Optimal LRC codes for all lengths $n \leq q$

Oleg Kolosov  Alexander Barg  Itzhak Tamo  Gala Yadgar

Abstract—A family of distance-optimal LRC codes from certain subcodes of $q$-ary Reed-Solomon codes, proposed by I. Tamo and A. Barg in 2014, assumes that the code length $n$ is a multiple of $r + 1$. By shortening codes from this family, we show that it is possible to lift this assumption, still obtaining distance-optimal codes.

I. INTRODUCTION

Let $C$ be a $q$-ary code of length $n$ and cardinality $q^k$. We say that $C$ has locality $r$ if for every $i = 1, \ldots, n$ there exists a subset $I_i \subset \{1, \ldots, n\}\setminus\{i\}$, $|I_i| = r$ such that for every codeword $c = (c_1, \ldots, c_n)$ and every $i = 1, \ldots, n$ the coordinate $c_i$ is a function of the coordinates $c_{I_i}$. We call $C$ an $(n, k, r)$ LRC code, and call the subsets $A_i := I_i \cup \{i\}$ repair groups.

Codes with the locality property were introduced in [1], which also proved the following upper bound on the minimum distance of an $(n, k, r)$ LRC code:

$$d_{\min}(C) \leq n - k - \left\lceil \frac{k}{r} \right\rfloor + 2. \quad (1)$$

To call an LRC code optimal if its distance is the largest possible given the other parameters. Several constructions of optimal LRC codes were proposed in the literature, among them [3], [9]. In particular, [8] suggested a family of $q$-ary $(n, k, r)$ LRC codes for any $n \leq q$ such that $(r + 1)n$ that are optimal with respect to the bound [1]. As shown recently in [2], in some cases it is possible to extend this construction to the case $n \leq q + 1$ (still assuming the divisibility).

The codes in [8] are constructed as certain subcodes of Reed-Solomon (RS) codes. Namely, for a given $n$ the code is constructed as a subcode of the RS code of length $n$ and dimension $k + \left\lceil \frac{k}{r} \right\rfloor - 1$. While the “parent” RS code is obtained by evaluating all the polynomials of degree $\leq k + \left\lceil \frac{k}{r} \right\rfloor - 2$, the LRC codes in [8] are isolated by evaluating the subset of polynomials of the form

$$f_a(x) = \sum_{i=0}^{\left\lceil \frac{k}{r} \right\rfloor - 1} \sum_{j=0}^{i} a_{ij}g(x)^j x^i,$$

where $\deg(f_a) \leq k + \left\lceil \frac{k}{r} \right\rfloor - 2$ and where $g(x)$ is a polynomial constant on each of the repair groups $A_i$.

As pointed out in [8], it is possible to lift the condition $(r + 1)|n$, obtaining LRC codes whose distance is at most one less than the right-hand side of (1). At the same time, [8] did not give a concrete construction of such codes, and did not resolve the question of optimality. In this note we point out a way to lift the divisibility assumption, constructing optimal LRC codes for almost all parameters.

Our results can be summarized as follows.

**Theorem I.1.** Suppose that the following assumptions on the parameters are satisfied:

1. Let $s := n \mod (r + 1)$ and suppose that $s \neq 1$;
2. Let $m = \left\lceil \frac{n}{r + 1} \right\rfloor$.

We assume that $\bar{n} := m(r + 1) \leq q$;

Then there exists an explicitly constructible $(n, k, r)$ LRC code $\bar{C}$ whose distance is the largest possible for its parameters $n, k$, and $r$.

Remark: After this note was completed, we became aware that most of its results are implied by an earlier work by A. Zeh and E. Yaakobi [10]. Specifically, we prove a bound on the distance of LRC codes of length $n$ given in Theorem II.3 which is sometimes stronger than the bound [1]. We also construct a family of LRC codes obtained as shortenings of the codes in [8] and use the bounds (1), (11) to show that they have the largest possible minimum distance for their parameters. It turns out that our strengthened bound is a particular case of [10] Thm.6], and that the fact that shortening optimal LRC codes preserves optimality is shown in [10] Thm.13]. This implies that the codes in [8] can be shortened without sacrificing the optimality property.

In this note we give an explicit algebraic construction of the shortened codes from [8], which is not directly implied by [10]. We believe that the construction of codes presents some interest. We also give an independent, self-contained proof of the needed particular case of the bound on their distance.

II. THE CONSTRUCTION

Let $F = \mathbb{F}_q$, let $n$ be the code length and let $r$ be the target locality parameter. As stated above, we assume that $s \neq 1$ (the case $s = 0$ accounts for the original construction in [8] and is included below). Let $t := r + 1 - s$.

