Lifted Reed-Solomon Codes with Application to Batch Codes

Lukas Holzbaur[*], **Rina Polyanskaya**[†], **Nikita Polyanskii**[*‡], and **Ilya Vorobyev**[‡]
[*]Technical University of Munich, Germany
[†]Institute for Information Transmission Problems, Russia
[‡] Skolkovo Institute of Science and Technology, Russia
Emails: lukas.holzbaur@tum.de, rev-rina@yandex.ru, nikita.polyansky@gmail.com, vorobyev.i.v@yandex.ru

***Abstract*—Guo, Kopparty and Sudan have initiated the study of error-correcting codes derived by lifting of affine-invariant codes. Lifted Reed-Solomon (RS) codes are defined as the evaluation of polynomials in a vector space over a field by requiring their restriction to every line in the space to be a codeword of the RS code. In this paper, we investigate lifted RS codes and discuss their application to batch codes, a notion introduced in the context of private information retrieval and load-balancing in distributed storage systems. First, we improve the estimate of the code rate of lifted RS codes for lifting parameter $m \geq 4$ and large field size. Second, a new explicit construction of batch codes utilizing lifted RS codes is proposed. For some parameter regimes, our codes have a better trade-off between parameters than previously known batch codes.**

I. Introduction

The concepts of *locality* and *availability* of codes have been subject to intensive studies. Informally, the locality of a code refers to the number of codeword symbols that need to be accessed in order to recover a single codeword or information symbol and availability is the number of such (disjoint) recovery sets. These properties are of interest in a variety of applications, such as load balancing in distributed data storage, cryptography, and low-complexity error correction/detection. Several different notions related to these parameters have been considered in the literature, including, but not limited to, locally recoverable codes, locally decodable/correctable codes, batch codes, private information retrieval (PIR) codes, and codes with the disjoint repair group property.

Lifting of codes is a technique that allows for the design of codes with strong locality and availability properties. It was first studied in [2] in the context of LDPC codes and later employed to design locally correctable codes [3], [4] and codes with the disjoint repair group property [5], [6]. Lifted codes can also be used to construct batch codes, a code class for which several explicit and non-explicit constructions of have been proposed, employing methods based on generalizations of Reed-Muller (RM) codes [7], [8], unbalanced expanders [7], graph theory [9], array and multiplicity codes [10], and finite geometries [8]. In this work, we consider a special notion of batch codes, namely *primitive multiset batch codes*. Informally, a primitive multiset k-batch code (in what follows, we simply write a k-batch code to refer to this class of codes) of length N and dimension n allows for the recovery of any set of k message symbols, possibly with repetition, in k disjoint ways, i.e., for any k-tuple (*batch*) of message symbols $x_{i_1}, ..., x_{i_k}$ with $i_1, ..., i_k \in [n]$ there exist k non-intersecting sets $R_1, ..., R_k \subset [N]$ such that the message symbol x_{i_j} can be recovered from the codeword symbols indexed by the set R_j.

This work is mainly based on the results of [1] (presented at ISIT 2020, available at https://arxiv.org/abs/2001.11981).
L. Holzbaur's work was supported by the Technical University of Munich – Institute for Advanced Study, funded by the German Excellence Initiative and European Union 7th Framework Programme under Grant Agreement No. 291763 and the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Grant No. WA3907/1-1. Rina Polyanskaya and Ilya Vorobyev were supported in part by the Russian Foundation for Basic Research through grant no. 20-01-00559. N. Polyanskii's research was supported in part by the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) under Grant No. WA3907/1-1.

A. Our approach and contribution

We construct batch codes from Reed-Solomon codes by lifting them to a higher dimension, while requiring the restriction of each codeword to a line to be a codeword of the RS code. This has been shown [3] to be equivalent to generating a code by evaluating the polynomials over a vector space $\mathbb{F}_q^m$ from the linear span of all m-variate monomials, such that, when restricted to a line in the space, the resulting univariate polynomial is of degree at most $d < q-1$. An m-variate Reed-Muller (RM) code of order d over a field $\mathbb{F}_q$ restricts the degree of the multivariate polynomials to be at most d and thereby naturally provides this property. However, this causes the rate of the RM code to be very small. Lifted RS codes include not only the multivariate monomials of low degree, as RM codes do, but all polynomials which fulfill the required property. This results in codes with locality properties similar to RM codes, but of significantly higher rate. In Section III we provide a tight estimate of the rate of these codes, thereby improving upon the results of [3]. Additionally, we show that a lifted-RS code is also a k-batch code. This improves the known upper bounds on the redundancy of batch codes in some parameter regimes of sublinear k, i.e., $k = n^{\varepsilon}$ with $n \to \infty$ and $0 \leq \varepsilon \leq 1$.

