Lecture 11: Matching Vector Families

Sivakanth Gopi

October 30, 2019

In the last lecture, we have seen that we can construct locally decodable codes from matching vector families. In this lecture, we will construct matching vector families. Let us recall the definition again.

Definition 1 (MVF). *Let $S \subset \mathbb{Z}_m \setminus \{0\}$ and let $\mathcal{F} = (\mathcal{U}, \mathcal{V})$ where $\mathcal{U} = (\mathbf{u}_1, \cdots, \mathbf{u}_k), \mathcal{V} = (\mathbf{v}_1, \cdots, \mathbf{v}_k)$ are lists of vectors $\mathbf{u}_i, \mathbf{v}_i \in \mathbb{Z}_m^d$. Then $\mathcal{F}$ is called an S-MVF over $\mathbb{Z}_m^d$ of size k (and dimension d) if $\forall$ i, j,*

$$\langle \mathbf{u}_i, \mathbf{v}_j \rangle \begin{cases} = 0 & \textit{if } i = j \\ \in S & \textit{if } i \neq j \end{cases}$$

If S is omitted, it is implied that $S = \mathbb{Z}_m \setminus \{0\}$.

We want to construct MVFs with $|S|, m, d$ small while k should be large. Typically, $|S|, m$ are some fixed constants and k, d are growing.

The following lemma shows that when m is a prime, one cannot have too many matching vectors.

Proposition 2. *If m is a prime , then any MVF over $\mathbb{Z}_m^d$ must have size $k \leq 1 + d^{m-1}$.*

Proof. Let $\mathcal{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$ and $\mathcal{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$ be the MVF. Consider the $k \times k$ matrix A given by $A_{ij} = \left\langle \mathbf{u}_i^{\otimes(m-1)}, \mathbf{v}_j^{\otimes(m-1)} \right\rangle = \left\langle \mathbf{u}_i, \mathbf{v}_j \right\rangle^{m-1}$. It is clear that $\text{rank}(A) \leq d^{m-1}$. By Femat's little theorem, A is equal to $J_k - I_k$ where J_k is the all ones matrix of size $k \times k$ and I_k is the identity matrix of size $k \times k$. Therefore $\text{rank}(A) \geq \text{rank}(I_k) - \text{rank}(J_k) = k - 1$. Combining both the bounds we get $k \leq 1 + d^{m-1}$. □

Given two vectors x, y of dimensions d_1, d_2 respectively, the tensor product $x \otimes y$ is a $d_1 d_2$-dimensional vector given by $(x \otimes y)_{ij} = x_i y_j$. $x^{\otimes \ell}$ denotes $x \otimes x \otimes \cdots \otimes x$ tensored ℓ times which will have dimension d_1^ℓ. Also note that $\langle x_1 \otimes y_1, x_2 \otimes y_2 \rangle = \langle x_1, x_2 \rangle \cdot \langle y_1, y_2 \rangle$.

With a little more effort, we can extend Lemma 2 to any prime power m and also improve the bound slightly.

Proposition 3. *If m is a prime power, then any MVF over $\mathbb{Z}_m^d$ must have size $k \leq 1 + \binom{d+m-2}{m-1}$.*

Thus for any constant prime power m, the size of an MVF can be only be polynomially larger than the dimension. The following construction (which works even when m is not a prime power) shows that the lower bound in Lemma 3 is nearly tight for constant m.

Proposition 4. *Let $m \geq 2$ be any positive integer and let $d \geq m$. Then there exists an MVF over $\mathbb{Z}_m^d$ of size $k = \binom{d-1}{m-1}$.*

Proof. Let $\mathbf{w}_1, \ldots, \mathbf{w}_k$ be the set of all vectors in $\{0,1\}^{d-1}$ with Hamming weight exactly $m-1$. Let $\mathbf{u}_i = \mathbf{v}_i = (1, \mathbf{w}_i)$ for $i \in [k]$. It is easy to see that this is an MVF of size $\binom{d-1}{m-1}$. □

Surprisingly, we can do much better if m is not a prime power!

Theorem 5 ([Gro99]). *Let $m = p_1 p_2 \cdots p_t$ where $p_1, p_2 \cdots, p_t$ are distinct primes with $t \geq 2$, then there exists an explicitly constructible S-MVF $\mathcal{F}$ in $\mathbb{Z}_m^d$ of size $k \geq \exp\left(\Omega\left(\frac{(\log d)^t}{(\log\log d)^{t-1}}\right)\right)$ for some set S of size $|S| = 2^t - 1$.*

The set S in Theorem 5 can be described explicity as $S = \{a \in \mathbb{Z}_m : a \mod p_i \in \{0,1\}\ \forall\, i \in [t]\} \setminus \{0\}$.

We will now prove this. The main ingredient is a polynomial which represents the OR (disjunction) function. The OR function is 1 if at least one of its inputs is 1, and 0 otherwise.

