On Locally Repairable Codes and Related Incidence Structures

Ziquan Yang

December, 2014

1 Introduction

1.1 Locally Repairable Codes

A locally repairable code (LRC) is a code of length n over a finite field $\mathbb{F}$ such that a failed disk can be reconstructed by accessing at most r other disks. For linear codes, the class of codes of primary concern here, this is equivalent to requiring that for each coordinate c_i, the dual code contains a vector of length at most $r + 1$ whose support contains the given coordinate. Much research has been devoted to understanding the tradeoff between the code dimension k and minimum distance d given a locality r.

1.2 Multiple Availability

A locally repairable code is said to have multiple availability if one failed disk can be reconstructed by more than one ways with the additional restriction that the sets of other disks accessed are disjoint. More precisely we can define [6]

Definition 1.1. A binary linear code C of length n is called s t-available-r-locally repairable code (LRC) if every coordinate i has at least t parity checks of weight $r + 1$ which intersect pairwise in and only in $\{i\}$.

This class of codes are studied in [1], [2], and the explicit construction of these LRCs invariably relies on incidence structures, especially paritial geometries or semiparitial geometries, in which the small weight parity checks used in the definition incident on each coordinate *exactly* t times and satisfy certian connectinity properties brought by the underlying finite geometry structure.

1.3 Contribution

It is conjectured in [6] that the optimal rate for these codes 3-available-2-local LRCs is 3/7. In this report we try to reach a partial answer to this open question by restricting ourselves to the cases when the small weight vectors used for local repairability incident on each coordinate *exactly* 3 times. In this particular case we reformulate the conjecture as:

Conjecture 1.2. *Let M in an $n \times n$ matrix over $\mathbb{F}_2$. Let $\{r_i\}_{i=1}^n$ be its row vectors and $\{c_j\}_{j=1}^n$ be its column vectors. If M satisfies the following conditions:*

1. $\mathrm{wt}(r_i) = 3$
2. $|supp(r_i) \cap supp(r_j)| \leq 1$
3. $\mathrm{wt}(c_j) = 3$

Then $\mathrm{rank}M \geq 4n/7$ *and equality holds only when $n = 7m$ and M is a direct sum of Hamming [7, 4, 3] Code.*

Altough a powerful machinery for induction has yet to be found, by analysing the left null space of M and applying known facts about linear codes of small length, we are able to prove this conjecture in a more general setting for $n = 7, 14, 21$. Hence we make a second conjecture on $n = 7m$ and show that if it is true, then the bound in conjecture 1.2 is ture for $n \equiv 1, 3, 5 \mod 7$ and remain asymptotically good for other congruence classes as well.

2 Optimality at Small Lengths

2.1 At Multiples of 7

As a first step, we restrict ourselves to multiples of 7 but strive to prove in a more general setting, in which we allow extra room to place the vectors:

Conjecture 2.1. *Let M in an $n \times l$ matrix over $\mathbb{F}_2$, where $l \geq n$ and $n = 7m$. Let $\{r_i\}_{i=1}^n$ be its row vectors and $\{c_j\}_{j=1}^l$ be its column vectors. If M satisfies the following conditions:*

1. $\mathrm{wt}(r_i) = 3$
2. $|supp(r_i) \cap supp(r_j)| \leq 1$
3. $\mathrm{wt}(c_j) \leq 3$

Then $\text{rank}M \geq 4n/7$ *and equality holds only when* M *is a direct sum of Hamming Code when the empty columns are removed.*

Now we prove a simple linear algebraic fact, which we will rely on in the subsequent proofs.

Lemma 2.2. *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* $\text{span}_{(C^T)^\perp} B \subset \text{span}\, A$*.*

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

$$M = \ker\varphi \oplus \text{Im}\varphi, \text{ and } M' = \ker\varphi \oplus \text{Im}\varphi'$$

Clearly $\text{Im}\varphi' \subset \text{Im}\varphi$ and equality holds only when $\text{span}_{(C^T)^\perp} B \subset \text{span}\, A$, where

$$\text{span}_{(C^T)^\perp} B := \{\sum_i \mathbf{b}_i c_i : \mathbf{b}_i \in B, (c_i) \in (C^T)^\perp\}$$