In Section IV we briefly discuss how the estimate of the rate of lifted RS codes can be used to obtain the best known estimate of the rate of lifted multiplicity codes, a generalization of RM/lifted RS codes.

II. Preliminaries

We start by introducing some notation that is used throughout the paper. Let $[n]$ be the set of integers from 1 to n. A vector is denoted by bold lowercase letters such as $\mathbf{d}$. Let $q = 2^{\ell}$ and $\mathbb{F}_q$ be a field of size q. We write $\log x$ to denote the logarithm of x in base two. In what follows, we fix m to be a positive integer representing the number of variables. For $\mathbf{d} = (d_1, \ldots, d_m) \in \mathbb{Z}_q^m$ and $\mathbf{X} = (X_1, \ldots, X_m)$, let $\mathbf{X}^{\mathbf{d}}$ denote the monomial $\prod_{i=1}^{m} X_i^{d_i}$ from $\mathbb{F}_q[\mathbf{X}]$.

For a function $f : \mathbb{F}_q^m \to \mathbb{F}_q$ and a set $S \subset \mathbb{F}_q^m$ let $f|_S$ denote the restriction of f to the domain S. Abbreviate a set of lines in $\mathbb{F}_q^m$ by

$$\mathcal{L}_m := \left\{ (\mathbf{a}T + \mathbf{b})|_{T \in \mathbb{F}_q} \text{ for } \mathbf{a}, \mathbf{b} \in \mathbb{F}_q^m \right\}.$$

For a positive integer $d < q$, denote the set of univariate polynomials of degree less than d by

$$\mathcal{F}_{d,q} := \{ f(T) \in \mathbb{F}_q[T] : \ \deg(f) < d \}.$$

A. Lifted Reed-Solomon codes

Let us recall the definition of lifted Reed-Solomon codes introduced in [3] in a more general form.

Definition 1 (Lifted Reed-Solomon code, [3])**.** For an integer $m \geq 1$, the *m-dimensional lift of a Reed-Solomon code* (or *$[m, d, q]$-lifted-RS code*) is the code

$$\left\{ (f(\mathbf{a}))|_{\mathbf{a} \in \mathbb{F}_q^m} : \begin{array}{l} f(\mathbf{X}) \in \mathbb{F}_q[\mathbf{X}] \text{ such that} \\ \forall L \in \mathcal{L}_m : \ f|_L \in \mathcal{F}_{d,q} \end{array} \right\}.$$

III. Properties of lifted RS codes

In this section, we investigate the code dimension of lifted RS codes. Our estimate improves upon the result presented in [3, Sections 3.2, 3.4] for $m \geq 3$ and is consistent with the result for $m = 3$ provided in [6].

Let $\binom{b}{\geq a}$ denote the number of ways to choose an (unordered) subset of at least a elements from a fixed set of b elements. For $a < 0$ or $a > b$, we assume that $\binom{b}{a} = 0$.

Theorem 1 (Properties of lifted RS codes)**.**
Code rate and distance: *The rate R and relative distance δ of the $[m, q - r, q]$-lifted-RS code are*

$$R = 1 - \Theta\left((q/r)^{\log \lambda_m - m} \right), \ \delta \geq \frac{r}{q} \quad \text{as } q \to \infty \ ,$$

where λ_m is the largest eigenvalue of the $m \times m$ matrix

$$A_m := \begin{pmatrix} \binom{m}{\geq 1} & \binom{m}{0} & 0 & 0 & \dots & 0 \\ \binom{m}{\geq 3} & \binom{m}{2} & \binom{m}{1} & \binom{m}{0} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{m}{\geq 2j+1} & \binom{m}{2j} & \binom{m}{2j-1} & \binom{m}{2j-2} & \dots & \binom{m}{2j-m+2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \binom{m}{\geq 2m-1} & \binom{m}{2m-2} & \binom{m}{2m-3} & \binom{m}{2m-4} & \dots & \binom{m}{m} \end{pmatrix}.$$

Availability: *Each symbol of a codeword of the $[m, q-r, q]$ lifted RS code can be reconstructed in q^{m-1} different ways, each of which involves a disjoint set of coordinates of the codeword with cardinality $q - 1$.*
Batch code: *The code is a k-batch code for $k = q^{m-2}r$.*

IV. Lifted Multiplicity Codes

Multiplicity codes [11] are another recently introduced class of codes with good locality properties based on RM codes. Here, each codeword symbol not only consists of the evaluation of a degree-restricted multi-variate polynomial, but it also contains the evaluation of all the derivatives of this polynomial up to some order. Similar to the concept of lifting, this generalization provides codes with significantly better rate than RM codes, while providing good locality properties. In particular, it was proved [11] that multiplicity codes represent a family of high-rate locally correctable codes that have very efficient local decoding algorithms.

