restriction on disjointness we could consider any length. Again we attack the general case by considering linear dependence relations of two repairable groups P and Q. Without loss of generality we can assume that P is set of linearly independent vectors in $C^{\perp}$ with weight 3 and similarly we assume linear independence within Q. Using the same notation, we put vectors in P and Q to a matrix M in a similar way as in section 1. Each linear dependence relation corresponds to a vector in $w \in \ker M^T$ and also to a pair $(P', Q') \in 2^P \times 2^Q$, which satisfies:

$$\sum_{\mathbf{p} \in P'} \mathbf{p} = \sum_{\mathbf{q} \in Q'} \mathbf{q}$$

We call (P', Q') *associated pair of subsets* of w and the support of both side of the above equation *associated coordinates* to w. Note it is defferent from support of w as a vector in $\ker M^T$. In the special case of disjoint repair groups the above condition is equivalent to

$$\bigcup_{\mathbf{p} \in P'} \operatorname{supp}(\mathbf{p}) = \bigcup_{\mathbf{q} \in Q'} \operatorname{supp}(\mathbf{q})$$

With each pair of repair groups, we see that in $2^{[n]}$ there is an interesting subset to consider: the subset of all possible $I \subset [n]$, where I is the set of the associated coordinates to some linear dependence relations. To simplify further discussion, I want to define a function $\zeta_3 : \mathbb{N} \mapsto \mathbb{N}$ by mapping $n \in \mathbb{N}$ to the minimal number of linearly independent vectors in $\mathbb{F}_2^n$ that sum up to $\mathbf{1}_n$, the all 1 vector of length n. Now I would like to leave $\zeta_3(1)$ and $\zeta_3(2)$ undefined since I don't know if certain properties of this function will suggest some canonical extension to $\{1, 2\}$. Clearly we can observe that $\zeta_3(3m) = m$ for any m. I want to break down the design of a pair of repair groups that minimizes dimenstion into three steps:

-