□

2.1.1 Optimality at n=7

Consider the left null space $(M^T)^\perp$ Note that $d((M^T)^\perp) > 2$ since no two of v_i's are identical and that $d(\Lambda) \neq 3$ since the all-one vector lies in M^T. Therefore $d((M^T)^\perp) \geq 4$. It can be easily observed from the optimality and uniqueness of Hamming code that $\text{rank}((M^T)^\perp) \leq 3$. Thus $\dim\text{span}\,(S) \geq 4$. Actually Hamming Code also provides the unique matrix in conjecture 1.2 when $n = 7$. There are at most 21 pairs of coordinates and each row vector contains 3 pairs in its support. Therefore each pair of coodinate is contained in exactly one row vector. Hence the row vectors form a 2-$[7, 3, 1]$ design, which we know is unique.

2.1.2 Optimality at n=14

We study the left null space of the matrix M. Suppose $\dim(M^T)^\perp \geq 6$. The table [5] tells us that the minimum distance of $(M^T)^\perp$ is at most 5. Since we know that $d(M^T)^\perp \geq 4$ and it must be even we have that $d(M^T)^\perp = 4$. Now select a vector, say $w \in (M^T)^\perp$ with $\text{wt}(w) = 4$ and project the space $(M^T)^\perp$

onto w and we see that the project has a nontrivial kernel, implying that there exists a vector w' disjoint from w.

If there are two disjoint sets of row vectors $\{v_i\}_{i\in I}$ and $\{v_j\}_{j\in J}$ in M with $\sum v_i = 0$ and $\sum v_j = 0$, then we know that

$$(\bigcup_{i\in I} \operatorname{supp}(v_i)) \cap (\bigcup_{j\in J} \operatorname{supp}(v_j)) = \emptyset$$

Suppose $p \in (\bigcup_{i\in I} \operatorname{supp}(v_i)) \cap (\bigcup_{j\in J} \operatorname{supp}(v_j))$, then we see that there p lies in at least two of the vectors in I and two of the vectors in J and hence p is incident on at least four of the row vectors in M, contradiction.

We assume that w' is a *circuit*, i.e. $\nexists w'' \in (M^T)^\perp, w'' \neq w'$ and $\operatorname{supp}(w'') \subset \operatorname{supp}(w')$. Therefore

$$\dim \operatorname{span} \{v_i\}_{i\in \operatorname{supp}(w')} = \operatorname{wt}(w') - 1$$

Now we discuss $\operatorname{wt}(w')$. If $\operatorname{wt}(w') = 6$, then up to permutation we can divide M in blocks:

$$M = \left[\begin{array}{c|c|c} \Lambda_4 & O & O \\ \hline O & \Lambda_6 & O \\ \hline ? & ? & R \end{array}\right]$$

Where Λ_4 (resp. Λ_6) is the matrix in which the 4 (resp. 6) linearly dependent vectors corresponding to w (resp. w') packed together. Note that $\operatorname{rank} R = 0$. Hence all other nonzero entries should lie below $\Lambda_4 \oplus \Lambda_6$. However, due to weight distribution considerations it is impossible that the rest 4 vectors are all contained in the span of $\Lambda_4 \oplus \Lambda_6$. Hence $\operatorname{wt}(w') = 4$ and we replace Λ_6 with another copy of Λ_4.

$$M = \left[\begin{array}{c|c|c} \Lambda_4 & O & O \\ \hline O & \Lambda_4 & O \\ \hline ? & ? & R \end{array}\right]$$

Now we analyse the residue part R. We see that $\operatorname{rank} R \leq 2$ and R has r rows. Since the span of $\Lambda_4 \oplus \Lambda_4$ does not contain any more vectors of weight 3, if R has a zero row, i.e. if there exists a vector v_i whose support is contained in $\Lambda_4 \oplus \Lambda_4$, then $\operatorname{rank} R \leq 1$. R should at least have 6 nonzero entries and 2 nonzero columns. Due to the constraints given it is impossible that $\operatorname{rank} R \leq 1$. Hence we know that R does not have a zero row. It is not hard to verify that the only possibility of R in this case is

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

At this stage the only way to add in the rest of the nonzero entries would produce a direct sum of two copies of Hamming $[7, 4, 3]$ code.

