A Study of $(3, 2)$-LRCs

Ziquan Yang

October 14, 2014

1 Introduction

In Sree's paper, a l-available r-locally repairable code is defined as follows:

Definition 1.1. A code of length n is called *l-available r-locally repairable* if for each coordinate $k \in [n]$, there exists l parity checks $p_1, p_2, \cdots, p_l$, each of weight at most $r + 1$, such that for each pair of these parity checks, we have

$$\{k\} = \operatorname{supp}(p_i) \cap \operatorname{supp}(p_j)$$

For simplicity l-available r-locally repairable codes will be denoted as (l, r)-LRCs. The Hamming $[7, 4, 3]$ code gives the dual of a strong 3-available 2-LRC. I call the dual of Hamming $[7, 4, 3]$ code HD code (short for *Hamming Dual*). It is conjectured to have the best rate among all strong 3-available 2-LRC's.

2 Examples

2.1 Hamming [7, 4] Code

2.1.1 Optimality at n=7

We show the optimality of HD code in the following sense:

Theorem 2.1. *Given a set S of 7 vectors $v_0, v_1, \cdots, v_6 \in \mathbb{F}_2^7$, and $wt(v_i) = 3$, then $\dim \operatorname{span}(S) \geq 4$.*

Proof. Let $M = [v_0; v_1; \cdots; v_6]$. Then consider the row space of $(M^T)^\perp$ Note that $d((M^T)^\perp) > 2$ since no two of v_i's are identical. Now note that $d(\Lambda) \neq 3$ since if we have

$$v_i + v_j + v_k = 0$$

then each coordinate that appears in the union of their supports has to appear for an even number of times, however the orders of their supports add up to 9, a contradiction. Therefore $d((M^T)^\perp) \geq 4$. It can be easily observed from the optimality and uniqueness of Hamming code that $\mathrm{rank}((M^T)^\perp) \leq 3$. Thus $\dim \mathrm{span}\,(S) \geq 4$. $\square$

2.1.2 Uniqueness at n=7

Hamming Code is the unique dual code to a $(3, 2)$-LRC of length 7. On each edge, there are 3 pairs of vertices and a pair of vertices determine an edge. Therefore there are 21 pairs of vertices connected by some edge. However, among 7 vertives there are 21 pairs of vertices in total, i.e. the dual code to a $(3, 2)$-LRC has to be a $2 - [7, 3, 1]$ design, which is know to be unique.

2.2 Trivial Code at $n = 8$

2.2.1 Combinatorial Proof

It turns out the only $(3, 2)$-LRC of length 8 is the trivial code. Without loss of generality, let us start from the matrix below:

$$\begin{bmatrix} 1 & 1 & 1 & & & & & \\ 1 & & & 1 & 1 & & & \\ 1 & & & & & 1 & 1 & \end{bmatrix}$$

Now note that there are only 3 vectors incident on the last coordinate. However, we still need 5 vectors. Therefore there are two vectors whose supports are contained in the first seven coordinates. Without loss of generality, we may assume the matrix looks like:

$$\begin{bmatrix} 1 & 1 & 1 & & & & & \\ 1 & & & 1 & 1 & & & \\ 1 & & & & & 1 & 1 & \\ & 1 & & 1 & & 1 & & \\ & & 1 & & 1 & & 1 & \end{bmatrix}$$

At this stage, there is only one way to fill in the other three vectors:

$$\begin{bmatrix} 1 & 1 & 1 & & & & & \\ 1 & & & 1 & 1 & & & \\ 1 & & & & & 1 & 1 & \\ & 1 & & 1 & & 1 & & \\ & & 1 & & 1 & & 1 & \\ & 1 & & & 1 & & & 1 \\ & & 1 & 1 & & & & 1 \\ & & & & 1 & 1 & & 1 \end{bmatrix}$$

This matrix has full binary rank, and the code is also trivially cyclic.

2.2.2 Linear Algebraic Proof

Let $M = [r_0; \cdots; r_7]$ be the incidence matrix of the desired hypergraph. Consider the matrix $A = MM^T$. Then $a_{ij} = \langle r_i, r_j \rangle$, which is 1 if $r_i \cap r_j \neq \emptyset$ and 0 otherwise. Therefore there are exactly one zero on each row or column. Up to permutation of columns, we can rewrite the matrix A as $I_8 + \mathbf{1}_8$, which has full rank. Therefore M must also has full rank.

