Correcting on Curves and Highly Sound Locally Correctable Codes of High Rate

Yeow Meng Chee, Wu Liyasi, and Chaoping Xing
School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore
emails: {ymchee, lwu008, xingcp}@ntu.edu.sg

***Abstract*—Locally correctable codes have found numerous applications in complexity theory, cryptography and the theory of fault tolerant computation. Recently, Guo et al. [1], discovered a family of high rate locally correctable codes by considering lifting of multivariate polynomials. In this paper, we extend their method by lifting multivariate polynomials on curves, and generalize the "decoding on curve" algorithm from Reed-Muller codes to these lifted codes to provide correcting algorithms with success probability arbitrarily approaching 1. This gives a family of high rate locally correctable codes that is highly sound.**

I. Introduction

A locally decodable code is a code such that any coordinate of a message can be determined with high probability by looking up a small subset of coordinates of the message's corresponding codeword, even when the codeword has a constant fraction of its coordinates corrupted. In certain applications, a stronger property, called local correctability, is required. A locally correctable code (LCC) is a code such that any coordinate of a corrupted codeword can be determined with high probability by looking up a small subset of coordinates of the corrupted codeword. The concept of LCC originates in the works of Blum and Kannan [2], and Lipton [3] on program checking. We now introduce LCC more formally.

Let X and Y be finite sets. The set of all functions from X to Y is denoted Y^X. The finite field of q elements is denoted by $\mathbb{F}_q$. A *q-ary code* on coordinate set D is a function space $C \subseteq {\mathbb{F}_q}^D$. The *length* of the code C is $L(C) = |D|$, the size of D, and its *rate* is $R(C) = \frac{\log_q |C|}{|D|}$. A family of q-ary codes is said to be of *high rate* if $\lim_{L(C)\to\infty} R(C) \to 1$. For $f, g \in {\mathbb{F}_q}^D$, the *relative distance* between f and g, denoted $\Delta(f, g)$, is the fraction of coordinates in D where f and g differ:

$$\Delta(f, g) = \frac{|\{i \in D : f(i) \neq g(i)\}|}{|D|}.$$

Throughout this paper, we take $D = \mathbb{F}_q^m$, for some positive integer m.

A q-ary code C is (r, δ, ξ)-*locally correctable* if there exists a randomized correcting algorithm $\mathcal{A}$ such that the following two conditions hold:

(i) For all $f \in C$, $\boldsymbol{w} \in \mathbb{F}_q^m$, and all functions $g : \mathbb{F}_q^m \to \mathbb{F}_q$ such that $\Delta(f, g) \leq \delta$, we have

$$\Pr[\mathcal{A}^g(\boldsymbol{w}) = f(\boldsymbol{w})] \geq 1 - \xi,$$

where the probability Pr is taken over the random coin tosses of the algorithm $\mathcal{A}$.

(ii) $\mathcal{A}$ makes at most r queries to g.

We call ξ the *soundness error* of C. A locally correctable code with soundness error $\xi \to 0$ (as its length tends to infinity) is said to be *highly sound*.

In a surprising piece of work, Kopparty, Saraf and Yekhanin [4] showed that for arbitrary $\alpha, \epsilon > 0$, there exists an infinite family of $(O(N^\epsilon), \delta, 0.2)$-locally correctable codes of length N with rate $1-\alpha$, for some constant δ. These codes have come to be known as *multiplicity codes* and they are derived from evaluations of high-order Hasse derivatives of multivariate polynomials.

Guo, Kopparty, and Sudan [1] presented an alternative construction for high-rate locally correctable codes. Recall that in a Reed-Muller codes, all multivariate polynomials of total degree at most d are collected to produced codewords and hence these polynomials become univariate polynomials of degree at most d. The idea employed by Guo et al. [1] is to collect a set of multivariate polynomials with total degree that may be greater than d but that the restriction of every polynomial in this set to a line is still a univariate polynomial of degree at most d. As a result, Guo et al. is able to produce a new and natural family of high-rate locally correctable codes with a sublinear time correcting algorithm. The correcting strategy used is the method of "decoding on lines", namely restricting polynomials to a line. The result of Guo et al. [1] may be summarized as follows.