2.1.3 Optimality at n=21

The table [5] gives a maximum minimum distance of 8 in this case so we have the extra work to eradicate the $d = 8$ and $d = 6$ case. The techniques used are similar to those in previous section so details are spared in the interest of space.

2.2 At Other Congruence Classes

If conjecture 2.1 is true, then we are able to prove conjecture 1.2 for $n \equiv 1, 3, 5 \mod 7$ and the 4/7 bound remains at least asymptotically good for $n \equiv 2, 4, 6 \mod 7$.

2.2.1 Bound at $n \equiv 1 \mod 7$

Let $n = 7m + 1$. Then clearly $\text{rank}M \geq 4m + 1$, and note that

$$4m + 1 > \frac{4}{7}(7m + 1)$$

2.2.2 Bound at $n \equiv 3, 5 \mod 7$

In this case we are able to divide M as in the lemma. When $n \equiv 3 \mod 7$, we let

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

and note that we can apply conjecture 2.1 to C^T. Note that it is impossible for C^T to be a direct sum of Hamming $[7, 4, 3]$ code when empty columns are deleted and hence $\text{rank}C = \text{rank}C^T \geq 4m + 1$. Therefore

$$\text{rank}M \geq \text{rank}A + \text{rank}C \geq 4m + 2 > \frac{4}{7}(7m + 3)$$

Equality for the first two inequalities can actually be achived. See Pair Code discussed in the next section.
When $n \equiv 5 \mod 7$, we let

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

and fortunately we still have

$$\text{rank}M \geq \text{rank}A + \text{rank}C \geq 4m + 3 > \frac{4}{7}(7m + 5)$$

3 Other Examples

It turns out that the matrices M in conjecture 1.2 are interesting objects themselves, although sometimes they do not give interesting 3-available-2-locally repairable codes. When n is large, we can form such matrices by taking direct sums but when n is small, the number of such matrices is pretty limited. Some examples of such matrices are presented here. For some matrices we can also draw the corresponding nice linear hypergraph. For example, the Hamming [7, 4, 3] matrix is the incidence matrix of the following hypergraph:

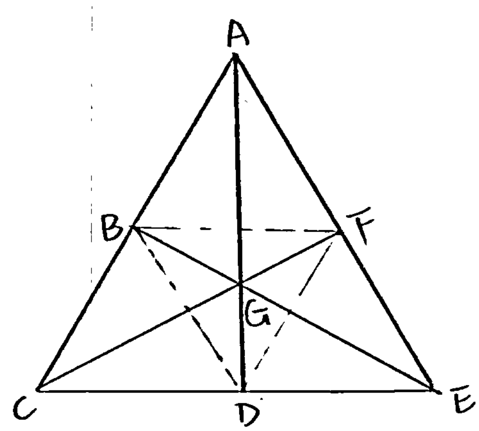

Figure 1: The Fano Plane

3.1 Trivial Code at n = 8

It turns out the only M that satisfies the conditions in conjecture 1.2 at $n = 8$ has full rank. 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:

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

3.2 Two Codes at n = 9

Here we present two matrices satisfying the conditions in conjecture 1.2 and prove the 4/7 bound is valid for $n = 9$.

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

This matrix has rank 8. Its linear hypergraph looks like:

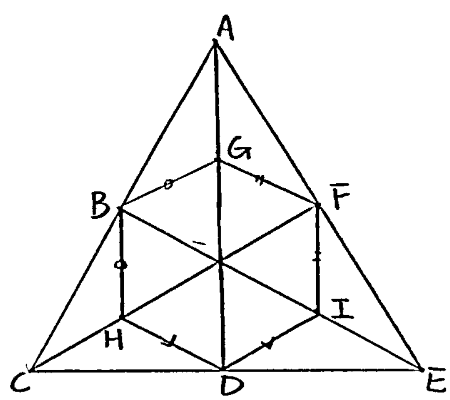

Figure 2: Hypergraph for a Rank 8 Matrix at n=9

Here is another matrix at $n = 9$.

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

This matrix has full rank.

3.2.1 Proof for the Bound

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 $\dim A \geq 6$.