2.3 Codes of Length 9

2.3.1 Two Examples

$$\begin{bmatrix}
1 & 1 & & & & 1 & & & \\
 & 1 & & 1 & & & & & 1 \\
 & & & & & 1 & & 1 & 1 \\
1 & & & & 1 & & & 1 & \\
 & 1 & 1 & & 1 & & & & \\
 & & & & 1 & & 1 & & 1 \\
 & & 1 & & & 1 & 1 & & \\
 & & & 1 & & & 1 & 1 & \\
1 & & 1 & 1 & & & & &
\end{bmatrix}$$

This code has dimension 1.

$$\begin{bmatrix}
1 & 1 & & & & 1 & & & \\
 & 1 & & & & & 1 & 1 & \\
 & & & & 1 & 1 & 1 & & \\
1 & & & & & & 1 & & 1 \\
 & 1 & 1 & & 1 & & & & \\
 & & & 1 & 1 & & & 1 & \\
 & & 1 & & & & & 1 & 1 \\
 & & & 1 & & 1 & & & 1 \\
1 & & 1 & 1 & & & & &
\end{bmatrix}$$

This code is trivial. I have been trying to classify the codes of length 9.

2.3.2 A Proof for the Upper Bound

If $3/7$ is the optimal rate, then at length 9 the best we can get is rate $3/9 = 1/3$. Now we show it is indeed the case. Again consider the matrix $A = MM^T$. Since $\text{rank}A \leq \text{rank}M$, it suffices to show that $\text{rank}A \geq 6$. A is clearly symmetrix and has exactly two zeros on each row and each column. Let

$B = A + \mathbf{1}_9$. Note that the all one vector lies in the row space of A, so $A^\perp$ only has vectors of even weight. Therefore $A^\perp \subset B^\perp$. The supports for the row vectors of B are pairs that are clearly distinct. Denote these pairs by $p_0, \cdots, p_8$ and consider the graph Γ whose vertices are indexed by these 9 pairs and two vertices are connected by an edge if and only if the two pairs intersect. It is not hard to see that Γ is a disjoint union of polygons and the dimension of $B^\perp$ is exactly the number of polygons. Hence $\dim A^\perp \leq \dim B^\perp \leq 3$ and $\dim A \geq 6$.

2.4 Pair Code

2.4.1 Basic Construction

Let $A = [v_0; v_1; \cdots; v_9]$, there v_i's are the weight 2 codewords in I_5. We claim that the row space of $(A^T)^\perp$ offers the dual to a $(3,2)$-LRC of length 10, i.e. A^T is a generator matrix of $(3,2)$-LRC.
If, say, $v_0 + v_1 + v_2 = 0$, then the three vectors must be of the form $\mathrm{supp}(v_0) = \{i,j\}, \mathrm{supp}(v_1) = \{j,k\}, \mathrm{supp}(v_2) = \{i,k\}$ for some triple $\{i,j,k\}$. In other words, each weight 3 codeword in Λ_S corresponds to a triangle. Since vertices and lines are dual to each other, the triangle can be thought of as having $\{i,j,k\}$ as vertices and their unions (pairs) as edges, or as having pairs as vertices, and their intersections as edges. Each coordinate in Λ_S is contained in exactly 3 such triangles. Furthermore, given two pairs, the third pair in the triangle is fixed. Therefore, two codewords of weight 3 in $(A^T)^\perp$ cannot intersect at more than one coordinates.
The matrix is explicitly given here:

$$\begin{bmatrix}
1 & 1 & & & 1 & & & & & \\
1 & & 1 & & & 1 & & & & \\
1 & & & 1 & & & 1 & & & \\
& 1 & 1 & & & & & 1 & & \\
& 1 & & 1 & & & & & 1 & \\
& & 1 & 1 & & & & & & 1 \\
& & & & 1 & 1 & & 1 & & \\
& & & & 1 & & 1 & & 1 & \\
& & & & & 1 & 1 & & & 1 \\
& & & & & & & 1 & 1 & 1
\end{bmatrix}$$

2.4.2 Properties

Note that $\mathrm{rank}(A) = 4$ since the row space of A is simply the space of even weight vectors in I_5, so the rate of this $(3,2) - LRC$ is $2/5$. I once thought

that it is intuitive that the Hamming Dual has the best rate, but this rate differs from 3/7 only by 1/35.
Moreover, up to permutation of columns, this code turns out to be cyclic:

$$\begin{bmatrix} 1 & 1 & & & 1 & & 1 & & & \\ & 1 & 1 & & & 1 & & 1 & & \\ & & 1 & 1 & & & 1 & & 1 & \\ & & & 1 & 1 & & & 1 & & 1 \end{bmatrix}$$

Hence we have a second example of a cyclic $(3, 2)$-LRC. (I was wrong, I just discovered that this generator polynomial does not divide $x^{10} + 1$).

