Alphabet-size dependent bounds for Exact Repair in Distributed Storage

Viveck Cadambe
Department of Electrical Engineering
Pennsylvania State University
University Park, PA 16802
email: viveck@engr.psu.edu

Arya Mazumdar
Department of Electrical and Computer Engineering
University of Minnesota - Twin Cities
Minneapolis, MN 55455
email: arya@umn.edu

***Abstract*—We use some simple techniques to derive a universal alphabet-size dependent bound on the parameters of erasure-correcting regenerating codes. It can be shown that the well-known previous bounds are special cases of our more general bound. In certain cases our bound leads to improvement over prior results.**

I. Introduction

When erasure codes are used for distributed storage systems, repairing failed nodes with a minimal amount of bandwidth is an important criterion that helps reduces network traffic in the system. Motivated by the goal to reduce the network traffic, the repair bandwidth problem for exact repair has been formulated in [5], [18]. Codes that achieve the minimum repair bandwidth are commonly termed *regenerating codes*. Following the formulations of [5], [18], there has been much interest in constructing regenerating codes and studying their properties. The main goal of this paper is to study the impact of size of the alphabet on regenerating codes.

The formulation of [5] involves the characterization of the trade-off between the repair bandwidth and amount of data to be stored. Code constructions corresponding to the two extremal points of the trade-off are termed minimum storage regenerating (MSR) codes and minimum bandwidth regenerating (MBR) codes respectively. Optimal linear code constructions for the MBR point are provided in [9], [10], [12].

MSR codes are maximum distance separable codes with the minimum possible repair bandwidth. Commonly studied MSR codes are vector linear code constructions, where each codeword symbol is interpreted as a vector. The length of the vector corresponding to each symbol is frequently referred to as the *sub-packetization level* of the code. Following the result of [4] which showed the existence of MSR codes using interference alignment, the sub-packetization required for MSR code constructions has been a subject of significant interest. If the length of the code n and the dimension of the code k are given, then the sub-packetization level of MSR code constructions of [16], [17] are expressions that are exponential in the dimension k[1]. Recently, code constructions of [1], [3], [13] have shown that if the rate R of the code and the dimension of the code k are fixed, then there exist code constructions whose sub-packetization is polynomial in the dimension k. Reference [7] provides a lower bound on the sub-packetization level of vector linear MDS codes.

In contrast to previous works which mainly study the sub-packetization level, we study the impact of the size of the alphabet of the code in this paper. In Theorem 1, we provide a bound on the minimum distance of a code in terms of the length, dimension, the repair bandwidth, and the alphabet size of the code. Our bound is information-theoretic, and is therefore applicable for both linear and non-linear code constructions. For a vector linear code, the alphabet size is a product of the sub-packetization level and the field size. Thus, unlike the works of [1], [3], [7], [13], our bound implicitly shed light on the minimum field size required for a given sub-packetization level, when a vector-linear code is used.

Our approach is similar to our previous work [2] that developed an upper bound on the rate of locally recoverable codes, which depended on the size of the alphabet. Like the result of [2], we can use any bound on the minimum distance of an erasure code, in terms of its length and dimension, to obtain a bound on the minimum distance of a regenerating code. The result of [2] matched previous bounds [6], [8] by using the Singleton bound. Similarly, we show that our bounds match the bounds of [5] at the MSR and MBR points by using the Singleton bound on the minimum distance. We show a comparison with the bounds of [5] and provide some new insights in Section IV.

II. System Model

For integers n, k, d, q, a q-ary (n, k, d) code (over an alphabet of size q) consists of a set of codewords $\mathcal{C} \subseteq \{0, 1, \ldots, q-1\}^n$ of cardinality $|\mathcal{C}| = q^k$ such that the Hamming distance between any two elements of $\mathcal{C}$ is at least d. The parameters n, k, d are respectively referred to as the length, the dimension, and the minimum distance of the code.

Arya Mazumdar's research is supported in parts by an NSF CAREER award CCF 1453121 and NSF grant CCF 1318093. Viveck Cadambe's research is supported partially by NSF award CCF 1464336.