3.3 Pair Code at n = 10

3.3.1 Basic Construction

Let $A = [v_1; v_2; \cdots; v_{10}]$, where v_i's are the weight 2 codewords in I_5. We claim that the left null space $(A^T)^\perp$ provides a M in conjecture at $n = 10$.
If, say, $v_0 + v_1 + v_2 = 0$, then the three vectors must be of the form $\text{supp}(v_0) = \{i, j\}, \text{supp}(v_1) = \{j, k\}, \text{supp}(v_2) = \{i, k\}$ for some triple $\{i, j, k\}$. In other words, each weight 3 codeword in $(A^T)^\perp$ corresponds to a triangle. Each coordinate in $(A^T)^\perp$ 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:

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

This matrix is particularly important as a non-Hamming example. This matrix has rank 4, meet the bound for $n \equiv 3 \mod 7$ with equality and $6/10$ differs from $4/7$ only by $1/35$. On the other hand, the incidence structure does not resemble that of the Haming [7, 4, 3] code. For example, every pair of weight 4 vectors in its left null space intersect at exactly one point, so the left null space can be equivalently spanned by one weight 4 vectors and several weight 6 vectors. We can actually read off those weight 4 vectors directly from its hypergraph. A vector in the left null space is a selection of vertices such that each hyperedge contains an even number of choosen vertices. These vectors are $ABJF$, $CBJD$, $EFJD$, $GHJI$, as shown in the following figure:

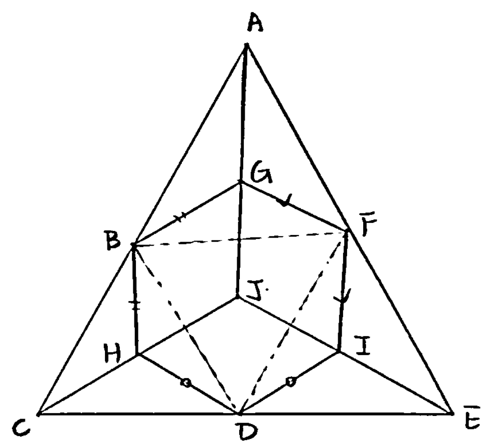

Figure 3: Hypergraph for the Pair Code

3.3.2 General Parameters

In general, let A be the matrix whose rows are the weight 2 vectors from I_n. (A^T). We can form a $(n-2)$-available-2-LRC by taking A^T as the generator matrix. An alternative construction of $(A^T)^\perp$ is to take the inclusion matrix of 2-subsets versus 3-subsets of an n set. The mod 2 rank of such matrices can be obtained by applying the formula given by Wilson [4]. The rate of such code is $2/n$.

3.4 Another Code at n = 10

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:

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

Like the Pair Code, this code also has a very symmetric hypergraph representation: Its linear hypergraph is shown in Figure 4.

3.4.1 Relation to Pair Code

This code has one codeword of weight 4 and 6 codewords of weight 6: $ABHJIF$, $CBGJID$, and $EDHJGF$. 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. Its hypergraph also has an interesting connection to that of Pair Code: one can be obtained from another by rotating the inner hexagon and triangle together.