2.4.3 Alternative Construction

Construct a hypergraph as follows: Let the vertices be 2-subsets of a 5-set S and let the lines be 3-subsets. The 2-subset U is incident on a 3-subset V if and only if $U \subset V$. The binary rank of the incidence matrix of this type of hypergraph is thoroughly studied.

2.5 Telescope Code

The telescope code is also of length 10, but it is non-isomorphic to the pair code and has rank 3. It is named as such since its hypergraph looks like a telescope. Clearly direct sums of $(3, 2)$-LRCs are still $(3, 2)$-LRCs. We are interested in those irreducible ones (those whose corresponding hypergraphs are connected) This example shows that there may be two non-isomorphic irreducible $(3, 2)$-LRC structures for the same length. The matrix for its dual is given here:

$$\begin{bmatrix} 1 & 1 & 1 & & & & & & & \\ & 1 & & & & & & 1 & 1 & \\ & & 1 & & & & & & 1 & 1 \\ 1 & & & & & & 1 & & 1 & \\ & & & & 1 & & & 1 & & 1 \\ & & 1 & & 1 & 1 & & & & \\ & & & & & 1 & 1 & & & 1 \\ & 1 & & 1 & 1 & & & & & \\ & & & 1 & & & 1 & 1 & & \\ 1 & & & 1 & & 1 & & & & \end{bmatrix}$$

2.5.1 Properties

This code has one codeword of weight 4 and 6 codewords of weight 6. This code has a relationship with the pair code. Let the parity check matrix of telescope be denoted by T and that of the pair code by P, we have

$$T^T T + \mathbf{1} = P$$

where $\mathbf{1}$ is the all 1 10×10 matrix.

2.6 Grid Code

2.6.1 Basic Construction

The Grid Code is constructed by considering the 27 points in the $\mathbb{Z}^3$ lattice whose coordinates are ± 1 or 0, and let the 27 lines parallel to x, y or z axis be parity checks. The parity check matrix is hence $M = [l_1; l_2; \cdots; l_{27}]$. Each l_i has weight 3 by construction and two l_i's intersect at at most one coordinate. Therefore $M^\perp$ is a $(3, 2)$-LRC.

2.6.2 Properties

The calculation of rank(M) has an algebraic topological flavor. Consider the row space of $(M^T)^\perp$. We can think of the coordinates of this space as labelled by lines. (I need to refine the following arguments, but still I belive 8 is the answer.)

2.7 Monster Code

This is a $(3, 2)$-LRC of length 16 I constructed when playing around with hypergraphs. I call it monster since it only has rank 1 and is practically useless. The matrix of the dual is given here:

$$\begin{bmatrix}
1 & 1 & 1 & & & & & & & & & & & & & \\
 & & 1 & 1 & 1 & & & & & & & & & & & \\
1 & & & & 1 & 1 & & & & & & & & & & \\
1 & & & & & & 1 & & & 1 & & & & & & \\
 & & & & 1 & & & & 1 & & & & & 1 & & \\
 & & 1 & & & & & 1 & & & & 1 & & & & \\
 & & & & & 1 & 1 & & 1 & & & & & & & \\
 & 1 & & & & & 1 & 1 & & & & & & & & \\
 & & 1 & & & & & 1 & 1 & & & & & & & \\
 & 1 & & & & & & & & & 1 & & & & & 1 \\
 & & & & & 1 & & & & & & & & 1 & & 1 \\
 & & & 1 & & & & & & & & & 1 & & & 1 \\
 & & & & & & & & & 1 & & & & 1 & 1 & \\
 & & & & & & & & & 1 & 1 & 1 & & & & \\
 & & & & & & & & & & & 1 & 1 & 1 & & \\
 & & & & & & & & & & 1 & & 1 & & 1 &
\end{bmatrix}$$

3 Some Progress

Sometimes after a matrix M is defined, the same letter also denotes the row space of the matrix or the set of row vectors. Therefore $\text{rank}M$ is also often written as $\dim M$ and M is sometimes called a subspace.

Definition 3.1. A set of vectors is called *linear* if they intersect at at most one point pairwise. This term is borrowed from hypergraph theory.

Here I present two versions of optimality of Hamming:

Theorem 3.2.