[1]If the rate of the code is smaller than $1/2$, then code constructions that have sub-packetization levels that are linear in k have been constructed in [10], [14], [15].

Informally speaking, an (n,k,d) code $\mathcal{C}$ over an alphabet of size $q^m, m \in \mathbb{Z}_+$ has a repair bandwidth of β if every erased symbol can be recovered from downloading β elements from $\{0,1,2,\ldots,q-1\}$ from each of the $n-1$ other symbols. Formally, the repair bandwidth of an (n,k,d) is β if, for every $i,j \in \{1,2,\ldots n\}, i \neq j$ there exist functions $g_{i,j} : \{0,1,\ldots,q^m-1\} \to \{0,1,\ldots,q^\beta-1\}$ and $\chi_i : \{0,1,\ldots,q^\beta-1\}^{n-1} \to \{0,1,2,\ldots,q^m-1\}$ such that for every codeword $(x_1,x_2,\ldots,x_n) \in \mathcal{C}$, the following holds

$$x_i = \chi_i\big(g_{i,1}(x_1), g_{i,2}(x_2), \ldots g_{i,i-1}(x_{i-1}), g_{i,i+1}(x_{i+1}), g_{i,i+2}(x_{i+2}) \ldots, g_{i,n}(x_n)\big).$$

III. Main Result

Given parameters n,k,q, let

$$d^{(q)}_{\text{opt}}(n,k) = \max\{d : \text{ a } q\text{-ary } (n,k,d) \text{ code exists } \},$$

i.e., the maximization is over all possible n-length codebooks $\mathcal{C}$ with dimension k, over an alphabet of size q. We now state our main result.

Theorem 1: An (n,k,d) code over an alphabet of size q^m and with repair bandwidth β satisfies

$$d \leq \min d^{(q^{m-t\beta})}_{\text{opt}}\left(n-t, \frac{mk-t(n-1)\beta+t(t-1)\beta/2}{m-t\beta}\right)$$

where the minimization is over $0 \leq t \leq \min\{n-d, m/\beta\}$.

The proof of the above theorem requires the following lemma.

Lemma 1: Assume, $m \geq \beta$. If there exists an (n,k,d) code over an alphabet of size q^m with repair bandwidth β, then there exists an $(n-1, \frac{mk-\beta}{m-\beta}, d)$ code over an alphabet of size $q^{m-\beta}$ with repair bandwidth β.

Proof: Consider an (n,k,d) code over an alphabet of size q^m with repair bandwidth β. We view each codeword symbol, whose element comes from $\{0,1,2,\ldots,q^m-1\}$, as a vector of length m with entries from $\{0,1,2,\ldots,q-1\}$. Without loss of generality, we can assume that the first q^m-ary symbol can be recovered from the last β entries of each of the remaining $n-1$ vectors. More precisely, we can assume without loss of generality that the function $g_{1,j}, j \neq 1$ (defined in Section II) simply projects the last β entries of jth symbol vector.

The codeword symbols of $\mathcal{C}$ are denoted as m length q-ary vectors $\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n$. Let us denote the first $m-\beta$ symbols of these vectors as $\overline{\mathbf{v}}_1,\overline{\mathbf{v}}_2,\ldots,\overline{\mathbf{v}}_n$ and the last β symbols as $\underline{\mathbf{v}}_1,\underline{\mathbf{v}}_2,\ldots,\underline{\mathbf{v}}_n$. Note that $\mathbf{v}_1 = \chi_1(\underline{\mathbf{v}}_2,\ldots,\underline{\mathbf{v}}_n)$.