Polynomial representations of OR $\mod m$

Definition 6 (Polynomial representation of OR $\mod m$). *A polynomial $p(x_1, \ldots, x_n)$ represents $\mathrm{OR}_n \mod m$ (over $\{0,1\}$ basis) if:*

1. *$p(0,0,\ldots,0) = 0 \mod m$ and*
2. *$p(x) \neq 0 \mod m$ for all non-zero $x \in \{0,1\}^n$.*

We will now show how to get MVFs over $\mathbb{Z}_m$ from polynomial representations of OR $\mod m$.

Lemma 7. *Suppose $p(x_1, \ldots, x_n)$ is a polynomial representation of $\mathrm{OR}_n \mod m$ of degree r. Then there exists an MVF over $\mathbb{Z}_m^d$ of size $k = 2^n$ and dimension $d = \binom{n+r}{r}$.*

Proof. Let us define the matrix M whose rows and columns are indexed by $x, y \in \{0,1\}^n$ as

$$M(x,y) = p(x \oplus y) \mod m = p(x_1 \oplus y_1, x_2 \oplus y_2, \ldots, x_n \oplus y_n) \mod m.$$

Note that the matrix M has 0's on the diagonal and non-zero values everywhere else. Therefore if rank of the matrix M is d, we can write $M(x,y) = \langle \mathbf{u}_x, \mathbf{v}_y \rangle$, where $\mathcal{U} = \{\mathbf{u}_x : x \in \{0,1\}^n\}$ and $\mathcal{V} = \{\mathbf{v}_x : x \in \{0,1\}^n\}$ form an MVF over $\mathbb{Z}_m^d$ of size $k = 2^n$.

We are now left to show that the rank of M is at most $d = \binom{n+r}{r}$. Note that we can write $x_i \oplus y_i$ as a degree 2 multilinear polynomial,

$$x_i \oplus y_i = x_i + y + i - 2x_i y_i.$$

Therefore $p(x \oplus y)$ is a degree r polyominial in x variables (also a degree r polynomial in y variables). So we can write $p(x \oplus y) =$

$\sum_{\alpha:|\alpha|\le r} x^\alpha q_\alpha(y)$ for some polynomials $q_\alpha(y)$ depending only on y. If we set $\mathbf{u}_x = \langle x^\alpha\rangle_{|\alpha|\le r}$ and $\mathbf{v}_y = \langle q_\alpha(y)\rangle_{|\alpha|\le r}$, we have $M(x,y) = p(x \oplus y) = \langle \mathbf{u}_x, \mathbf{v}_y\rangle$. The dimension d is the number of monomials in $x_1, \dots, x_n$ of degree at most r, which is $\binom{n+r}{r}$. □

Thus low degree polynomial representations of OR $\mod m$ give us good MVFs. Again, when m is a prime power, we cannot have very good MVFs and thus there cannot be any low degree polynomial representations of OR $\mod m$.

Proposition 8. *If m is a prime power and $p(x_1, \dots, x_n)$ is a polynomial representation of* $\mathrm{OR}_n \mod m$*, then* $\deg(p) \ge n/(m-1)$.

Proof Sketch. We will only prove this for prime m. By Fermat's little theorem $q(x) = p(x)^{m-1}$ is exactly equal to the OR function i.e. $q(0,0,\dots,0) = 0 \mod m$ and $q(x) = 1$ for all non-zero $x \in \{0,1\}^n$. Every function $f : 0,1^n \to \mathbb{F}_m$ has a unique multilinear polynomial representation. Therefore $q(x) = 1 - \prod_{i=1}^n (1 - x_i)$ which has degree n. Therefore $p(x)$ has degree at least $n/(m-1)$. □

When m is not a prime power, there are surprisingly low degree polynomial representations for $\mathrm{OR}_n \mod m$. For example when $m = 6$, there is degree $O(\sqrt{n})$ polynomial representation!

Theorem 9 ([BBR94])**.** *Let $m = p_1 p_2 \cdots p_t$ be a product of t distinct primes. Then there exists a polynomial representation of* $\mathrm{OR}_n \mod m$ *of degree $O_m(n^{1/t})$.*

Combining Theorem 9 with Lemma 7, we get the MVF promised in Theorem 5. The best known lower bound on the degree of polynomial representations of $\mathrm{OR}_n \mod m$ is much weaker. Therefore it is still possible to construct much better LDCs using this approach.

Theorem 10 ([TB98])**.** *Suppose m has t distinct prime factors. Then the degree of a polynomial representating* $\mathrm{OR}_n \mod m$ *is at least* $\Omega_m((\log n)^{\frac{1}{t-1}})$.

We will prove Theorem 9 in the next lecture.

References

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

[Gro99] Vince Grolmusz. Superpolynomial size set-systems with restricted intersections mod 6 and explicit ramsey graphs. *Combinatorica*, 20:2000, 1999.

[TB98] Gabor Tardos and DA Mix Barrington. A lower bound on the mod 6 degree of the or function. *Computational Complexity*, 7(2):99–108, 1998.