1 Introduction

This write-up will summarize the methodology and results of Ghasemi, Kopparty and Sudan's 2024 construction of imporved PIR schemes and the results leading up to that, starting with the complexity OR polynomials representations over n elements.

Strong representation of n elements requires a degree n polynomial which cannot be reduced. However, weak representations share a lowest degree with one-sided representations and they can be reduced. This lower degree polynomial serves as the basis for Grolmusz's construction of his uniformed super-polynomial set families over $\mathbb{F}_m$ which follow the following rules provided that m is a composite, non-prime power:

1. $\forall H \in \mathcal{H} : |H| \equiv 0 \pmod{m}$
2. $\forall G, H \in \mathcal{H}, G \neq H : |G \cap H| \not\equiv 0 \pmod{m}$

This construction is especially surprising because over $\mathbb{F}_p$ where p is prime, the set families cannot scale super-polynomialy. The Oddtown example given by László Babai and Péter Frankl is an example of this over $\mathbb{F}_2$.

Grolmusz' super-polynomial sets is then used as the basis of Efremenko's 3-query locally decodable codes (2008).

At the time of this write-up, I have yet to fully understand the role of Efremenko's 3-query locally decodable codes in Ghasemi, Kopparty and Sudan's construction of imporved PIR schemes. I aim to have this document appended before the fall with completely understanding of their paper.

2 OR Polynomials Representations

We can represent the OR-Function as a polynomial of some degree. Tardos and Barrington give 3 candidate definitions of this polynomial:

1. Strong representation: $f(x) = P(x) \in \{0, 1\}$.
2. One-sided representation: f(x) = 0 iff P(x) = 0. If f(x) = 1, P(x) = any non-zero value in Z_m.
3. Weak representation: $\forall$ x and y, P(x) $\neq$ P(y) whenever f(x) $\neq$ f(y). Equivalently, there is a subset S $\subset Z_m$ s.t f(x) = 0 iff P(x) $\in$ S.

Remark 2.1. From Barrington, Beigel, and Rudih (1994), we get the result that over Z_m, the minimum degree of a one-sided representation is $\mathcal{O}\left(n^{1/r}\right)$, where r is the number of unique prime factors of m.

3 Polynomial Sets over Z_q

Given a uniform set system, $\mathcal{H}$, obeying the following rules, how large can the system grow?

1. $\forall H \in \mathcal{H},\ |H| \equiv 0 \pmod{q}$
2. $\forall H, G \in \mathcal{H},\ \text{s.t } H \neq G,\ |H \cap G| \not\equiv 0 \pmod{q}$

The following "Eventown/Oddtown" example is from László Babai and Péter Frankl show that this set is very limited in size when q is prime, more specifically when q = 2.

Eventown

The n residents of Eventown want to form clubs and it is their life's purpose to form as many as possible under the laws of their town which is as follows:

1. There must be an even number of people in each club
2. There must be an even amount of overlap between any two clubs
3. All clubs are require to be distinct

One way to form these clubs is by marrying the residents of Eventown and having an additional couple requirement for membership (both spouses have to be part of any club). This forces the membership of any club and the overlap between any two clubs to be an even number.

Representing each club as an indicator vector: $v_i \in \{0,1\}^{\frac{n}{2}}$, we see that there exist $2^{\frac{n}{2}}$ possible clubs, which is huge.

Oddtown

It turns out maintaining these clubs is costing the town a fortune and therefore lawmakers decide to limit this number by changing the laws to the following (they also change the name of the town to Oddtown after public criticism):

1. There must be an odd number of people in each club
2. There must be an even amount of overlap between any two clubs

Theorem 3.1. For n residents of Oddtown, there can exist at most n clubs following the revised laws.

