Locally decodable codes

These notes are based on the note due to Gopalan [Gop09] and the lecture notes of Sudan [Sud12, Lectures 23-24].

Definition 0.1. *Let $\ell \in \mathbb{Z}^{\geq 0}$ and $\delta \in (0,1)$. A code $\mathcal{C} : \Sigma^k \to \Sigma^n$ is said to be (ℓ, δ)-locally decodable if there exists a (probabilistic) decoder D such that on oracle access to any $y \in \Sigma^n$ that satisfies $\Delta(y, \mathcal{C}(m)) \leq \delta n$, we have*

- $\forall i \in [k], \Pr[D^y(i) = m_i] \geq \frac{2}{3}$.
- *D makes at most ℓ probes into y on any input i and internal random coins.*

In these notes, we will discuss Efremenko's construction [Efr12] of sub-exponential locally decodable codes using matching vector families.

1 Matching vector codes

Let $\mathbb{F}_q$ be a finite field ($q > 2$) and $\gamma \in \mathbb{F}_q^*$ an element of order m in $\mathbb{F}_q^*$ (hence, $m|(q-1)$). We will consider both the field $\mathbb{F}_q$ and the ring $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ below.

Definition 1.1. *Let $n \in \mathbb{Z}^{\geq 0}$. For any set $L \subseteq \mathbb{Z}_m \setminus \{0\}$, $(\mathcal{U}, \mathcal{V})$, a pair of f-long vector sequences in $\mathbb{Z}_m^n$ (i.e., $\mathcal{U} = (u[1], \ldots, u[f])$ and $\mathcal{V} = (v[1], \ldots, v[f])$ where each $u[i], v[i] \in \mathbb{Z}_m^n$) is said to be an L-matching vector family if the following conditions are true.*

- *For all $i \in [f]$, $u[i] \cdot v[i] = 0$.*
- *For all $i \neq j \in [f]$, $u[i] \cdot v[j] \in L$.*

Grolmusz [Gro00] showed that for composite m, there exist vector families where f is superpolynomial in n.

Theorem 1.2. *Let m be a composite with t distinct prime factors. Then for every $n \in \mathbb{Z}^{\geq 0}$, there exists an (explicit) construction of L-matching vector family in $\mathbb{Z}_m^n$ satisfying $\ell := |L| \leq (2^t - 1)$ and $f \geq \exp\left(\frac{(\log n)^t}{(\log \log n)^{t-1}}\right)$.*

Efremenko [Efr12] showed that this (explicit) construction of matching vector family can be used to construct $(\ell+1, \frac{1}{3(\ell+1)})$-locally decodable codes which has only a sub-exponential blowup. Prior constructions required an exponential blowup.

Theorem 1.3. *Let $\mathbb{F}_q$ be a finite field ($q > 2$) and γ an element of order m in $\mathbb{F}_q^*$. Let $(\mathcal{U}, \mathcal{V})$ be a L-matching vector family of f vectors in $\mathbb{Z}_m^n$ where $L \subseteq \mathbb{Z}_m \setminus \{0\}$ and $\ell := |L| + 1$. Then there exists an $(\ell+1, \frac{1}{3(\ell+1)})$-locally decodable code $\mathcal{C}_{\mathcal{U},\mathcal{V}} : \mathbb{F}_q^f \to \mathbb{F}_q^{(q-1)^n}$.*

Observe that the blowup is $f \mapsto (q-1)^n$. Therefore, for constant q, the fact that Grolmusz's construction yields superpolynomial f implies that the blowup is at most sub-exponential.

The code $\mathcal{C}_{\mathcal{U},\mathcal{V}}$ will be a Reed-Muller-like code in the following sense. Based on the matching vector family $(\mathcal{U}, \mathcal{V})$ (in fact, just $\mathcal{V}$), we will define a set of monomials $\chi_i, i \in [f]$ as follows:

$$\chi_i(x_1, \ldots, x_n) := \prod_{k=1}^{n} x_k^{v[i]_k}.$$

Corresponding to any message $m \in \mathbb{F}_q^f$, we will construct the polynomial $P_m(x)$ as follows:

$$P_m(x_1, \ldots, x_n) := \sum_{i \in [f]} m_i \chi_i(x_1, \ldots, x_n).$$

The encoding of m will be the evaluation of the polynomial P_m on all points in $(\mathbb{F}_q^*)^n$. In other words, $\mathcal{C}_{\mathcal{U},\mathcal{V}}(m) = (P_m(x))_{x \in (\mathbb{F}_q^*)^n}$.