Theorem 1 (Guo et al. [1])**.** *For arbitrarily small $\delta, \epsilon > 0$ and for infinitely many N, there is a family of $(N^\delta, \tau, 2/3)$-locally correctable code of length N with rate $1 - \epsilon$, where $\tau = 2^{-c}/6$, $c = \lceil b 2^{b\lceil 1/\delta \rceil} \log 1/\epsilon \rceil$, and $b = 1 + \lceil \log \lceil 1/\delta \rceil \rceil$, i.e., $\tau = \Omega(1/\exp(\exp(\log(1/\delta)/\delta) \log(1/\epsilon)))$.*

A. Our Contribution

Recall that a q-ary Reed-Muller code consists of the evaluations on $\mathbb{F}_q^m$ of all m-variable polynomials of total degree at most d. Given a point $\boldsymbol{w} \in \mathbb{F}_q^m$, we want to recover the evaluation of a polynomial f at $\boldsymbol{w}$. To do so, we can shoot a random line ℓ through $\boldsymbol{w}$ and consider the restriction of the polynomial to ℓ. By querying values at $d+1$ points on the line, we can recover the evaluation of f at $\boldsymbol{w}$. This is the idea of "decoding on line" for a Reed-Muller code. One drawback for this "decoding on line" method is that the success probability $1 - \xi$ is small, namely $\xi = (d + 1)\delta$ (see [5]). In order to

978-1-4799-5186-4/14/$31.00 ©2014 IEEE

improve this probability, a so-called "decoding on curves" strategy can be used. In the "decoding on curves" algorithm, the decoder uses a random curve instead of a random line. As a result, it can decode with success probability arbitrarily approaching 1 (see [6], [5]), even when the fraction of errors is nearly 1/2.

In this paper, we explore the lifting of multivariate polynomial codes on curves and design a correcting algorithm based on "decoding on curves". By choosing a random parametric curve of degree two over $\mathbb{F}_q$ and considering the restriction of functions to the curve, the corrector can recover any bit of the codeword with probability approaching 1 arbitrarily close. This gives a family of highly sound locally correctable codes that have high rate (rate tending to 1). More precisely, we have the following result

Main Theorem. *For arbitrarily small $\delta, \epsilon > 0$ and for infinitely many N, there is a family of $(N^\delta, \tau, O_{\epsilon,\delta}(1/N^\delta))$-locally correctable code of length N and rate $1-\epsilon$, where $\tau = 2^{-c}/6$, $c = \lceil b2^{b\lceil 1/\delta \rceil} \log 1/\epsilon \rceil$, and $b = 2 + \lceil \log \lceil 1/\delta \rceil \rceil$, i.e., $\tau = \Omega(1/\exp(\exp(\log(1/\delta)/\delta)\log(1/\delta)\log(1/\epsilon)))$.*

II. Preliminaries

We first introduce terminologies required in this paper. Most of the notations here follow [5], [1]. Let us first recall the construction of the codes given by Guo et al. [1].

For a vector $\mathbf{d} = (d_1, \ldots, d_m) \in \mathbb{Z}^m_{\geq 0}$, we use $\mathbf{x}^{\mathbf{d}}$ to denote the monomial $x_1^{d_1} \cdots x_m^{d_m}$. The code C consists of all polynomials which have degree at most d when restricted to a line. In other words, C is defined as the set

$$\Big\{ f = \sum \gamma x_1^{d_1} \cdots x_m^{d_m} \in \mathbb{F}_q[X_1, \ldots, X_m] : d_i \leq q-1 \text{ and for all } \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{F}_q^m, \text{ we have } \deg(f|_{\boldsymbol{\alpha}+\lambda\boldsymbol{\beta}=\mathbf{0}}) \leq d \Big\}.$$