Fact 3.1. $rank(AB) \leq min\{rank(A),\ rank(B)\}$

Proof. WLOG let $n = 32$. Let each club be represented by a indicator vector $v_i \in \{0,1\}^{32}$. We define the inner product as follows:

$$v_i \cdot v_j = v_{i_1} \cdot v_{j_1} + \cdots + v_{i_{32}} \cdot v_{j_{32}}$$

Evidently we see that $v_H \cdot v_G$ is equivalent to $|H \cap G|$ where H and G represent any two clubs.

$$v_i \cdot v_j = \begin{cases} Odd & \text{if } i = j \\ Even & \text{if } i \neq j \end{cases}$$

We can simplify this by performing our calculations over $\mathbb{F}_2$.

$$v_i \cdot v_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

Let m be the number of possible law-abiding clubs formed from 32 residents. There exist m unique indicator vectors for each such club. Let those vectors form the rows of matrix $M \in M_{m\times 32}(\mathbb{R})$.

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix} \cdot \begin{bmatrix} v_1 & v_2 \dots & v_m \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Using Fact 3.1, we see that:

$$32 \geq rank(M) = rank(M^\top) \geq rank(M \cdot M^\top) = m$$

m is therefore limited to 32. □ □

4 Grolmusz's Supersized Polynomial Sets over Z_m

Theorem 4.1. Let m be a positive integer, and suppose that m has $r > 1$ different prime divisors: $m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r}$. Then there exists $c = c(m) > 0$, such that for every integer $h > 0$, there exists an explicitly constructible uniform set-system $\mathcal{H}$ over a universe of h elements, such that

1. $|\mathcal{H}| \geq \exp(c\frac{(\log h)^r}{(\log\log h)^{r-1}})$
2. $\forall H \in \mathcal{H} : |H| \equiv 0 \pmod m$
3. $\forall G, H \in \mathcal{H}, G \neq H : |G \cap H| \not\equiv 0 \pmod m$

Remark 4.1. The value of c is roughly p_r^{-r} where p_r is the largest prime factor.

Lemma 4.2. For every integer $n > 0$, there exists a uniform set-system $\mathcal{H}$ over a universe of $2(m-1)n^{2d}/d!$ elements which is explicitly constructible from the polynomial Q and satisfies

1. $|\mathcal{H}| = n^n$
2. $\forall H \in \mathcal{H} : \ |H| \equiv 0 \pmod{m}$
3. $\forall G, H \in \mathcal{H}, G \neq H : |G \cap H| \not\equiv 0 \pmod{m}$

Remark 4.2. Lemma 4.2 easily yields Theorem 4.1 by setting $d = \Theta(n^{\frac{1}{r}})$ and using elementary estimations for the binomial coefficients.

Proof of Lemma 4.2. Let $P(\vec{z})$ be a d-degree representation of OR_n s.t $P(\vec{0}) \equiv 0 \pmod{m}$.

$$Q(\vec{z}) = P(1 - z_1, 1 - z_2, \ldots, 1 - z_n)$$

We can represent $Q(\vec{z})$ as the sum of an arbitrary number of monomials of some degree less than or equal to d.

$$Q(\vec{z}) = \sum_{i_1 i_2 \ldots i_l} a_{i_1, i_2, \cdots, i_l} \cdot z_{i_1} z_{i_2} \ldots z_{i_l}$$

Since we are working over $\mathbb{F}_m$, we can separate the monomial s.t $\forall i_1 i_2 \ldots i_l, \ a_{i_1, i_2, \cdots, i_l} = 1$. Note that there does not exist more than m copies of any monomial because we are working in $\mathbb{F}_m$.

$$\tilde{Q}(\vec{z}) = \sum_{i_1 i_2 \ldots i_l} z_{i_1} z_{i_2} \ldots z_{i_l}$$

$Q(\vec{1}) \equiv \tilde{Q}(\vec{1}) \equiv 0 \pmod{m}$.

Now consider one of the monomials, say $z_1 z_2$. We can map this monomial to the set of its indices, which in this case is $\{1, 2\}$. Applying this to all of the monomials, we get a multiset of subsets of $[n]$, notated as $\mathcal{S}(\tilde{Q})$, representing the monomials that form $\tilde{Q}$.

Note that $|\mathcal{S}(\tilde{Q})|$ is equivalent to the sum of the coefficients of $\tilde{Q}$.

