Notes on Kumar's Paper

Ziquan Yang

September 7, 2014

1 Introduction

This report is mainly a reading note of [1]. The main contribution of this paper is to give a bound on the minimum distance of 2-reconstructible codes with locality r and a given rate and to give a construction for certain combinations of these metrics.

2 Notations and Definitions

Notation 2.1.

$$\mathcal{A}_i := \{c \in C : i \in \mathrm{supp(c)}\}$$

Notation 2.2.

$$\mathcal{B}_0 = \mathrm{span}\{\mathrm{c} \in \mathrm{C}^{\perp}, |\mathrm{supp(c)}| \leq \mathrm{r}+1\}$$

Notation 2.3. C_I denotes the shortened code of C on I and C^I denotes the punctured code of C on I.

Definition 2.4. A code C is 2-reconstructible with locality r if for any pair of disks fails, then there exists a sequence of two parity checks with weight at most $r+1$ that can be used to recover the pair.

Remark 2.5. In practice the above requirements amount to the following conditions: (i) $|\mathcal{A}_i| \geq 1$ and (ii) $\mathcal{A}_i \neq \mathcal{A}_j, \forall i \neq j$. Proof is straightforward.

Definition 2.6. A subset $I \subset [n], |I| = k$ is said to be a k-core of code C if it does not contain the support of any codewords in C. In other words, $C_I = 0$.

3 Theorems and Results

Lemma 3.1. *$\forall I \subset [n], |I| = g_k, \exists I' \subset I$ that is a k-core of C.*

Proof. First observe that $\dim C_I \leq g_k - k$. Suppose not, then $\dim C_I \geq g_k - k + 1$, which implies $d_{g_k-k+1} \leq g_k$, contradicting with the fact that the number of GHWs up to g_k is exactly $g_k - k$.
Let $b' = \dim C_I$ and up to permutation of columns we can represent the generator matrix of C as $[I_{b'} \mid P_{b' \times (g_k - b')}]$. Let T be the support of of $P_{b' \times (g_k - b')}$, then any $I' \subset T$ is a k-core. □

Theorem 3.2. *If $B_0 < C^{\perp}$ and any I that is a l-core of B_0 is also a l-core of $C^{\perp}$, then $d_{\min}(C) = n + 1 - g_k(B_0)$.*

Proof. By the previous lemma, for any $I \subset [n], |I| = g_k(\mathcal{B}_0)$, there exists $I' \subset I, |I| = k$ that is a k-core of $\mathcal{B}_0$. This implies that $\dim C^I = k, \forall |I| = g_k$, i.e. C punctured on any $I, |I| = g_k$ still has full rank, thus any $I, |I| = g_k$ cannot hold any codewords. Thus $d_{\min}(C) \geq n + 1 - g_k(\mathcal{B}_0)$, however since $d_{\min}(C) = n - g_k(C^\perp) + 1 \leq n - g_k(B_0) + 1$, we have equality. □

Theorem 3.3. *For a 2-reconstructible LRC with locality r, $b = \dim \mathcal{B}_0 \geq 2n/(r+2)$.*

Proof. Let $B_0 = \{p_0, \cdots, p_{b-1}\}$ be a basis for $\mathcal{B}_0$ and $S_i = \mathrm{supp}(\mathrm{p_i})$. Now let

$$s_i := |S_i - \bigcup_{j \neq i} S_j|, 1 \leq i \leq b$$

. In other words, s_i is the number of coordinates covered uniquely by p_i. We immediately have $s_i \leq 1$. If the $s_i \geq 2$, then if a pair in S_i fails, there is no way to recover it. Coordinates not in any S_i are covered by more than one vectors, thus we have

$$\sum_{i=0}^{b-1} s_i + 2(n - \sum_{i=0}^{b-1} s_i) \leq b(r+1)$$

$$2n - b \leq 2n - \sum_{i=0}^{b-1} s_i \leq b(r+1)$$

$$b \geq \frac{2n}{r+2}$$

□

Remark 3.4. The use of s_i also appeared in the proof of theorem 3 in Sree's paper.

The following is a pure set-theoretic fact:

Lemma 3.5. *Let T be any set such that $|T| = n \geq r+1$ and let $S_i, 0 \leq i \leq b-1$ be subsets of T such that $\bigcup_{i=0}^{b-1} S_i = T$ and $|S_i| = r+1$. Define*

$$f_m = \min_{I \subset [b], |I| = m} |\bigcup_{i \in I} S_i|, 1 \leq m \leq b$$

then $f_m \leq e_m$ where e_m is defined recursively as:

$$e_b = n$$

$$e_{m-1} = e_m - \lceil \frac{2e_m}{m} \rceil + (r+1), 2 \leq m \leq b$$

Proof. Suppose $f_m \leq e_m$ is true, then we show that $f_{m-1} \leq e_{m-1}$. Now suppose f_m is achieved by $\{S_i | 0 \leq i \leq m-1\}$. Again we consider s_i that is the number of points uniquely covered by S_i. Then we have

$$\sum_{i=0}^{m-1} s_i + 2(f_m - \sum_{i=0}^{m-1} s_i) \leq m(r+1)$$

which gives

$$\sum_{i=0}^{m-1} s_i \geq 2f_m - m(r+1)$$

Now remove the max s_i. Assume s_0 is the largest. Then

$$|\bigcup_{i=1}^{m} S_i| = f_m - s_0 \leq f_m - \frac{\sum_{i=0}^{m-1} s_i}{m} \leq \frac{m-2}{m} e_m + (r+1)$$

□

Applying the above lemma to the set $\mathcal{B}_0$, we see that $d_m(\mathcal{B}_0) \leq e_m, 1 \leq m \leq b$. To minimize $g_k(\mathcal{B}_0)$ we want to maximize all the $d_m(\mathcal{B}_0)$. [1]

4 Construction of Optimal Code Based on Turan Graph

The class of optimal code construction has length with the form:

$$n = \frac{(r+\beta)(r+2)}{2}$$

where $\beta \mid r$.

