

Two applications of the Pattern Matrix Method

Madhur Tulsiani

1 Introduction

We study two recent results which use the “Pattern Matrix Method” of Sherstov to prove lower bounds on quantum communication complexity [She07] and on complexity of learning \mathbf{AC}^0 functions by a threshold function [RS08].

Pattern matrices are a special construction of matrices from a given function ϕ with the property that the singular values of the matrices correspond nicely to the Fourier coefficients of ϕ . Also, the eigenvalues “corresponding” to $\hat{\phi}(S)$ have an inverse exponential dependence on the size of S . Hence, if ϕ has mass only on the high-degree Fourier coefficients, then the corresponding pattern matrix has very small singular values.

Both the results in [She07] and [RS08] that use these matrices have a very similar flavor. In both results, the goal is to prove a certain kind of hardness result for a function, say f . Also, in both cases, there is a fairly natural matrix M associated with the f such that the problem is equivalent to proving a different “hardness” result for M . Furthermore, the problem is hereditary, in the sense that a hardness result for any submatrix of M implies a hardness result for M . The proof then proceeds along the following steps:

1. Reduce the problem to that of finding a matrix “correlated” with a submatrix of M and having small spectral norm.
2. Show that for the given choice of f , there is a function ϕ that correlates with f and has all the mass on Fourier coefficients of high degree.
3. Show that the pattern matrix of ϕ correlates well with the pattern matrix of f , which in turn is a submatrix of M (The pattern matrix of ϕ will have small spectral norm by the discussion above).

The notions of correlation between matrices (and hence also functions) are slightly different in both the proofs. However, the appropriate correlation conditions are established by LP-duality in both the cases¹. The first step is most problem-specific and relies on older results, while the third step is a consequence of a good choice for f .

We first present the construction and analysis for pattern matrices. Later, we describe both the applications individually, giving details for each of the steps.

¹In fact, in the LP formulation, the primal for one correlation condition can be seen to be closely related to the dual for another.

2 Construction and spectral analysis of Pattern Matrices

Before giving the general construction of a pattern matrix, we consider a special case. For a given function $\phi : \{0, 1\}^n \rightarrow \{0, 1\}$, let $A(\phi)$ be the $2^n \times 2^n$ matrix given by $A_{x,w} = \phi(x \oplus w)$. If we think of ϕ as the characteristic function of a set U , then $A(\phi)$ is the adjacency matrix of the Cayley graph generated by U . In this case the eigenvalues of $A(\phi)$ are precisely equal to the Fourier coefficients of ϕ . The last property is also true if ϕ is any real valued function.

In a pattern matrix, instead of taking x and w both from the same domain, we consider x coming from a larger domain (say $\{0, 1\}^N$ for $N \geq n$) and then consider a projection π of x to $\{0, 1\}^n$. The rows of the matrix are indexed by the inputs x and the columns by *both* the projection π and the other input w . Thus, $A_{x,(\pi,w)} = \phi(\pi(x) \oplus w)$. The fact that π is a many to one mapping results in the fact that eigenvalue corresponding to $\hat{\phi}(S)$ decrease exponentially with $|S|$.

Finally, the family of projections is not completely arbitrary but has a special form. Let n and N be such that $n|N$. We then divide the positions $1, \dots, N$ into n contiguous blocks of length N/n each. Let V be a set containing exactly one position from each block. For $x \in \{0, 1\}^N$, we denote by $x|_V$ the projection of x to the positions in V , and by $\mathcal{V}(N, n)$ the family of such sets V . Note that $|\mathcal{V}(N, n)| = (N/n)^n$. In Sherstov's constructions, $\pi(x) \equiv x|_V$ for some $V \in \mathcal{V}(N, n)$.

Definition 1 For $\phi : \{0, 1\}^n \rightarrow \mathbb{R}$, the (N, n, ϕ) -pattern matrix $A(N, n, \phi)$ is defined by

$$A_{x,(V,w)} = \phi(x|_V \oplus w)$$

where $x \in \{0, 1\}^N$, $w \in \{0, 1\}^n$ and $V \in \mathcal{V}(N, n)$.

By decomposing ϕ in the basis of characters over $\{0, 1\}^n$, it is easy to observe the following decomposition for $A(N, n, \phi)$

$$A(N, n, \phi) = \sum_{S \subseteq [n]} \hat{\phi}(S) \cdot A(N, n, \chi_S) \quad (1)$$

Let A_S denote matrix $A(N, n, \chi_S)$. We first show that each row (column) of A_S is orthogonal to each row (column) of A_T if $S \neq T$

Claim 2 For $S \neq T$, $A_S A_T^\top = A_S^\top A_T = 0$

PROOF: The proof is simply by orthogonality of the character functions

$$(A_S A_T^\top)_{x,y} = \sum_{(V,w)} \chi_S(x|_V \oplus w) \chi_T(y|_V \oplus w) = \sum_V \chi_S(x|_V) \chi_T(y|_V) \sum_w \chi_S(w) \chi_T(w) = 0$$

For the second part, we think of $\chi_S(x|_V)$ as $\chi_{S_V}(x)$, where $S_V \subseteq [N]$ only contains for each block included in S , the element from that block included in V .

$$(A_S^\top A_T)_{(V_1, w_1), (V_2, w_2)} = \chi_S(w_1) \chi_T(w_2) \sum_x \chi_S(x|_{V_1}) \chi_T(x|_{V_2}) = \chi_S(w_1) \chi_T(w_2) \sum_x \chi_{S_{V_1}}(x) \chi_{T_{V_2}}(x) = 0$$

□

Using the above decomposition and orthogonality properties, we can now characterize the spectrum of a pattern matrix.