Now we define $\delta(u, v) = \begin{cases} 0 & u = v \\ 1 & u \neq v \end{cases}$

Using δ we build a square matrix $\mathcal{A} \in M_{n^n \times n^n}(R^n)$ of vectors $x \in [n]^n$ such that each entry is denoted by:

$$a_{xy} = (\delta(x_1, y_n), \delta(x_2, y_2), \ldots, \delta(x_n, y_n))$$

We see that $\mathcal{A}$ is composed of matrices created with respects to correlating elements of $\mathcal{S}(\tilde{\mathcal{Q}})$: $B_{i_1, i_2, \ldots, i_l}$ where $S = \{i_1, i_2, \ldots, i_l\}$ such that each entry

$$B_{xy}^{i_1,i_2,\ldots,i_l} = \delta(x_1, y_1) \times \delta(x_2, y_2) \times \ldots \times \delta(x_l, y_l)$$

We can now see that the diagonals of $\mathcal{A} = a_{xx} = 0 \pmod m$ due to all l-sized blocks of B having a value of 1, giving us the full dimension.

Now we can construct a hypergraph $\mathcal{G}$. Let the vertices be the vectors $x \in [n]^n$. Let the edges be the B-blocks. The maximum number of edges is represented by the equation (assuming $n \geq 2d$)

$$Q(n, n, \ldots, n) = \sum_{l<\leq d} \sum a_{i_1,i_2,\ldots,i_n} n^l < (m-1) \sum_{l \leq d} \binom{n}{l} n^l < \frac{2(m-1)n^{2d}}{d!}$$

This equations represents both the number of variations, $\binom{n}{l}$, and the unique variations of each, n^l.

Now we consider the dual graph of $\mathcal{G}$ which we will denote as $\mathcal{H}$. Let the universe be the set of B-blocks and the members of the set be the vectors $x \in [n]^n$.

$$|\mathcal{H}| = n^n$$

$\forall H \in \mathcal{H}, |H| = D$, the number of multisets, making this a uniform system with $Q(1, 1, \ldots, 1)$ elements.

Overlap between any two vertices: x and y, can we represented by the number of B-Blocks the two have in common by construction of the original matrix $\mathcal{A}$:

$$\forall H_x \neq H_y \in \mathcal{H},\ |H_x \cap H_y| = a_{xy} \neq 0 \pmod m$$

The intersection of the same vector $x = y$ can be represented the same way

$$\forall H_x \in \mathcal{H},\ |H_x| = |H_x \cap H_x| = a_{xx} = 0 \pmod m$$

□

Proof of Remark 4.2. Assume we want a universe of $h = \frac{n^{2\sqrt{n}}}{(\sqrt{n})!}$ elements. WLOG let $m = p_1^{\alpha_1} p_2^{\alpha_2}$ and by Grolmusz's assertion, let $d = \Theta(n^{\frac{1}{2}})$.

By Lemma 4.2, for any $n > 0$, there exists $\mathcal{H}$ over a universe of $2(m-1)\frac{n^{2(\sqrt{n})}}{(\sqrt{n})!}$ elements. Since we only care about the asymptotes, we can ignore the constant, leaving us with a universe of $\frac{n^{2(\sqrt{n})}}{(\sqrt{n})!}$ elements.

Therefore, we have found a n that gives us the h we're looking for. We can repeat this process for any arbitrary $h > 0$ and find such n. Now we can build a set-system using that n, and by the properties of Lemma 4.2, that set-system already satisfies (2) and (3) of Theorem 4.1.