Definition 1 (Support)**.** *Let $f = \sum_{\boldsymbol{d}} c_{\boldsymbol{d}} \boldsymbol{x}^{\boldsymbol{d}}$ be a function from $\mathbb{F}_q^m$ to $\mathbb{F}_q$. The* support *of f is defined as*

$$\mathrm{supp}(f) = \{\boldsymbol{d} \mid c_{\boldsymbol{d}} \neq 0\}.$$

Given a vector $\mathbf{d} = (d_1, \ldots, d_m) \in \mathbb{Z}^m_{\geq 0}$, we say it is in the degree set of C if there exists a function $f = \sum_{\boldsymbol{d}} c_{\boldsymbol{d}} \boldsymbol{x}^{\boldsymbol{d}} \in C$ such that the monomial corresponding to $\mathbf{d}$ has nonzero coefficient. More formally, we have the following definition.

Definition 2 (Degree set)**.** *Let C be a set of functions from $\mathbb{F}_q^m$ to $\mathbb{F}_q$. The* degree set *of C is*

$$\mathrm{Deg}(C) = \bigcup_{f \in C} \mathrm{supp}(f).$$

Let p be a prime number. For a nonnegative integer a, we denote its base p expansion by $a^{(0)}, a^{(1)}, a^{(2)}, \ldots$, i.e., $a = \sum_i a^{(i)} p^i$ and $0 \leq a^{(i)} < p$. If a and b are non-negative integers and satisfy that $b^{(i)} \leq a^{(i)}$ for every i, then we say that b is in the *p-shadow* of a, denoted $b \leq_p a$. For vectors $\boldsymbol{a} = (a_1, \ldots, a_n)$ and $\boldsymbol{b} = (b_1, \ldots, b_n)$, we say $\boldsymbol{b}$ is in the p-shadow of $\boldsymbol{a}$ if $b_i \leq_p a_i$ for every $i \in \{1, \ldots, n\}$. (see [1].)

We also need Lucas' Theorem to determine the size of codes. Recall Lucas' Theorem, which expresses $\binom{a}{b}$ modulo p in terms of the base p expansions of the integers a and b.

Theorem 2. (Lucas' Theorem.) *For nonnegative integers a and b, we have*

$$\binom{a}{b} \equiv \prod_i \binom{a^{(i)}}{b^{(i)}} \bmod p.$$

Note that $\binom{a}{b} \not\equiv 0 \bmod p$ if and only if $b \leq_p a$.

The following operation introduced by Guo et al. [1] is useful when dealing with degree sets. Let q be a prime power. Then $\bmod^* q$ denotes the operation which maps nonnegative integers to the set $\{0, \ldots, q-1\}$ as follows: for an integer $a \geq 0$,

$$a \bmod^* q = \begin{cases} a \bmod (q-1) & \text{if } a \bmod (q-1) \neq 0 \\ q-1 & \text{if } a \neq 0 \text{ and } (q-1) | a \\ 0 & \text{if } a = 0. \end{cases}$$

The motivation behind the $\bmod^* q$ operation is so that if $a \bmod^* q = b$, then $x^a \equiv x^b \bmod (x^q - x)$.

III. Decoding on curves

Consider the code C consisting of all polynomials which have degree at most d when restricted to a curve. In other words, C is the set

$$\Big\{ f = \sum \gamma x_1^{e_1} \cdots x_m^{e_m} \in \mathbb{F}_q[X_1, \ldots, X_m] : e_i \leq q-1 \text{ and for all } \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{F}_q^m, \deg(f|_{\boldsymbol{\alpha}+\lambda\boldsymbol{\beta}+\lambda^2\boldsymbol{\gamma}=\mathbf{0}}) \leq d \Big\}.$$

Given $\epsilon, \delta < 1$, we set $m = \lceil 1/\delta \rceil$, $b = 2 + \lceil \log m \rceil$ and $c = b2^{bm} \lceil \log 1/\epsilon \rceil$. Suppose that $q = 2^s$ for some positive integer s. Let $N = q^m$, $\gamma = 2^{-c}$, $d = (1 - 2^{-c})q$, and $\tau = \gamma/6$. Suppose that C is the code defined above and g is a received function such that $\Delta(f, g) \leq \tau$ for some $f \in C$. Given a point $\boldsymbol{\alpha} \in \mathbb{F}_q^m$, we want to recover the evaluation of f at this point by looking at the value of g at $O(N^\delta)$ positions. To this end, the corrector picks points $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbb{F}_q^m$ uniformly at random and consider the curve

$$\chi = \{\boldsymbol{\alpha} + \lambda\boldsymbol{\beta} + \lambda^2\boldsymbol{\gamma} : \lambda \in \mathbb{F}_q\}.$$

By using evaluations of function g on this curve, the restriction of f can be recovered. Thus the value at the desired point can be computed easily.