Let $\bar{A} \subset F$ be a subset of size $\bar{n}$. Suppose that $\bar{A}$ is partitioned into disjoint subsets of size $r + 1$:

$$\bar{A} = \bigcup_{i=1}^{m} A_i.$$  \hspace{1cm} (2)

The set $A$ of $n$ coordinates of the code $C$ is formed of arbitrary $m - 1$ blocks in this partition, say $A_1, \ldots, A_{m-1}$, and an
arbitrary subset of the block $A_m$ of size $s$ (our construction includes the case of $(r+1)n$ in [8], in which case this subset is empty). Denote by $B$ the subset of $A_m$ that is not included in $A$, so that

$$A = \bigcup_{i=1}^{m-1} A_i \cup (A_m \setminus B).$$

Let $g(x) \in \mathbb{F}[x]$ be a polynomial of degree $r+1$ that is constant on each of the blocks of the partition (2) and let $\gamma$ be the value of $g(x)$ on the points in the set $A_m$. Without loss of generality we will assume that $\gamma = 0$ (if not, we can take the polynomial $g(x) - \gamma$ as the new polynomial $g(x)$). A way to construct such polynomials relies on the structure of subgroups of $F$ and was presented in [8] (see also [4]).

The codewords of $C$ are formed as evaluations of specially constructed polynomials $f(x)$ on the set of points $A$. To define the polynomials, let $k' := k + t$ and define the quantity

$$S_{k',r}(i) = \begin{cases} \lfloor k' \mod r \rfloor & i < k' \mod r, i = 0, \ldots, r - 1. \\ \lfloor k' \mod r \rfloor - 1 & i \geq k' \mod r. \end{cases}$$

Next let $a \in F^k$ be a data vector. Write $a$ as a concatenation of two vectors:

$$(a_{ij}, i = 0, \ldots, r - 1, j = 1, \ldots, S_{k',r}(i)) \quad (b_m, m = 0, \ldots, s - 2).$$

The total number of entries in the vectors in (3) equals

$$\begin{align*} (k' \mod r) \left\lfloor \frac{k'}{r} \right\rfloor + r - (k' \mod r) \left\lfloor \frac{k'}{r} - 1 \right\rfloor &= \left\lfloor \frac{k'}{r} \right\rfloor - r + (k' \mod r) + r - t = k' - t = k, \\
&= (k' \mod r) - r + (k' \mod r) + r - t = k' - t = k, \\
&= \text{a valid representation of the k-dimensional vector } a. \\
\end{align*}$$

Given $a$, let us construct the polynomial

$$f_a(x) = \sum_{i=0}^{r-1} f_i(x)x^i + h_B(x) \sum_{m=0}^{r-t-1} b_m x^m,$$

where $h_B(x) = \prod_{\beta \in B} (x - \beta)$, $\deg(h_B) = t$ is the annihilator polynomial of $B$ and

$$f_i(x) := \sum_{j=1}^{S_{k',r}(i)} a_{ij} x^j.$$
Noting that \( \left\lfloor \frac{n}{r+1} \right\rfloor = \frac{n}{r+1} \) and using (6), we obtain

\[
n - k - \left\lceil \frac{k'}{r} \right\rceil \geq \left\lfloor \frac{n}{r+1} \right\rfloor - \left\lceil \frac{k + t}{r} \right\rceil \geq \left\lfloor \frac{n}{r+1} \right\rfloor - \frac{n - \left\lfloor \frac{n}{r+1} \right\rfloor + t}{r} = \left\lfloor \frac{n}{r+1} \right\rfloor - \frac{n - \frac{n}{r} + t}{r} = \left\lfloor \frac{n}{r+1} \right\rfloor - \frac{n + t}{r} = 0.
\]

Thus the number of nonzero values of \( f_a(x) \) within the support of the codeword is at least two. Hence the mapping (5) is injective on \( F^k \), which proves that \( \dim(C) = k \). The weight of a nonzero vector satisfies \( \text{wt}(c_{a_i}) \geq n - \deg(f_a) \), which together with (9)-(10) proves inequality (7) for the distance of \( C \).

**C. Optimality**

Finally, let us prove that the constructed codes are distance-optimal. The following upper bound on the distance of LRC codes tightens the bound (1) in some cases.

**Theorem III.3.** Let \( C \) be an \((n, k, r)\) LRC such that \( s := n \mod (r + 1) \neq 0 \). Suppose that \( C \) has \( m := \left\lfloor \frac{n}{r} \right\rfloor \) disjoint repair groups \( A_i \) such that \( |A_i| = r + 1, i = 1, \ldots, m-1 \) and \( |A_m| = s \).