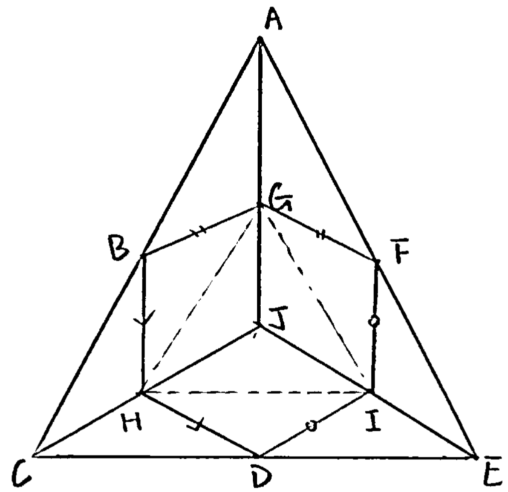

Figure 4: Hypergraph for Another Code at n=10

3.5 Lattice Code at n = 27

We form a vector space as follows: take a set of 3^n coordinates S and associate each of them to a point in $\mathbb{F}_3^n$. Let us denote this association by a map $\varphi : \mathbb{F}_3^n \to S$. To avoid confusion we call the coordinates in S "disks". Now we require $\varphi(v_i), \varphi(v_j), \varphi(v_k)$ to sum up to zero if v_i, v_j, v_k differ pairwise at only one coordinate in $\mathbb{F}_3^n$. Then the vector space generated by these disks in S over $\mathbb{F}_2$ gives us a n-available-2-LRC. To reconstruct one disk $\varphi(v_i)$, we can choose a coordinate of v_i and sum up the two disks whose preimages under φ differ from v_i at precisely the choosen coordinate. At each coordinate in $\mathbb{F}_3^n$ we clearly gain 2 binary degrees of freedom, so the dimension of the vector space is 2^n. The rate of such code is hence $(2/3)^n$. In particular, when $n = 3$ we obtain a 3-available-2-LRC with rate 8/27.

3.6 Visualizing Circuits in Left Null Space

In order to study the generalized conjecture 2.1 we may naturally want to characterize the linear dependence relations among weight 3 vectors that intersect pairwise at at most 1 point, or, equivalently, the left null space of M. As we have seen, the $n = 7, 14, 21$ cases are all proved in this way. Sometimes we can also draw a diagram to represent the incidence relationships of vectors in the left null space. For example, the linear dependence relations of the 7 weight 3 vectors in the Hamming [7, 4, 3] code can be represented by three squares: The vertices are labelled by the support of the vector they represent. It is often not the case that the linear dependence relations can be represented in a neat way. However, we can sometimes use these pictures to deduce some

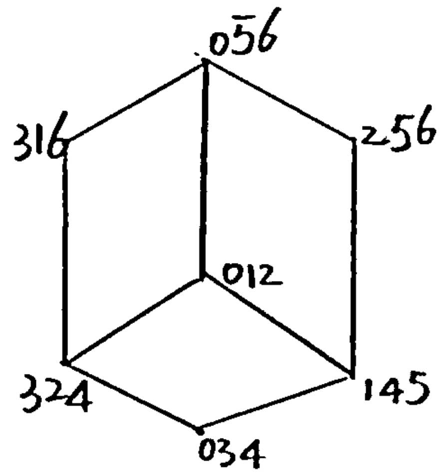

Figure 5: Left Null Space for the Matrix Given by Hamming

forbidden structures. Here is an example:

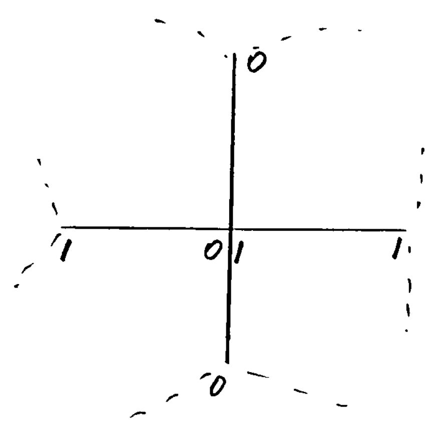

Figure 6: An Example of Forbidden Structure

If we put any coordinate in the center, then we must assign the coordinate twice again in the manner indicated in the diagram. We cannot place a third coordinate at the vertex in the center such that the coordinate appears in the four polygons an even number of times. In other words, in the left null space of M in conjecture 2.1 there cannot be four vectors, such that when punctured onto 5 coordinates, they look like:

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

and when punctured onto the complement of these 5 coordinates they are disjoint.

4 Related Results on Incidence Structures

The main problem that this report is trying to solve lies in a broader context of research: the study of incidence matrices of a collection of sets whose intersections are subject to certain restrictions. Rank, abundance, and existence of these matrices are all frequently frequently discussed topics. This section collects some portions of various researches that are relevant to this theme.