We partition the q^{mk} codewords $(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n)$ into $q^{(n-1)\beta}$ partitions. For notational convenience, we index the $q^{(n-1)\beta}$ partitions by the set $\mathcal{I} = \{(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_{n-1}) : \forall \mathbf{w}_i \in \{0,1,\ldots,q-1\}^\beta, i \in \{1,2,\ldots,n-1\}\}$. A codeword $(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n)$ belongs to partition corresponding to vector $(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_{n-1}) \in \mathcal{I}$ if and only if $(\underline{\mathbf{v}}_2,\underline{\mathbf{v}}_3,\ldots,\underline{\mathbf{v}}_n)$ is equal to $(\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_{n-1})$. By the pigeon-hole principle, there is at least one element $(\mathbf{w}^*_1,\mathbf{w}^*_2,\ldots,\mathbf{w}^*_{n-1}) \in \mathcal{I}$ whose corresponding partition set $\mathcal{P} \subset \mathcal{C}$ has $q^{mk-(n-1)\beta}$ codewords.

We use the codewords in partition set $\mathcal{P}$ to construct an $(n-1, \frac{mk-(n-1)\beta}{m-\beta}, d)$ code $\mathcal{C}_1$ over an alphabet of size $q^{m-\beta}$. The codeword $\mathcal{C}_1$ is constructed as follows.

$$\mathcal{C}_1 = \{(\overline{\mathbf{v}}_2,\ldots,\overline{\mathbf{v}}_n) : (\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n) \in \mathcal{P}\}$$

Clearly $\mathcal{C}_1$ is a codebook with length $n-1$ and dimension at least $\frac{mk-(n-1)\beta}{m-\beta}$, over an alphabet of size $q^{m-\beta}$. One thing that remains to be shown is that the minimum distance of $\mathcal{C}_1$ is d. We show this next.

For any two codewords $(\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_n)$ and $(\mathbf{v}^*_1,\mathbf{v}^*_2,\ldots,\mathbf{v}^*_n)$ in $\mathcal{P}$ we show that $\mathbf{v}_1 = \mathbf{v}^*_1$. Because the two codewords belong to the same partition $\mathcal{P}$, we know that

$$(\underline{\mathbf{v}}^*_2,\underline{\mathbf{v}}^*_3,\ldots,\underline{\mathbf{v}}^*_n) = (\underline{\mathbf{v}}_2,\underline{\mathbf{v}}_3,\ldots,\underline{\mathbf{v}}_n). \quad (1)$$

Because the code has repair bandwidth β, we know $\mathbf{v}_1 = \chi_1(\underline{\mathbf{v}}_2,\ldots,\underline{\mathbf{v}}_n)$ and $\mathbf{v}^*_1 = \chi_1(\underline{\mathbf{v}}^*_2,\ldots,\underline{\mathbf{v}}^*_n)$. Therefore, we infer that $\mathbf{v}_1 = \mathbf{v}^*_1$. Now, because any two elements of $\mathcal{P}$ have a distance of d and because all the elements of $\mathcal{P}$ have identical entries in the first co-ordinate, the distance of any two elements of $\mathcal{P}$ when restricted to the last $(n-1)$ co-ordinates is d. Because of (1), we infer that any two codewords of $\mathcal{C}_1$ have a distance of d.

It should also be noted that the repair bandwidth of $\mathcal{C}_1$ is β. Indeed, the repair bandwidth of $\mathcal{P}$ is still β, and we can shorten this code by puncturing the first coordinate which has a fixed value. This completes the proof. ∎

Now we prove Theorem 1 by using Lemma 1 recursively.

Proof of Theorem 1: Consider an (n,k,d) code $\mathcal{C}$ over an alphabet of size q^m with repair bandwidth β. Applying Lemma 1, we get a $(n-1, \frac{mk-(n-1)\beta}{m-\beta}, d)$ code $\mathcal{C}_1$ over an alphabet of size $q^{m-\beta}$ with repair bandwidth β. Now, applying Lemma 1 again to code $\mathcal{C}_1$, we get a $(n-2, \frac{mk-2(n-1)\beta+\beta}{m-2\beta}, d)$ code $\mathcal{C}_2$ over an alphabet of size $q^{m-2\beta}$ with repair bandiwdth β. Proceeding similarly, applying Lemma 1 repeatedly, t times, we get a sequence of codes $\mathcal{C}_3,\mathcal{C}_4,\ldots,\mathcal{C}_t$. The code $\mathcal{C}_t$ has length $n-t$, dimension $\frac{mk-t(n-1)\beta+t(t-1)\beta/2}{m-t\beta}$, distance d, and uses an alphabet of size $q^{m-t\beta}$. The minimum distance d of code $\mathcal{C}_t$ has to satisfy

$$d \leq d^{q^{m-t\beta}}_{\text{opt}}\left(n-t, \frac{mk-t(n-1)\beta+t(t-1)\beta/2}{m-t\beta}\right).$$