We will use the matching vector $u[i]$ to decode m_i, the coefficient of the monomial χ_i (which was defined using the vector $v[i]$). First for some notation. Given two $x, y \in (\mathbb{F}_q^*)^n$, define $x \odot y \in (\mathbb{F}_q^*)^n$ to be the vector obtained by component-wise product (i.e., $(x \odot y)_i = x_i y_i$). Given a vector $x \in (\mathbb{F}_q^*)^n$ and $h \in \mathbb{Z}_m$, let $x^h := (x_1^h, \ldots, x_n^h)$. Given a $a \in \mathbb{F}_q^*$ and a vector $u \in \mathbb{Z}_m^n$, let a^u be the vector in $(\mathbb{F}_q^*)^n$ defined as follows: $(a^u)_i := a^{u_i}$.

Let $B := \gamma^L := \{\gamma^c \in \mathbb{F}_q^* | c \in L\}$. Observe that $1 \notin B$ since $0 \notin L$ and $|B| = |L| = \ell$. This immediately implies the following claim

Claim 1.4. *There exists elements $c_i, i = 0, \ldots, \ell$ in $\mathbb{F}_q$ such that $\sum_{h=0}^{\ell} c_h = 1$ while for every $\beta \in B$, $\sum_{h=0}^{\ell} c_h \beta^h = 0$.*

Proof. Let c_i's be the coefficients of the polynomial $\prod_{\beta \in B} \frac{(x-\beta)}{(1-\beta)}$. □

We now define a "multiplicative line" through the point $x \in (\mathbb{F}_q^*)^n$ and direction $y \in \langle \gamma \rangle \subseteq (\mathbb{F}_q^*)^n$ as follows:

$$l_{x,y} = \{x \odot y^t \in (\mathbb{F}_q^*)^n | t \in \mathbb{Z}_m\}.$$

The following claim shows that among the monomials $\{\chi_j\}_j \in [f]$, χ_i is the only monomial that is constant along any multiplicative line in the direction of $\gamma^{u[i]}$.

Claim 1.5. *For any $i, j \in [f]$, $x \in (\mathbb{F}_q^*)^n$ and $h \in \mathbb{Z}_m$, we have*

$$\chi_j(x \odot \gamma^{hu[i]}) = \begin{cases} \chi_i(x) & \text{if } i = j, \\ \chi_j(x) \cdot \beta_{i,j}^h & \text{if } i \neq j \text{ where } \beta_{i,j} \in B. \end{cases}$$

Proof. $\chi_j(x \odot \gamma^{hu[i]}) = \prod_k \left(x_k \gamma^{hu[i]_k}\right)^{v[j]_k} = \chi_j(x) \cdot \gamma^{h(u[i] \cdot v[j])}$. □

We are now ready to define the local decoder D for the code $\mathcal{C}_{\mathcal{U},\mathcal{V}}$.

Decoder D:
Input (oracle access): $y : (\mathbb{F}_q^*)^n \to \mathbb{F}_q$ such that there exists a $m \in \Sigma^k$ such that $\Delta(y, \mathcal{C}_{\mathcal{U},\mathcal{V}}(m)) \leq \delta n$
Input (explicit): $i \in [k]$

1. Choose a random $x \in_R (\mathbb{F}_q^*)^n$ and query y at $x, x \odot \gamma^{u[i]}, x \odot \gamma^{2u[i]}, \ldots, x \odot \gamma^{\ell u[i]}$.
2. Output $\left(\sum_{h=0}^{\ell} c_h y\left(x \odot \gamma^{hu[i]}\right)\right) \cdot \chi_i(x)^{-1}$.

We will show that the above decoder proves (ℓ, δ)-local decodability of $\mathcal{C}_{\mathcal{U},\mathcal{V}}$ for $\delta \leq \frac{1}{3(\ell+1)}$.