The only thing left is to prove (1):

$$|H| \geq exp(c\frac{(\log h)^r}{(\log\log h)^{r-1}}$$

By Lemma 4.2, $|\mathcal{H}| = n^n$, and using Stirling's approximation: $\ln(n!) = n\ln(n) - n + O(\ln n)$, we can verify that this set system holds under (1) at the asymptotes:

$$\begin{aligned}\log h &= 2\sqrt{n}\log n - \log(\sqrt{n}!) \\ &= 2\sqrt{n}\log n - \left(\frac{1}{2}\sqrt{n}\log n + \sqrt{n}\right) \\ &= 1.5\sqrt{n}\log n\end{aligned}$$

$$\log\log h = \log(1.5) + \frac{1}{2}\log\ n + \log\log n$$

Our claim: $|\mathcal{H}| = n^n \geq \exp(c\frac{(\log h)^2}{(\log\log h}$). By applying log to both sides we get:

$$\begin{aligned}n\log n &\geq c\frac{(\log h)^2}{\log\log h} \\ &= c\frac{(1.5\sqrt{n}\ \log n)^2}{\log(1.5) + \frac{1}{2}\log n + \log\log n} \\ &\geq \frac{c * 1.5^2 * n\log^2 n}{\frac{1}{2}\log\ n + \log n} \\ &= c * n\log n\end{aligned}$$

which holds due to $c < 1$ always. □

Corollary 4.3. Results from Theorem 4.1 remains valid if we add the following condition:

(*d*) $\forall G, H \in \mathcal{H},\ G \neq H$ and $\forall i \in \{1, 2, \ldots, r\}$, we have $|G \cap H| \equiv 0 \pmod{p_i^{\alpha_i}}$ or $|G \cap H| \equiv 1 \pmod{p_i^{\alpha_i}}$.

5 3-Query Locally Decodable Codes

Locally decodable codes (LDCs) are codes that allow us to retrieve any part of the original message by reading only a constant number of symbols from the code word.

Definition 5.1. A code C is said to be locally decodable with parameters (q, δ, ϵ) if it is possible to recover any bit, x_i, of message x by making at most q queries to $C(x)$. Such that if up to a δ fraction of $C(x)$ is corrupted, then the decoding algorithm will return the correct answer with probability at least $1 - \epsilon$.

Definition 5.2. A code C is said to have a perfectly smooth decoder if $d_i(C(\vec{x})) = x_i,\ \forall \vec{x}$ and each query of d_i is uniformly distributed over $[N]$.

Fact 5.1. *Any code with a perfectly smooth decoder which makes q queries is also $(q, \delta, q\delta)$ locally decodable.*

5.1 Matching sets of vectors

Definition 5.3. The family of vectors $\{u_i\}_{i=1}^n \in (\mathbb{Z}_m)^h$ is said to be $S - matching$ if the following conditions hold over Z_m with $m = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r},$ s.t $r > 1$:

1. $\langle u_i, v_i \rangle = 0,\ \forall i \in [n]$
2. $\langle u_i, v_j \rangle = 1,\ \forall i \neq j \in [n]$

Let $S(m) = \{u \in (Z_m)^h \mid u\ \%\ p_i = \{0, 1\},\ \forall i \leq r\}$. By the Chinese Remainder Theorem and Corollary 4.3, $\exists\ 2^r - 1$ such non-zero integers $\in (Z_m)^h$. Therefore, $|S(m)| = 2^r - 1$

Definition 5.4. A polynomial $P \in \mathbb{F}[x]$ is an S-decoding polynomial if

1. $\forall s \in S, P(\gamma^s) = 0$
2. $P(\gamma^0) = P(1) = 1$

Remark 5.1. *For any S such that $0 \notin S, \exists$ an S-decoding polynomial with at most $|S| + 1$ monomials.*

Proof. Let $\tilde{P} = \prod_{s \in S}(x - \gamma^s)$. Then we can define $P(x) = \frac{\tilde{P}(x)}{\tilde{P}(1)}$ to be a S decoding polynomial. The degree of P is $|S|,$ therefore P has at most $|S| + 1$ monomials. □

5.2 The code and decoding algorithms

Fix some set S and construct an $S - matching$ family of vectors: $\{u_i\}_{i=1}^n \in (\mathbb{Z}_m)^h$, and a $S - decoding\ polynomial$: P. We define a code $C : \mathbb{F}^n \to \mathbb{F}^{m^h}$ which produce a codeword $\omega : \mathbb{F}^{m^h} \to \mathbb{F}$.

