Electronic Colloquium on Computational Complexity, Report No. 69 (2009)

A note on Efremenko's Locally Decodable Codes

Parikshit Gopalan
MSR-Silicon Valley
parik@microsoft.com

There have been three beautiful recent results on constructing short locally decodable codes or LDCs [Yek07, Rag07, Efr09], culminating in the construction of LDCs of subexponential length. The initial breakthrough was due to Yekhanin who constructed 3-query LDCs of sub-exponential length, assuming the existence of infinitely many Mersenne primes [Yek07]. Raghavendra presented a clean formulation of Yekhanin's codes in terms of group homomorphisms [Rag07]. Building on these works, Efremenko recently gave an elegant construction of 3-query LDCs which achieve sub-exponential length unconditionally [Efr09].

In this note, we observe that Efremenko's construction can be viewed in the framework of Reed-Muller codes: the code consists of a linear subspace of (multilinear) polynomials in $\mathbb{F}_q[X_1, \ldots, X_n]$, evaluated at all points in $(\mathbb{F}_q^\star)^n$. We stress that this is not a new construction, but just a different view of [Efr09]. In this view, the decoding algorithm is similar to traditional local decoders for Reed-Muller codes, where the decoder essentially shoots a line in a random direction and decodes along it (see for instance [STV01]). The difference is that the monomials which are used are not of low-degree, they are chosen according to a suitable set-system. Further, the lines for decoding are *multiplicative*, a notion we will define shortly.

The Code Construction. Let $\mathbb{F}_q$ be a finite field with q elements, $\mathbb{F}_q^\star$ its multiplicative group, and let $m = |\mathbb{F}_q^\star|$. We think of q and m as constants (say 7 and 6 for concreteness). Given $L \subset \mathbb{Z}_m$ and an integer x, we say $x \in L \bmod m$ if $x \bmod m \in L$.

Definition 1. Let $L \subseteq \mathbb{Z}_m \setminus \{0\}$. A set system $\mathcal{F}$ consisting of subsets of a universe $[n]$ is said to be L-intersecting if the following conditions hold:

- For every set $S \in \mathcal{F}$, $|S| \equiv 0 \bmod m$.
- For every $S \neq T \in \mathcal{F}$, $|S \cap T| \in L \bmod m$.

If m is a prime power, then $|\mathcal{F}|$ can be at most polynomial in n [Gop06]. For composite m with two or more prime factors, Grolmusz shows that $|\mathcal{F}|$ can be super-polynomial in n [Gro00].

Lemma 2. *If m has t distinct prime factors, then there is an (explicit) L-intersecting family $\mathcal{F}$ of subsets of $[n]$ such that $\ell = |L| \leq 2^t - 1$ and $f = |\mathcal{F}| \geq \exp\left(\frac{(\log n)^t}{(\log\log n)^{t-1}}\right)$.*

We now describe the code $\mathcal{C}_{\mathcal{F}}$.

- **Message Space:** For each set $S \in \mathcal{F}$, define a monomial $X_S = \prod_{i \in S} X_i$. The messages in $\mathcal{C}_{\mathcal{F}}$ correspond to polynomials of the form $P(X) = \sum_{S \in \mathcal{F}} \lambda_S X_S$ where $\lambda_S \in \mathbb{F}_q$.

ISSN 1433-8092

- **Encoding:** The encoding is the evaluation of the polynomial P at all points in $(\mathbb{F}_q^\star)^n$.

It follows that $\mathcal{C}_\mathcal{F}$ is linear over $\mathbb{F}_q$, it has dimension f and length $(q-1)^n$. We will give a local decoder for it with query complexity $\ell + 1$.

The Local Decoder. Let γ be a generator of $\mathbb{F}_q^\star$. Let $B = \{\gamma^c | c \in L\} \subset \mathbb{F}_q^\star$. Note that $1 \not\in B$. For a scalar $\lambda \in \mathbb{F}_q$, a vector $a \in (\mathbb{F}_q^\star)^n$, and $T \subset [n]$ let $\lambda \odot_S a$ denote the vector obtained by multiplying co-ordinates of a in S by λ (and leaving the rest unchanged).

The following lemma is the key to decoding.

Lemma 3. *Let $S, T \in \mathcal{F}$. Then for any $i \geq 0$,*

- $X_S(\gamma^i \odot_S a) = X_S(a)$
- $X_T(\gamma^i \odot_S a) = \mu^i X_T(a)$ *where* $\mu = \gamma^{|S \cap T|} \in B$.