The code C has properties given in the Main Theorem. The following subsections establish this fact.

A. Proof of Main Theorem: Part I

Here, we show that C is an $(N^{\delta}, \tau, O_{\epsilon,\delta}(1/q))$-locally correctable code.

Recall that $d = (1 - 2^{-c})q$. If the fraction of errors on a curve is less than 6τ, we can recover the polynomial successfully. We now show that this event occurs with high probability with Chebyshev's inequality.

Lemma 1 (Chebyshev's Inequality)**.** *Let X be a random variable with mean and variance μ and η^2, respectively. Let k be a positive constant. Then the following inequality holds:*

$$\Pr[|X - \mu| \geq k] \leq \eta^2/k^2.$$

Consider the random variable X denoting the number of corruptions on the set $\chi^* = \{\boldsymbol{w} + \lambda \boldsymbol{v}_1 + \lambda^2 \boldsymbol{v}_2 : \lambda \in \mathbb{F}_q^*\}$. It is easy to verify that

$$E[X] = (q-1)\tau, \qquad \text{and}$$
$$D[X] \leq (q-1)(\tau - \tau^2).$$

An application of Chebyshev's inequality gives

$$\begin{aligned}\Pr[X \geq 6\tau(q-1)] &= \Pr[X - E[X] \geq 6\tau(q-1) - E[X]] \\ &\leq \frac{D[X]}{5\tau(q-1)^2} \\ &= O_{\tau}\left(\frac{1}{q}\right) \\ &= O_{\epsilon,\delta}\left(\frac{1}{q}\right)\end{aligned}$$

which is the desired result.

B. Proof of Main Theorem: Part II

Here, we prove that the code C has high rate.

We begin by determining the degree set of C. Note that

$$\begin{aligned}&(\alpha_i + \beta_i\lambda + \gamma_i\lambda^2)^{e_i} \\ &= \sum_{u_i \leq_2 e_i} \binom{e_i}{u_i} \alpha^{e_i - u_i} (\beta_i\lambda + \gamma_i\lambda^2)^{u_i} \\ &= \sum_{u_i \leq_2 e_i} \binom{e_i}{u_i} \alpha^{e_i - u_i} \left(\sum_{w_i \leq_2 u_i} \binom{u_i}{w_i} \beta_i^{u_i - w_i} \lambda^{u_i - w_i} \gamma_i^{w_i} \lambda^{2w_i} \right).\end{aligned}$$

Thus, by Lucas' Theorem, we can verify that the degree set D is

$$\Bigg\{ \boldsymbol{e} \in \{0, \ldots, q-1\}^m : \text{for all } \boldsymbol{u}, \boldsymbol{w} \in \{0, \ldots, q-1\}^m$$
$$\text{satisfying } \boldsymbol{u} \leq_2 \boldsymbol{e} \text{ and } \boldsymbol{w} \leq_2 \boldsymbol{u}, \text{ we have}$$
$$\sum_{i=1}^{m} (u_i + w_i) \mod^* q \leq d \Bigg\}.$$

Lemma 2. *Let $b = 2 + \lceil \log m \rceil$ and $\boldsymbol{e} = (e_1, \ldots, e_m) \in \{0, \ldots, q-1\}^m$. If there exists an integer $j \in \{s-c, \ldots, s-b\}$ such that for every $i \in \{1, \ldots, m\}$ we have*

$$e_i^{(j)} = e_i^{(j+1)} = \cdots = e_i^{(j+b-1)} = 0,$$

then $\boldsymbol{e}$ is in the degree set of the code.