4.1 Handling Two Erasures Locally

Kumar's paper deals with the type of locally repairable codes that can reconstruct two failed disks with the low weight parity checks in the repair group sequentially. Mathematically this amounts to requiring the set of low weight parity checks that covers a given coordinate is unique. This is not a strong enough restriction to imply multiple availability but at least the simple minded nearly pairwise disjoint repair group does not work.
The paper develops interesting machineries relating generalized Hamming weight, minimum distance and k-core of a code, which is a subset of coordinates that does not contain the support of any codewords. One of the interesting theorems related to theme

Theorem 4.1. *Suppose C is a $[n, k]$ linear code and let $\{c_1, \cdots, c_k\}$ be a basis. Let $R_i = supp(c_i)$. If R_i's satisfy the following three requirements:*

- $|R_i \bigcap R_j| \leq 1, \forall i \neq j$
- *any $l \in [n]$ belongs to at most two sets among R_i's*
- $|R_i - \bigcup_{j \neq i} R_j| \geq 1$

then the generalized Hamming weights are given by:

$$d_m(C) = \min_{I \subset [k], |I| = m} |\bigcup_{i \in I} R_i|, 1 \leq m \leq k$$

Proof. Consider any m dimensional subcode of C and suppose $\{v_1, \cdots, v_m\}$ be a set of basis. Without loss of generality, we can reduce the change of basis matrix (this is a slight abuse since the dimensions on both sides do not match) to the standard form by renumbering c_i's if necessary.

$$\begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} = [I_m \mid R_{m\times(b-m)}] \begin{bmatrix} c_1 \\ \vdots \\ c_m \\ c_{m+1} \\ \vdots \\ v_b \end{bmatrix}$$

We can think of $v_1, \cdots, v_m$ as generated by starting with $c_1, \cdots, c_m$ and adding a linear combination of the rest of the $b - m$ vectors to each of them. Again, let $R_i = \text{supp}(c_i)$ and $R'_j = \text{supp}(v_j)$ then we claim that

$$|\bigcup_{i=1}^{m} R'_i| \geq |\bigcup_{i=1}^{m} R_i|$$

Suppose $x \in \bigcup_{i=1}^m R_i$, we observe that x is either preserved by the subsequent linear transformations or is replaced by another coordinate to appear in $\bigcup_{i=1}^m R'_i$. Clearly, if x only belongs to $\bigcup_{i=1}^m R_i$ and does not appear in $\bigcup_{i=m+1}^k R_i$ at all then clearly it is preserved, we only need to discuss the case when $x \in R_i, R_j$ where $1 \leq i \leq m$ and $m+1 \leq j \leq k$. In this case we focus on the jth column vector in $[I \mid B]$ and consider three subcases: $|supp(b)| = 0, 1$ or ≥ 2. In each case we reason that a coordinate is either preserved or compensated by a new coordinate after linear combinations.

Remark 4.2. This result has the significance that it reduces the study of generalized Hamming weight of a code to the study of a particular generator matrix. It is not very common for a vector space to possess a set of basis that reveals its properties better than another.

□

4.2 (0,1)-Matrices with Constant Line Sums 3

In this section we are interested in the following number:

$$\mathcal{A}(n, 3) = |\{M \in \mathbb{F}_2^{n\times n} : \text{wt}(r_i) = \text{wt}(c_j) = 3\}|$$

where r_i and c_j run through all rows and columns respectively.
This is relevant to the main problem since we are studying the bound on the

rank of a paricular subclass of these matrices, i.e. the linear ones. ("linear" in the sense of hypergraphs) Although some examples of this class of matrices are presented in the previous sections, at small length the number of such matrices are still quite limited. I once strived to understand something of the abundance of such linear matrices. The formula for the number of matrices with constant line sum 3 is already hard to prove; however, we not only want to restrict ourselves to the linear ones, but only diffrentiate the matrices up to permutations of rows or columns as well. It will be a hard but interesting counting problem.

Here we present a sketch of the proof of an estimate of $\mathcal{A}(n,3)$ given in van Lint and Wilson's "Introduction to Combinatorics" [?]:

$$\mathcal{A}(n,3) = \frac{(3n)!}{(3!)^{2n}} e^{-2}(1 + O(\frac{1}{n}))$$

Consider $3n$ elements labeled as:

$$1_a, 1_b, 1_c, 2_a, 2_b, \cdots, n_a, n_b, n_c$$

We want to assign three of these elements to each of the n row vectors in a $n \times n$ matrix M. To achieve this we form a permutation of these $3n$ elements and form the corresponding ordered partition into triples $(x, y, z)(u, v, w) \cdots$. In each triple (x, y, z) we do not care about ordering, therefore a matrix M can be given by $(3!)^{2n}$ different permutations. The next step is to eradicate those permutations that assign one coordinate twise to a triple, i.e. a triple like $(5_a, 3_b, 5_c)$ occurs. Now let $1 \le r \le n$, count the permutations that give repetitions in r specified triples and let N_r be the sum of this number over all choices of r triples. Now apply inclusion-exclusion principle:

$$\sum_{r=0}^{\lfloor\sqrt{n}/2\rfloor+1} (-1)^r N_r \le P \le \sum_{r=0}^{\lfloor\sqrt{n}/2\rfloor} (-1)^r N_r$$

and then combine this with the estimate

$$(3n)!\frac{2^r}{r!}(1 + \frac{8r^2}{3n}) \ge N_r \ge (3n)!\frac{2^r}{r!}(1 - \frac{3r^2}{n})$$

whose proof is omitted here in the interest of space.

4.3 Incidence Matrix for Uniform Hypergraphs

Anders Björner and Johan Karlander [?] discussed the mod p rank of incidence matrices for *connected* r-uniform hypergraphs. Connectedness means that given any two vectors v_i, v_j in the incidence matrix, there exists a sequence of vectors $v_i = v_0, v_1, \cdots, v_t = v_j$ such that $|\text{supp}(v_i) \cap \text{supp}(v_j)| = r - 1$. Interestingly, the rank of the matrix is computed by considering the right null space. The observation central to the proof is the following: if $\{[k_1], [k_2]\} = \text{supp}(v_i)\Delta\text{supp}(v_j)$, then for any $w \perp v_i, v_j$, we must have $w |_{k_1} = w |_{k_2}$. (Δ is denotes the symmetric difference) In otherwords, if two coordinates are *connected* by two vectors, then any vector in the right null space must be constant on them. Clearly we can extend this to an equivalence relation and the one equivalence class gives one degree of freedom for the right null space. A much less obvious observation, however, is that each vector has to incident on each equivalence class an equal number of times. Therefore, to each equivalence class we can record this number E_i, $1 \le i \le r$. If $p \mid E_i$ for all is, the mod p rank of the incidence matrix is just the $(n-r)$, otherwise the rank is $(n-r+1)$.

5 Concluding Remarks

This research depends a lot on the study of left null space, which is not very common. In the normal way to present a code (or a vector space) with a generator matrix, the right null space corresponds canonically to the perpendicular complement of the space; however, the left null space is completely determined by the generator matrix and does not relate well to the space spanned by the vectors in the generator matrix, except in their dimensional relations. (There is dual code, but there is no such thing as a transpose code.) In order set up the machinery for induction, I suppose we need to figure out a nice way to classify the forbidden structures so that we are garanteed a certain number of disjoint vectors in the left null space and then set up another machinery to analyze the residue part.

References

[1] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin, "On the Locality of Codeword Symbols", *IEEE Trans. Inform. Theory*, vol. 58, no. 11, pp.6925-6934, Nov. 2012

[2] L. Pamies-Juarez, H. D. Hollmann, and F. Oggier, Locally Repairable Codes with Multiple Repair Alternatives, in arxiv.org, May 2013. [Online]. Available: http://arxiv.org/abs/1302.5518

[3] N. Prakash, V. Lalitha and P. Vijay Kumar, "Codes with Locality for Two Erasures", *2014 IEEE International Symposium on Information Theory*, pp. 1962-1966

[4] Richard M. Wilson, "A Diagonal Form for the Incidence Matrices of t-Subsets vs. s-Susbets", *European Journal of Combinatorics*, vol. 11, pp. 609-615, 1990

[5] Grassl, Markus. "Bounds on the minimum distance of linear codes and quantum codes." Online available at http://www.codetables.de. Accessed on 2014-12-03

[6] S. Goparaju, R. Calderbank, "Binary Cyclic Codes that are Locally Repairable", *2014 IEEE International Symposium on Information Theory*, pp. 676-680

[7] J. van Lint, R. Wilson, *A Course in Combinatorics*, 2nd Edition, Cambridge University Press, 2011

[8] A. Björner, J. Karlander "The mod p Rank of Incidence Matrices for Connected Uniform Hypergraphs", *Europ. J. Combinatorics*, **14**, 151-155, 1993