If either \( r \mid k \), or \( r \nmid k \) and \( k \mod r \geq s \), then

\[
d_{\text{min}}(C) \leq n - k - \left\lceil \frac{k}{r} \right\rceil + 1. \tag{11}
\]

**Proof:** The minimum distance of a \( q \)-ary \((n, k)\) code (with or without locality) equals

\[
d = n - \max_{I \subseteq [n]} |I| : C_I < q^k \}.
\]

Let \( A'_i \subset A_i \) be an arbitrary subset of size \( |A_i| - 1 \) and let

\[I' = A'_1 \cup \cdots \cup A'_{\left\lfloor \frac{n}{r} \right\rfloor} \cup A_m \]

Note that

\[
|I'| = r \left\lceil \frac{k - 1}{r} \right\rceil + s - 1. \tag{12}
\]

If \( r \mid k \) then, since \( s \leq r \), (12) becomes

\[
k - r + s - 1 \leq k - 1
\]

Similarly, if \( r \nmid k \) and \( k \mod r \geq s \), then (12) becomes

\[
k - 1 - ((k \mod r) - 1) + s - 1 \leq k - 1.
\]

Thus in either case \( |I'| \leq k - 1 \).

If \( |I'| < k - 1 \), let us add to it arbitrary \( k - 1 - |I'| \) coordinates which are not in the set

\[A_1 \cup \cdots \cup A_{\left\lfloor \frac{n}{r} \right\rfloor} \cup A_m,\]

again calling the resulting subset \( I' \). By construction, \( |C_{I'}| \leq q^{k - 1} \).

\[\Delta := \left\lceil \frac{k + t}{r} \right\rceil - \left\lceil \frac{k}{r} \right\rceil \leq 1\]

(since \( t \leq r - 1 \)).

If \( \Delta = 0 \), then the code \( C \) is optimal by (7).

Let us prove that \( C \) is optimal even when \( \Delta = 1 \). Let us find the parameters for which this holds true. Let \( k = ru + v \), where \( 0 \leq v \leq r - 1 \). If \( v = 0 \) (i.e., \( r \mid k \)), then clearly \( \Delta = 1 \). Otherwise, suppose that \( v \geq 1 \) and compute

\[
\left\lceil \frac{k + t}{r} \right\rceil - \left\lceil \frac{k}{r} \right\rceil = u + 1 + \frac{v + 1 - s}{r} - (u + 1),
\]

which equals 1 if and only if \( v + 1 - s > 0 \), i.e., if and only if \( v = k \mod r \geq s \).

In summary, \( \Delta = 1 \) if and only if either \( r \mid k \), or \( r \nmid k \) and \( k \mod r \geq s \). However, in both these cases, according to (11), the maximum possible distance is one smaller than the bound (7). This again establishes optimality of the code \( C \).

**Remark:** In conclusion we note that the code family in [8] affords an easy extension to the case when each repair group is resilient to more than one erasure (i.e., it supports a code with distance \( \rho > 2 \)). The construction in [8] assumes that \((r + \rho - 1) n \). Using the ideas in the previous section, specifically, polynomials of the form (4), it is easy to lift this assumption, obtaining codes that support local correction of multiple erasures for any length \( n \leq q \) such that \( \left\lceil \frac{n}{r + \rho - 1} \right\rceil (r + \rho - 1) \leq q \).

**REFERENCES**

[1] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. IEEE Transactions on Information Theory, 58(11):6925–6934, November 2012.

[2] L. Jin, L. Ma, and C. Xing. Construction of optimal locally repairable codes via automorphism groups of rational function fields, 2017. arXiv:1710.09638.

[3] X. Li, L. Ma, and C. Xing. Optimal locally repairable codes via elliptic curves, 2017. arXiv:1712.03744.

[4] J. Liu, S. Mesnager, and L. Chen. New constructions of optimal locally recoverable codes via good polynomials. IEEE Trans. Inform. Theory, 2018. to appear.

[5] J. Ma and G. Ge. Optimal binary linear locally repairable codes with disjoint repair groups. arXiv:1711.07138.

[6] N. Prakash, G. M. Kamath, V. Lalitha, and P. V. Kumar. Optimal linear codes with a local-error-correction property. In Proc. 2012 IEEE Internat. Sympos. Inform. Theory, pages 2776–2780. IEEE, 2012.

[7] N. Silberstein, A. S. Rawat, O. Koyluoglu, and S. Vishwanath. Optimal locally repairable codes via rank-metric codes. In Proc. IEEE Int. Sympos. Inform. Theory, Boston, MA, pages 1819–1823, 2013.
[8] I. Tamo and A. Barg. A family of optimal locally recoverable codes. 
IEEE Transactions on Information Theory, 60(8):4661–4676, Aug 2014.

[9] I. Tamo, D. S. Papailiopoulos, and A. G. Dimakis. Optimal locally 
repairable codes and connections to matroid theory. In Proc. 2013 IEEE 
Internat. Sympos. Inform. Theory, pages 1814–1818, 2013.

[10] A. Zeh and E. Yaacobi. Bounds and constructions of codes with multiple 
localities, 2016. arXiv:1601.02763.