This completes the proof. ∎

IV. Discussion

We discuss some implications of Theorem 1 here. In particular, we argue that if we let the alphabet size grow large, our bound matches the well known bounds of [5].

For large alphabets, Singleton bound is the tight limit for distance of codes. Applying the Singleton bound [11, p. 94] to Theorem 1, we get

$$d \leq \min_t\left(n-t-\frac{mk-t(n-1)\beta+\frac{t(t-1)\beta}{2}}{m-t\beta}\right)+1. \quad (2)$$

where the minimization is carried over the constraints $0 \leq t \leq n-d$ and $0 \leq m-t\beta < mk-t(n-1)\beta+t(t-1)\beta/2$.

Simple rearrangement shows that the bound is equivalent to,

$$m \leq \min_t \frac{t\beta(d - \frac{t}{2} - \frac{1}{2})}{t + d - (n - k + 1)}. \quad (3)$$

We can use (2) and (3) depending on the set of parameters given to us.

We use the above to compare our bound with Dimakis et al.'s [5] and, in parallel generate insights into the role of alphabet size.

A. The Minimum Storage Regenerating (MSR) point

- If we set $t = 1$ in Eq. (2) we get,

$$d \leq n - 1 - \frac{mk - (n-1)\beta}{m - \beta} + 1$$

$$\Rightarrow \beta \geq \frac{m(k - (n - d))}{d - 1}. \quad (4)$$

  This is indeed Eq. (5) from [5] when we set $d = n - k + 1$, i.e., when we start with an MDS (maximum distance separable) code.
- We can demonstrate the usefulness of our bound by applying it in conjunction with the MDS conjecture [11, p. 342]. We claim that there does not exist an MDS code with parameters $n = 14, k = 10$ over $q = 2^m, m = 4$, achieving the bound of [5] on the repair bandwidth. The bound of [5] or Eq. (4) says that, for the aforementioned parameters $\beta \geq 1$. Suppose such a code exists. This implies that, using Lemma 1, a code exists with length equal to 13, dimension equal to $\frac{mk-(n-1)\beta}{m-\beta} = 9$ and distance equal to $n - k + 1 = 5$ over an alphabet of size $2^{4-1} = 2^3$. Since the alphabet size is smaller than the dimension, this code does not exist as per the MDS conjecture, which leads to a contradiction. Note that, such conclusion was not possible using only (4).
- More generally, if we set $\beta = \frac{m}{n-k}$, then we must need an $(n-1, k-1, d)$ code over alphabet $q^{m-\beta}$, with repair bandwidth of $\frac{m}{n-k}$ symbols per node.

  Applying the MDS conjecture, we get $q^{m(1-\frac{1}{n-k})} \geq n - 2$, which is a non-trivial, albeit small, improvement of the standard setting where we started with $q^m \geq n - 1$.

B. The Minimum Bandwidth Regenerating (MBR) point

In this section, we show that our bound is tight at the MBR point. In particular, if we set $t = n - d + 1$, then we know that the codebook obtained from removing t symbols should have at most 1 codeword. This is because, on removing $n - d + 1$ symbols, if we had 2 codewords, then we would have two codewords in the original codebook with distance $d - 1$. Therefore, we get

$$mk - (n - d + 1)(n - 1)\beta + (n - d + 1)(n - d)\beta/2 \leq 0$$

$$\Rightarrow \beta \geq \frac{2mk}{(n - d + 1)(n + d - 2)},$$

which matches the repair bandwidth bound of the MBR point of Dimakis et. al [5]. For a hypothetical code that satisfies the above bound with equality, we get, from the Singleton bound

$$d \leq (n - t) - \frac{mk - t(n-1)\beta^* + \frac{t(t-1)\beta^*}{2}}{m - t\beta^*} + 1$$

for all $t \leq n - d$, where $\beta^* = \frac{2mk}{(n-d+1)(n+d-2)}$.