Proof. Suppose $\boldsymbol{u}, \boldsymbol{w} \in \{0, \cdots, q-1\}^m$ and $\boldsymbol{u} \leq_2 \boldsymbol{e}$, $\boldsymbol{w} \leq_2 \boldsymbol{u}$. Let $z_i = u_i$ for $i = 1, \ldots, m$ and $z_i = w_{i-m}$ for $i = m+1, \ldots, 2m$. Due to the property of $\boldsymbol{e}$, we have $z_i^{(j)} = \cdots = z_i^{(j+b-1)} = 0$ for every $i \in \{0, \cdots, 2m\}$. Let $z = \sum_{i=1}^{2m} z_i \mod^* q$. We should show that $z \leq d$. Note that $d^{(0)} = \cdots = d^{(s-c-1)} = 0$ and $d^{(s-c)} = \cdots = d^{(s-1)} = 1$. Thus, $z < d$ if and only if at least one of $z^{(s-c)}, \cdots, z^{(s-1)}$ is 0.

Let $z_i' = 2^{s-j-b} z_i \mod^* q$ for $i = 1, \cdots, 2m$, and $z' = \sum_{i=1}^{2m} z_i' \mod^* q$. It is obvious that $z^{(k)} = z'^{(s-j-b+k \bmod s)}$. Therefore, we have $z_i'^{(s-b)} = \cdots = z_i'^{(s-1)} = 0$, which implies that $z_i' < 2^{s-b}$. Then $z' = \sum_{i=1}^{2m} z_i' < 2m \cdot 2^{s-b} \leq 2^{s-1}$. As a result, we have $z'^{(s-1)} = 0$ thus $z^{j+b-1} = 0$. The fact that $j+b-1$ lies in $\{s-c, \ldots, s-1\}$ completes the proof. □

Pick $\boldsymbol{u} \in \mathbb{F}_q^m$ randomly. For each $u_i = \sum_{j=0}^{s-1} u_i^{(j)} 2^j$, consider the last c coordinates of the expansion, $u_i^{(s-c)}, \ldots, u_i^{(s-1)}$, as $\frac{c}{b}$ blocks of size b each. The probability that there exists an all-zero block is $1 - (1 - 2^{-m(b)})$, which is at least $1 - e^{-\frac{c}{(b)2^{mb}}}$. Thus, the dimension of the code is at least $(1-\epsilon)N$, implying the rate of the code is $1 - \epsilon$.

IV. Conclusion

In this paper, we generalize the results of Guo et al. [1] by lifting multivariate polynomials on curves to construct high rate codes. A corresponding correcting algorithm that uses the "decoding on curves" strategy is developed to correct errors with high probability, even when the number of errors is a constant fraction of the code length. This gives a family of high rate locally correctable codes that are highly sound.

References

[1] A. Guo, S. Kopparty, and M. Sudan, "New affine-invariant codes from lifting," in *ITCS '13 – Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*. New York: ACM, 2013, pp. 529–540.

[2] M. Blum and S. Kannan, "Designing programs that check their work," *J. Assoc. Comput. Mach.*, vol. 42, no. 1, pp. 269–291, 1995.

[3] R. J. Lipton, "Efficient checking of computations," in *STACS 90 – Proceedings of the 7th Annual Symposium on Theoretical Aspects of Computer Science*, ser. Lecture Notes in Comput. Sci., C. Choffrut and T. Lengauer, Eds., vol. 415. Springer, 1990, pp. 207–215.

[4] S. Kopparty, S. Saraf, and S. Yekhanin, "High-rate codes with sublinear-time decoding," in *STOC '11 – Proceedings of the 43rd ACM Symposium on Theory of Computing*, 2011, pp. 167–176.

[5] S. Yekhanin, *Locally Decodable Codes*, ser. Foundations and Trends in Theoretical Computer Science. NOW, 2012, vol. 6, no. 3.

[6] P. Gemmell and M. Sudan, "Highly resilient correctors for polynomials," *Inform. Process. Lett.*, vol. 43, no. 4, pp. 169–174, 1992.