Since x is random in $(\mathbb{F}_q^*)^n$, so is $x \odot \gamma^{hu[i]}$ for each $h \in [\ell]$ (though they are not pairwise independent). Hence, by a union bound, we can assume that at the $\ell + 1$ points queried, with probability at least $1-(\ell+1)\delta \geq 2/3$, we have that y agrees with P_m. We now have

$$\begin{aligned}
\sum_{h=0}^{\ell} c_h y\left(x \odot \gamma^{hu[i]}\right) &= \sum_{h=0}^{\ell} c_h P_m\left(x \odot \gamma^{hu[i]}\right) \\
&= \sum_{h=0}^{\ell} c_h \sum_{j\in[f]} m_j \chi_j(x \odot \gamma^{hu[i]}) \\
&= \sum_{h=0}^{\ell} c_h m_i \chi_i(x) + \sum_{h=0}^{\ell} c_h \sum_{j\in[f]\setminus\{i\}} m_j \chi_j(x) \beta_{i,j}^{h} \\
&= m_i \chi_i(x) + \sum_{j\in[f]\setminus\{i\}} m_j \chi_j(x) \sum_{h=0}^{\ell} c_h \beta_{i,j}^{h} \\
&= m_i \chi_i(x) \ .
\end{aligned}$$

This completes the proof of Theorem 1.3

2 Construction of matching vector families

Grolmusz's construction is based on the representation of OR using low-degree polynomial over $\mathbb{Z}_m$.

Definition 2.1. *Let $m \in \mathbb{Z}^{\geq 0}$. We say that $f : \{0,1\}^r \to \{0,1\}$ has a polynomial representation of degree d over $\mathbb{Z}_m$, if there exists a polynomial $p \in \mathbb{Z}_m[x_1, \ldots, x_r]$ of degree d such that for all $x \in f^{-1}(0)$, we have $p(x) = 0$ and for all $x \notin f^{-1}(1)$, we have $p(x) \neq 0$. Furthermore, we will say that f's representation has a non-zero set of size ℓ if $\ell = |\{\alpha \in \mathbb{Z}_m \setminus \{0\} | \exists x \in f^1(1), p(x) = \alpha\}|$.*

Beigel, Barrington and Rudich showed the following about OR's representation.

Theorem 2.2 (OR representation [BBR94])**.** *Let $m \in \mathbb{Z}^{\geq 0}$ have t distinct prime factors. For each $r \in \mathbb{Z}^{\geq 0}$, there exists an (explicit) representation of the OR function of degree at most $O(r^{1/t})$ and non-zero set of size at most $(2^t - 1)$ over $\mathbb{Z}_m$.*

Proof of Theorem 1.2. Let $R \in \mathbb{Z}_m[x_1, \ldots, x_r]$ be the degree d multilinear polynomial representing OR with non-zero set L of size at most $(2^t - 1)$. For each $y \in \{0,1\}^r$, construct the polynomial R_y as follows: $R_y(x_1, \ldots, x_r) = R(x_1^{y_1}, \ldots, x_r^{y_r})$ where $x_i^{y_i} = x_i$ if $y_i = 0$ and $1 - x_i$ if $y_i = 1$. Just as R represents the function $x = \bar{0}$?, R_y represents the function $x = y$?. Let $R_y(x) = \sum_m R_y^m \cdot m(x)$ be the monomial expansion of the polynomial R_y.

Define n and f as follows:

$$n = \sum_{i \leq d} \binom{r}{i}, \qquad\qquad f = 2^r$$

Observe that n denotes the number of (multilinear) monomials of degree at most d while f denotes the number of inputs in $\{0,1\}^r$. Hence, we can view each vector $v \in \mathbb{Z}_m^n$ as indexed by the monomials of degree at most d. We will now define a L-matching vector family $(\mathcal{U}, \mathcal{V})$ of length f in $\mathbb{Z}_m^n$. $\mathcal{U}$ and $\mathcal{V}$ will contain f vectors each (indexed by the elements of $\{0,1\}^r$).

- $\mathcal{U} = (u[x])_{x\in\{0,1\}^r}$ where $u[x]_m := m(x)$ (i.e, the evaluation of the monomial m at the point x).
- $\mathcal{V} = (v[x])_{x\in\{0,1\}^r}$ where $v[x]_m := R_y^m$ (i.e., the coefficient of the monomial m in the polynomial R_y).