Let $e_i \in \mathbb{F}^n$ be the i-th unit vector. We define $C(a)$ by defining $C(a_i \cdot e_i) = a_i \cdot C(e_i)\ \forall i$:

$$C(e_i) \triangleq (\gamma^{<u_i, x>})_{x \in (\mathbb{Z}_m)^h}$$

We have the following equation: $C(\vec{a}) = C(a_1 \cdot e_1 + a_2 \cdot e_2 + \cdots + a_n \cdot e_n)$ and therefore:

$$C(\vec{a}) = \sum_{i=1}^{n} a_i \cdot C(e_i) = \omega_a$$

5.3 Decoder

Since P is an S-decoding polynomial and $\{u_i\}$ are S-matching vectors, $\langle u_i, u_j\rangle \in S$ for $i \neq j$, and therefore it follows that $P(\gamma^{\langle u_i,u_i\rangle}) = 1$ and $P(\gamma^{\langle u_i,u_j\rangle}) = 0$ for all $i, j \in [n],\ i \neq j$.

$$P(x) = a_0x^0 + a_0x^{b_1} + a_0x^{b_2} + \cdots + a_0x^{b_{q-1}}$$

Let C be a code word with less than δ fraction damaged coordinates and d_i be a decoder s.t $d_i(C) \Rightarrow x_i$.

The coding algorithm is as follows:

1. Start at a random vector $\vec{v}\ x \in (\mathbb{Z}_m)^h$
2. $\omega_x(v),\ \omega_x(v + b_{q-1}u_i),\ \omega_x(v + b_1u_i)$
3. Output:

$$c_i = \gamma - \langle u_i, v\rangle \left(a_0w(v) + a_1w(v + b_1u_i) + \cdots + a_{q-1}w(v + b_{q-1}u_i)\right)$$

Lemma 5.1. *The decoding algorithm d_i is a Perfectly Smooth Decoder.*

Proof. The algorithm d_i chooses v uniformly at random. Each of the queries $\omega_x(v),\ \omega_x(v + b_{q-1}u_i),\ \omega_x(v + b_1u_i)$ is uniformly distributed so therefore d_i is a Perfectly Smooth Decoder if $d_i(C(e_i)) = 1$ and $d_i(C(e_j)) = 0,\ \forall j \neq i$.

In order to prove $d_i(C(e_i)) = 1$, we use the decoder:

$$d_i(C(e_i)) = \gamma^{-\langle u_i,v\rangle}\left(a_0\gamma^{\langle u_i,v\rangle} + a_1\gamma^{\langle u_i,v+b_1u_i\rangle} + \cdots + a_{q-1}\gamma^{\langle u_i,v+b_{q-1}u_i\rangle}\right).$$

And since $\langle u_i, v + cu_i\rangle = \langle u_i, v\rangle + c\langle u_i, u_i\rangle = \langle u_i, v\rangle$, the equation becomes:

$$d_i(C(e_i)) = \gamma^{-\langle u_i,v\rangle}\left(a_0\gamma^{\langle u_i,v\rangle} + a_1\gamma^{\langle u_i,v\rangle} + \cdots + a_{q-1}\gamma^{\langle u_i,v\rangle}\right) = a_0+a_1+\cdots+a_{q-1} = P(1) = 1.$$

Now to prove $d_i(C(e_j)) = 0,\ \forall j \neq i$, recall that $P(\gamma^{\langle u_i,u_j\rangle}) = 0$ and therefore:

$$\gamma^{\langle u_i,v\rangle}\left(a_0 + a_1\gamma^{b_1\langle u_i,u_j\rangle} + \cdots + a_{q-1}\gamma^{b_{q-1}\langle u_i,u_j\rangle}\right) = \gamma^{\langle u_i,v\rangle}P(\gamma^{\langle u_i,u_j\rangle}) = 0.$$

□

5.4 Future Steps

I aim to conclude this report with a full and in-depth explanation of GKS's improved PIR schemes.