REFERENCES

[1] G. K. Agarwal, B. Sasidharan, and P. V. Kumar. An alternate construction of an access-optimal regenerating code with optimal sub-packetization level. *CoRR*, abs/1501.04760, 2015.

[2] V. Cadambe and A. Mazumdar. An upper bound on the size of locally recoverable codes. In *Proc. IEEE Int. Symp. Network Coding*, June 2013.

[3] V. R. Cadambe, C. Huang, J. Li, and S. Mehrotra. Compound codes for optimal repair in distributed storage. *Asilomar Conference on Signal Processing, 2011*, Nov 2011.

[4] V. R. Cadambe, S. Jafar, H. Maleki, K. Ramchandran, and C. Suh. Asymptotic interference alignment for optimal repair of mds codes in distributed storage. *Information Theory, IEEE Transactions on*, 59(5):2974–2987, 2013.

[5] A. G. Dimakis, P. B. Godfrey, Y. Wu, M. J. Wainwright, and K. Ramchandran. Network coding for distributed storage systems. *IEEE Trans. Inform. Theory*, 56(9):4539–4551, Sep. 2010.

[6] P. Gopalan, C. Huang, H. Simitci, and S. Yekhanin. On the locality of codeword symbols. *IEEE Trans. Inform. Theory*, 58(11):6925–6934, Nov. 2012.

[7] S. Goparaju, I. Tamo, and R. Calderbank. An improved sub-packetization bound for minimum storage regenerating codes. *IEEE Transactions on Information Theory*, 60(5):2770–2779, 2014.

[8] D. S. Papailiopoulos and A. G. Dimakis. Locally repairable codes. In *Proc. Int. Symp. Inform. Theory*, pages 2771–2775, Cambridge, MA, July 2012.

[9] K. Rashmi, N. B. Shah, P. V. Kumar, and K. Ramchandran. Explicit construction of optimal exact regenerating codes for distributed storage. In *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, pages 1243–1249. IEEE, 2009.

[10] K. V. Rashmi, N. B. Shah, and P. V. Kumar. Optimal exact-regenerating codes for distributed storage at the msr and MBR points via a product-matrix construction. *IEEE Trans. Inform. Theory*, 57(8):5227–5239, Aug. 2011.

[11] R. M. Roth. *Introduction to Coding Theory*. Cambridge University Press, 2006.

[12] S. Y. E. Rouayheb and K. Ramchandran. Fractional repetition codes for repair in distributed storage systems. In *Communication, Control, and Computing (Allerton), 2010 48th Annual Allerton Conference on*, pages 1510–1517. IEEE, 2010.

[13] B. Sasidharan, G. K. Agarwal, and P. V. Kumar. A high-rate msr code with polynomial sub-packetization level. *arXiv preprint arXiv:1501.06662*, 2015.

[14] N. B. Shah, K. Rashmi, P. V. Kumar, and K. Ramchandran. Interference alignment in regenerating codes for distributed storage: Necessity and code constructions. *Information Theory, IEEE Transactions on*, 58(4):2134–2158, 2012.

[15] C. Suh and K. Ramchandran. Exact-repair mds code construction using interference alignment. *IEEE Transactions on Information Theory*, 57(3):1425–1442, 2011.

[16] Z. Wang, I. Tamo, and J. Bruck. On codes for optimal rebuilding access. In *49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1374–1381. IEEE, 2011.

[17] Z. Wang, I. Tamo, and J. Bruck. Long mds codes for optimal repair bandwidth. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1182–1186. IEEE, 2012.

[18] Y. Wu and A. Dimakis. Reducing repair traffic for erasure coding-based storage via interference alignment. In *IEEE International Symposium on Information Theory*, pages 2276 –2280, 28 2009-july 3 2009.