As both lifted RS codes and multiplicity codes are based on generalizations of RM codes, it is a natural question whether these techniques can be combined to further improve the parameters of the respective codes. Some progress in the study of these *lifted multiplicity codes* has recently been made in [5], [12]. In [12], the authors show asymptotic results for any number of variables and [5] is devoted to improving the existing bounds on the required redundancy in the bi-variate case.

In [13] we use the results established in Section III to estimate the rate of lifted multiplicity codes by generalizing the results on the bi-variate case of [3], [5] to an arbitrary number of variables. Essentially, we investigate the same class of codes as defined in [5], [12]. Informally, the $[m, s, d, q]$ lifted multiplicity code consists of the evaluation (together with the derivatives up to the sth order) of polynomials from $\mathbb{F}_q[X_1, \dots, X_m]$ whose restriction to a line agrees with some polynomial of degree less than d on its first $s - 1$ derivatives. Following a standard approach, we consider a subcode of a lifted multiplicity code which is formed by the linear span of *good* monomials with this property. To count bad monomials, we make use of the result for lifted RS codes $(s = 1)$ derived in Section III and extend them for larger s. Roughly speaking, we prove that there exists a one-to-$\binom{s+m-1}{m-1}$ correspondence between bad monomials for lifted RS codes and groups of bad monomials for lifted multiplicity codes. This enables us to find the exact asymptotic order of the number of bad monomials when q is large. This fraction of good monomials serves as a lower bound on the rate of a lifted multiplicity code. Our estimate is consistent with [5] for $m = 2$ and better than the result of [12] for any $m \geq 2$. Using these results we show that lifted multiplicity codes are PIR codes and provide the best trade-off between length, redundancy, and availability for some parameter regime.

References

[1] L. Holzbaur, R. Polyanskaya, N. Polyanskii, and I. Vorobyev, "Lifted Reed-Solomon codes with application to batch codes," in *2020 IEEE Int. Symp. Inf. Theory (ISIT)*, 2020, pp. 634–639.

[2] E. Ben-Sasson, G. Maatouk, A. Shpilka, and M. Sudan, "Symmetric LDPC codes are not necessarily locally testable," in *IEEE 26th Annu. Conf. Comput. Complex.*, 2011, pp. 55–65.

[3] A. Guo, S. Kopparty, and M. Sudan, "New affine-invariant codes from lifting," in *Proc. 4th Conf. Innov. Theor. Computer Sci.*, 2013, pp. 529–540.

[4] A. Guo, "High-rate locally correctable codes via lifting," *IEEE Trans. on Inf. Theory*, vol. 62, no. 12, pp. 6672–6682, 2015.

[5] R. Li and M. Wootters, "Lifted multiplicity codes and the disjoint repair group property," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.

[6] N. Polyanskii and I. Vorobyev, "Trivariate lifted codes with disjoint repair groups," in *Proc. IEEE XVI Int. Symp. Probl. Redund. Inf. Contr. Syst.*, 2019, pp. 64–68.

[7] Y. Ishai, E. Kushilevitz, R. Ostrovsky, and A. Sahai, "Batch codes and their applications," in *Proc. 36th Annu. ACM Symp. Theory Comput.*, 2004, pp. 262–271.

[8] R. Polyanskaya, N. Polyanskii, and I. Vorobyev, "Binary batch codes with improved redundancy," *IEEE Trans. Inf. Theor.*, 2020.

[9] A. S. Rawat, Z. Song, A. G. Dimakis, and A. Gál, "Batch codes through dense graphs without short cycles," *IEEE Trans. Inf. Theory*, vol. 62, no. 4, pp. 1592–1604, 2016.

[10] H. Asi and E. Yaakobi, "Nearly optimal constructions of PIR and batch codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 947–964, 2018.

[11] S. Kopparty, S. Saraf, and S. Yekhanin, "High-rate codes with sublinear-time decoding," *J. Assoc. Comput. Mach.*, vol. 61, no. 5, p. 28, 2014.

[12] L. Wu, "Revisiting the multiplicity codes: A new class of high-rate locally correctable codes," in *Proc. IEEE 53rd Annu. Allerton Conf. Commun. Contr. Comput. (Allerton)*, 2015, pp. 509–513.

[13] L. Holzbaur, R. Polyanskaya, N. Polyanskii, I. Vorobyev, and E. Yaakobi, "Lifted multiplicity codes," *arXiv preprint arXiv:2008.04717*, 2020.