Proof. We prove the claim when $i = 1$, the case of general i follows by repeated application of this claim. It is easy to see that $X_T(\gamma \odot_S a) = \gamma^{|S\cap T|} X_S(a)$. If $S = T$, then $|S \cap T| = |S| \equiv 0 \bmod m$, hence $\gamma^{|S\cap T|} = 1$. Whereas if $S \neq T$, then $\gamma^{|S \cap T|} = \mu \in B$. □

Let us define the *multiplicative* line through $a \in (\mathbb{F}_q^\star)^n$ in the direction $S \subseteq [n]$ as the set of points $\{a, \gamma \odot_S a, \gamma^2 \odot_S a, \ldots\}$. Lemma 3 says that X_S is the unique monomial that stays constant along this line. The decoder uses this to recover λ_S. We need the following claim from [Efr09]

Claim 4. *There exist $c_0, \ldots, c_\ell \in \mathbb{F}_q$ such that $\sum_{i=0}^{\ell} c_i = 1$ and $\sum_{i=0}^{\ell} c_i \mu^i = 0$ for $\mu \in B$.*

The c_is are the coefficients of a univariate polynomial that vanishes on B, suitably rescaled.

We now state the decoding algorithm. The algorithm has query access to P and is given $S \in \mathcal{F}$ as input. The goal is to return λ_S.

1. Pick $a \in (\mathbb{F}_q^\star)^n$ at random, query the values $P(a), P(\gamma \odot_S a), \ldots, P(\gamma^\ell \odot_S a)$.
2. Return $(\sum_{i=0}^{\ell} c_i P_i(\lambda^i \odot_S a)) \cdot (X_S(a)^{-1})$.

In step 2, the algorithm needs to compute $X_S(a)^{-1}$, which is easy given S and a.

Theorem 5. *The Decoding Algorithm returns the coefficient λ_S.*

Proof. We have

$$\begin{aligned}
\sum_{i=0}^{\ell} c_i P_i(\gamma^i \odot_S a) &= \sum_{i=0}^{\ell} c_i \sum_{T \in \mathcal{F}} \lambda_T X_T(\gamma^i \odot_S a) \ = \sum_{T\in\mathcal{F}} \lambda_T \sum_{i=0}^{\ell} c_i X_T(\gamma^i \odot_S a) \\
&= \sum_{T \in \mathcal{F}; T \neq S} \lambda_T \sum_{i=0}^{\ell} c_i \mu^i X_T(a) + \lambda_S \sum_{i=0}^{\ell-1} c_i X_S(a) \qquad (1) \\
&= \sum_{T \in \mathcal{F}; T \neq S} \lambda_T X_T(a) \sum_{i=0}^{\ell} c_i \mu^i + \lambda_S X_S(a) \sum_{i=0}^{\ell} c_i \\
&= \lambda_S X_S(a) \qquad (2)
\end{aligned}$$

where Equation 1 uses Lemma 3, and Equation 2 uses Claim 4. We note that $\mu = \gamma^{|S \cap T|}$ in Equation 1 depends on the monomial T, but we suppress this for notational clarity. □

With Grolmusz's construction, the code $\mathcal{C}_{\mathcal{F}}$ gives encoding length $(q-1)^n$, dimension $f = n^{\omega}(1)$ and query complexity 2^t. Put differently, messages of length k are encoded by codewords of length $\exp(\exp(O((\log k)^{\frac{1}{t}}(\log\log k)^{1-\frac{1}{t}})))$, which can be decoded using 2^t queries.

Summary. A better construction of set-systems with restricted intersections will give LDCs with better parameters. The set-system construction due to Grolmusz in turn uses low-degree polynomials representing the OR function on $\{0,1\}^n$ modulo composites, which were discovered by Barrington *et al.* [BBR94]. These polynomials have now found diverse combinatorial applications; LDCs, set-systems and Ramsey graphs to name a few, yet there is an exponential gap in the known degree bounds for these polynomials [Gop06]. There is also no strong evidence for what the right bound should be. We pose closing this gap as a natural open question.

Acknowledgments. I thank Venkatesan Guruswami, Prasad Raghavendra, Sergey Yekhanin and Klim Efremenko for useful discussions, and Sergey again for encouraging me to write this note.

References

[BBR94] David A. Barrington, Richard Beigel, and Steven Rudich. Representing Boolean functions as polynomials modulo composite numbers. *Computational Complexity*, 4:367–382, 1994. 3

[Efr09] Klim Efremenko. 3-query locally decodable codes of subexponential length. In *Proceedings of the* 41^{st} *Annual ACM Symposium on Theory of Computing (STOC'09)*, pages 39–44, 2009. 1, 2

[Gop06] Parikshit Gopalan. *Computing with Polynomials over Composites*. PhD thesis, Georgia Institute of Technology, 2006. 1, 3

[Gro00] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs. *Combinatorica*, 20(1):71–86, 2000. 1

[Rag07] Prasad Raghavendra. A note on Yekhanin's locally decodable codes. *Electronic Colloqium on Computational Complexity (ECCC)*, TR07-016, 2007. 1

[STV01] Madhu Sudan, Luca Trevisan, and Salil P. Vadhan. Pseudorandom generators without the XOR lemma. *J. Comput. Syst. Sci.*, 62(2):236–266, 2001. 1

[Yek07] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of ACM*, pages 1–16, 2007. 1

ECCC ISSN 1433-8092

http://eccc.hpi-web.de