4.1 Description of Construction

Consider a graph of $b = r + \beta$ vertices. Divide the vertices into $x = (r+\beta)/\beta$ partitions, then place exactly one edge between any two vertices belonging to distinct partitions. Then we have exactly

$$\frac{x(x+1)}{2}\beta^2 = n - b$$

edges, and there are exactly r edges incident on one vertex. Label the vertices with indices $1, 2, \cdots, b$ and then label the edges with indices $b+1, b+2, \cdots, n$. Order does not matter. Now for each vertex, we associate a parity check p_i as follows:

$$\mathrm{supp(p_i)} = \{\mathrm{i}, \mathrm{j_1}, \cdots, \mathrm{j_r} : \mathrm{i} \in \partial(\mathrm{e_{j_k}}), 1 \leq \mathrm{k} \leq \mathrm{r}\}$$

and p_i is specified to be the all-1 vector with the given support.

4.2 Properties of the Constructed Code

Theorem 4.1. *Let* $S_i = \mathrm{supp(p_i)}$. *Then collection of sets* $\{S_i : 1 \leq i \leq b\}$ *achieves the bound* $f_m = e_m$.

Proof. We want to show that $f_b = n$ and f_m's satsify the same recursive relations as e_m, i.e.

$$f_m - f_{m-1} = \lceil \frac{2f_m}{m} \rceil - (r+1)$$

[1] There is a hand-drawn picture here.

Consider any subset M of vertices of order m. We have that

$$|\bigcup_{i=1}^{m} S_i| = m + |E|$$

where

$$E = \{e_j : v \in \partial(e_j) \text{ for some } v \in M\}$$

The rest of the edges are those who only incident on the rest of the $b - m$ vertices, so we want to maximize the number of the edges among any $b - m$ vertices. Thus we want them to be as evenly distributed to the partitions as possible. Apply division with remaider we have

$$b - m = r + \beta - m = ux + v, 0 \leq v \leq x - 1$$

then clearly we should distribute the vertices as illustrated below:

The above chart also clearly shows that if we add one more vertex, then it should at most form $(u + 1)v + u(x - v - 1)$ new edges. Thus we have

$$f_m - f_{m-1} = (u + 1)v + u(x - v - 1) + 1 = v + ux - u + 1$$

since to remove one vertex from M we have 1 fewer vertex and $(u + 1)v + u(x - v - 1)$ fewer edges. Note that in the above equation, u, v are both uniquely determined by m with division with remainder, so this allows us to calculate f_m recursively:

$$f_m = n - \sum_{i=r+\beta}^{m+1} (f_i - f_{i-1})$$

Let $d_i = f_i - f_{i-1}$, then we have that

$$d_i - d_{i-1} = \begin{cases} 0, \text{if } x \mid m \\ 1, \text{otherwise} \end{cases}$$

thus we have

$$\sum_{i=b}^{m+1} b_i = \sum_{i=1}^{ux-u+v} i + \sum_{i=1}^{u} (ix - i + 1)$$

where the first term is the sum of all unique d_i's and the second is that of all repeated terms. Now we apply division by m with remainder again to $2f_m$: $2f_m = mq' + r'$, where

$$q' = (r + 1) + (ux + v - u), r' = (r + \beta) - ux - v\beta + uv$$

. Thus we have

$$\lceil \frac{2f_m}{m} \rceil = q' + 1 = r + 1 + ux + v - u + 1$$

and the desired result follows. □

Theorem 4.2. *Suppose C is a $[n,k]$ linear code and let $\{c_1,\cdots,c_k\}$ be a basis. Let $R_i = \mathrm{supp(c_i)}$. 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:

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

Remark 4.3. The significance of the conclusion is that to calculate d_m we do not need to consider all m dimensional subspaces of C, instead we only need to consider those formed by m vectors in the given basis.

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 $R_i = \mathrm{supp(c_i)}$ and $R'_j = \mathrm{supp(v_j)}$ 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:

- $|\mathrm{supp(b)}| = 0$, in other words, c_j does not participate in the linear combinations operated on $c_1,\cdots,c_m$, then clearly x is preserved.
- $|\mathrm{supp(b)}| = 1$. Then we again divide the case into two subcases, as illustrated below:

- $|\text{supp(b)}| \geq 2$, then like the above case, a new element must be introduced.

□

From the geometric construction of B_0 it is clear that the supports in its parity checks satisfy the conditions in the above theorem, which gives $f_m = d_m$ for B_0.

References

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

[2] V. K. Wei, "Generalized Hamming weights for linear codes", *IEEE Trans. Inform. Theory*, vol. 37, pp.1412-1418, Sept. 1991

[3] W. C. Huffman and V. Pless, *Fundamentals of Error-Correcting Codes*. Cambridge University Press, UK 2003.