We now observe that

$$u[x] \cdot v[y] = \sum_m u[x]_m \cdot v[x]_m = \sum_m m(x) R_y^m = R_y(x)$$
$$= \begin{cases} 0 & \text{if } x = y, \\ \in L, & \text{if } x \neq y. \end{cases}$$

This completes the proof of Theorem 1.2 assuming Theorem 2.2. $\square$

3 OR represention

In this section, we prove the BBR construction for the case when $m = 6$ and has two distinct primes 2 and 3. We need to construct a polynomial $R \in \mathbb{Z}_6[x_1, \ldots, x_r]$ of degree $O(\sqrt{r})$ that represents the OR function over r bits. More precisely, we need to constuct a polynomial $R \in \mathbb{Z}_6[x_1, \ldots, x_r]$ of degree $O(\sqrt{r})$ such that $R(x) = 0$ if the Hamming weight of x is 0 and non-zero otherwise. To this end, we will construct two polynomials $R_2 \in \mathbb{Z}_2[x_1, \ldots, x_r]$ and $R_3 \in \mathbb{Z}_3[x_1, \ldots, x_r]$ both of degree $O(\sqrt{r})$ such that if x has Hamming weight 0, then both $R_2(x) = 0$ and $R_3(x) = 0$ and if the Hamming weight is non-zero, at most one of $R_2(x)$ and $R_3(x)$ vanishes (and the other or both as the case may be are 1 (mod 2 and mod 3 respectively).

Let us construct R_2. Choose the smallest power of 2, larger than $\sqrt{r}$. In otherwords, let $a \in \mathbb{Z}^{\geq 0}$ such that $\sqrt{r} \leq 2^a < 2\sqrt{r}$. Consider the univariate polynomial $h(z) = 1 - \binom{z-1}{2^a}$ with rational coefficients. Clearly, $h(0) = 0$ while $h(1) = h(2) = \cdots = h(2^a - 1) = 1$. We make the following observations about h.

- h takes on only integral values at integral inputs. In fact, $h \pmod 2$ has a period of 2^a. In other words, $h(z) = 0 \pmod 2$ if $z = 0 \pmod{2^a}$ and $h(z) = 1 \pmod 2$ otherwise.
- $\binom{z-1}{2^a}$ can be written as an integral linear combination of $\left\{\binom{z}{i}\right\}_{i=0}^{2^a}$. I.e., there exists integers c_i such that $\binom{(z-1)}{2^a} = \sum_{i=0}^{2^a} c_i \binom{z}{i}$. This can be proved by induction using the identity $\binom{n-1}{r} = \binom{n}{r} - \binom{n-1}{r-1}$.

To move from univariate to multivariate polynomials, we observe that the symmetric polynomials $S_i(x_1, \ldots, x_r) := \sum_{S:|S|=i} \prod_{j \in S} x_j$ satisfy the property that $S_i(x_1, \ldots, x_r) = \binom{|x|}{i}$ where $|x|$ denotes the Hamming weight of x. Putting all these together, we get that the polynomial $R_2(x) = 1 - \sum c_i S_i(x)$ has the required properties.

R_3 is constructed similarly using 3^b such that $\sqrt{r} \leq 3^b < 3\sqrt{r}$. We can now combine R_2 and R_3 to obtain a single polynomial R using the Chinese remainder theorem.

References

[BBR94] DAVID A. MIX BARRINGTON, RICHARD BEIGEL, and STEVEN RUDICH. *Representing Boolean functions as polynomials modulo composite numbers*. Comput. Complexity, 4:367–382, 1994. (Preliminary version in *24th STOC*, 1992). doi:10.1007/BF01263424.

[Efr12] KLIM EFREMENKO. *3-query locally decodable codes of subexponential length*. SIAM J. Comput., 41(6):1694–1703, 2012. (Preliminary version in *41st STOC*, 2009). eccc:TR08-069, doi:10.1137/090772721.

[Gop09] PARIKSHIT GOPALAN. *A note on Efremenko's locally decodable codes*. Technical Report TR09-069, Elect. Colloq. on Comput. Complexity (ECCC), 2009. eccc:TR09-069.

[Gro00] VINCE GROLMUSZ. *Superpolynomial size set-systems with restricted intersections mod 6 and explicit Ramsey graphs*. Combinatorica, 20(1):71–86, 2000. doi:10.1007/s004930070032.

[Sud12] MADHU SUDAN. *6.S897: Algebra and Computation*, 2012. (A course on algebraic methods in computation at MIT, Spring 2012).