Now I prove a lemma in linear algebra:

Lemma 3.3. *Suppose I have matrix M and M':*

$$M = \left[\begin{array}{c|c} A & O \\ \hline B & C \end{array}\right], \text{ and } M' = \left[\begin{array}{c|c} A & O \\ \hline O & C \end{array}\right]$$

Then $\text{rank}M' \leq \text{rank}M$*. Furthermore,* $\text{rank}M' = \text{rank}M$ *only when* $\text{span}_{\Lambda_C} B \subset \text{span}\, A$*.*

Proof. Let S be the set of coordinates occupied by A, and T be that of C. Then let φ and φ' be the projections of M and M' onto T respectively. Then we have decomposition

$$M = \ker\varphi \oplus \mathrm{I}m\varphi, \text{ and } M' = \ker\varphi \oplus \mathrm{I}m\varphi'$$

□

4 Questions to Look at

4.1 Geometry and Rank

I have been trying to find a method to calculate the binary rank of the incidence matrix of a hypergraph by using geometry. One the hypergraph, a codeword corresponds to a set of vertices, on which each line incident for an even number of times. Let's call such set of points cycles.

4.1.1 Enumeration

One stratedy that I thought of for a long time is just to count the number of cycles. Then the rank is simply $\log_2$ of this number. I tried to use polynomials for this enumeration and instead of using $\mathbb{Z}$ as coefficients, I tried using tensor or wedge product of some rings. But so far this trial has not been successful. If the construction of the coefficient cannot faithfully reflect the specialty of such matrices, then this problem is just as hard to get the weight distribution of a generic code, which is known to be hard.

4.2 Local Property of a Vertex

Another thing I have been considering is how to characterize the local property of a vertex, or how to classify vertices, and how to read off information about the automorphism group of the code. One particularly interesting phenomenon is that many vertices seem to be "special" on the hypergraph, but they really aren't. The pair code is such an example. The vertex in the middle seems to be special but the code is actually cyclic. In particular, this should help to decide whether a code is cyclic up to permutation of columns.

4.2.1 Local Rate

I dream of a local notion of rate of a code and then piece up the local rate to give the global rate. (Like we do in geometry or topology) The local rate at a vertex measures how active it is to participate in a parity check (the parity check of the dual is a codeword of LRC). I still cannot give an exact notion of local rate, but intuitively it should satisfy the following properties:

1. If the automorphism group is transitive on two vertices, then the two vertices should have the same local rate.

2. The global rate of a code lies in between the maximum local rate and the minimum local rate. In particular, if a LRC is cyclic, then the local rate of a vertex equals to the rate of the code.

4.3 Comparing two Codes

I often recall in representation theory how we decide if a representation is contained in another by calculating the inner product of characters. I am wondering if there are some algebraic munipulations that reveal how similar two matrices are. For example, the pair code and the telescope code are pretty similar on many vertices.
I also looked at some hypergraph theory. Although the dual of $(3, 2)$-LRCs can be very nicely described in the language of linear hypergraphs, the 2-rank of the incidence matrix does not seem to be an object of much interest in hypergraph theory.

4.4 Tensor Product Construction and Minimum Distance

If we assume that $3/7$ is the best rate, then pushing up distance only takes up to $\log_2 N$ dimensions, where N is the length of the code. I am still interested in some working tools to push up distance of multiple copies of a simple code, in particular by coset assignment argument. I have tried to do something to $n = 15$, $r = 4$, one repair group, and $d = 6$. On particular stratedy that I am trying to refine is how to induce such an assignment if we can do this for one copy.
At this stage I think what should be within reach is to discuss $(3, 2)$-LRCs are small lengths. I don't think they exist for any length. There may only be two non-isomorphic $(3, 2)$-LRCs of length 10. I wish to show that $\dim 4$ is the best you can get for length 10. If $3/7$ is optimal, then it has to be.

4.5 Missing Pairs

One particularly interesting property of the Hamming code is that every pair of vertices are connected by some hyperedge. Actually for a fixed length n the number of missing pairs is fixed:

$$\binom{n}{2} - 3n = \frac{1}{2}n(n-7)$$

How do the relationships among the missing pairs affect the dimension of the code? We can form a graph by drawing an edge between two intersecting (or disjoint) pairs. For the two codes of length 9, the first one has a hexagon and a triangle and the second one has a nonagon. Actually the zeros in the matrix $M^T M$ correspond to the missing pairs.

4.6 Faith

Still I have faith in Hamming.