Theorem 3 Let A be the (N, n, ϕ) -pattern matrix for a given function $\phi : \{0, 1\} \rightarrow \mathbb{R}$. Then the multiset of the non-zero singular values of A is given by

$$\bigcup_{S: \hat{\phi}(S) \neq 0} \left\{ \sqrt{2^{N+n} \left(\frac{N}{n}\right)^n \cdot |\phi(\hat{S})| \cdot \left(\frac{n}{N}\right)^{|S|/2}}, \text{ repeated } \left(\frac{N}{n}\right)^{|S|} \text{ times} \right\}$$

PROOF: From the decomposition, we know that $A = \sum_{S \subseteq [n]} \hat{\phi}(S) A_S$. Since every left (right) singular vector of a matrix is a linear combination of its columns (rows), we know that the left (right) singular vectors of all the matrices A_S form separate orthogonal subspaces. Hence, the multiset of non-zero singular values of A is the union of the the multisets of non-zero singular values of each matrix $|\hat{\phi}(S)| A_S$.² It only remains to estimate the singular values of A_S .

We compute the singular values A_S by computing those of $A_S^\top A_S$ and taking the square root. First note that $A_S^\top A_S = B_S \otimes C_S$, where B_S is a $2^n \times 2^n$ matrix with $(B_S)_{w_1, w_2} = \chi_S(w_1) \chi_S(w_2)$ and C_S is a $|\mathcal{V}(N, n)| \times |\mathcal{V}(N, n)|$ matrix with $(C_S)_{V_1, V_2} = \sum_x \chi_S(x|_{V_1}) \chi_S(x|_{V_2})$. Since, singular values of A_S are the pairwise products of singular values of B_S and C_S , we analyze the spectra of B_S and C_S separately.

B_S is a rank 1 matrix (the outer product of the table of χ_S written as a vector) having only one non-zero singular value equal to 2^n . For C_S , we analyze

$$(C_S)_{V_1, V_2} = \sum_{x \in \{0, 1\}^N} \chi_S(x|_{V_1}) \chi_S(x|_{V_2}) = \sum_{x \in B^N} \chi_{S_{V_1}}(x) \chi_{S_{V_2}}(x)$$

As before, $S_V \subseteq [N]$ is a set that only contains for each block included in S , the element from that block included in V . This gives $(C_S)_{V_1, V_2} = 2^N$ if $S_{V_1} = S_{V_2}$ and 0 otherwise. $S_{V_1} = S_{V_2}$ iff V_1 and V_2 differ in a block not included in S and for each V_1 there are $(N/n)^{n-|S|}$ such sets V_2 . The matrix can then be seen to be permutation-similar to the a block-diagonal matrix $2^N \cdot J \otimes I$ where J is the all-ones matrix of size $(N/n)^{n-|S|} \times (N/n)^{n-|S|}$ and I is the identity matrix of size $(N/n)^{|S|} \times (N/n)^{|S|}$. Again using pairwise multiplication, the singular values of C_S are $2^N \cdot (N/n)^{n-|S|}$ with multiplicity $(N/n)^{|S|}$. Multiplying these with the singular value of B_S and taking the square root proves the claim. \square

3 Lower bound on Quantum Communication Complexity

In the setting of communication complexity, we usually have a boolean function f with two inputs x, y . Two communicating parties, say Alice and Bob, are given one input each and the goal is compute $f(x, y)$. The cost of the protocol is the number of bits Alice and Bob need to exchange for both to know the value of $f(x, y)$. In the context of quantum communication complexity, they are allowed to exchange qubits instead of bits. The lower bounds presented here are for the even more general model where they are also allowed to have private qubits which may be entangled with the qubits they exchange.

For $x \in X$ and $y \in Y$ and $f : X \times Y \rightarrow \{0, 1\}$, we can define a natural $|X| \times |Y|$ matrix M with $M_{x,y} = (-1)^{f(x,y)}$. We denote by $Q_{1/5}^*(M)$, the cost of a protocol (with entangled qubits) which, given x and y computes the entry $M_{x,y}$ with error probability (for *each* input) at most $1/5$. It is immediate that if M' is a submatrix of M , then $Q_{1/5}^*(M') \leq Q_{1/5}^*(M)$.

²We use $|\hat{\phi}(S)|$ instead of $\hat{\phi}(S)$ since singular values are always positive by convention.

In the lower bound presented here, we will consider the setting where f is a function which takes only a *single* input in $\{0, 1\}^n$ i.e. $f : \{0, 1\}^n \rightarrow \{0, 1\}$. We will convert this to a two-input function by having $x \in \{0, 1\}^N$ and another input $V \in \mathcal{V}(N, n)$, where $\mathcal{V}(N, n)$ is as defined in the previous section. Given x and V , we then compute the function $f(x|_V)$. Sherstov [She07] uses this to prove lower bounds on other functions, but this case illustrates most of the ideas. The matrix M in our case will be given by $M_{x,V} = (-1)^{f(x|_V)}$. For given parameters N and n , we denote this matrix by $M(N, n)$.

Finally, we will prove this result by showing that the $(N, n, (-1)^f)$ pattern matrix $A(N, n, (-1)^f)$ is a submatrix of $M(2N, n)$ and then proving a lower bound on $Q_{1/5}^*(A(N, n, (-1)^f))$. To observe the submatrix property, we think of a generic entry of $M(2N, n)$ corresponding to $x \in \{0, 1\}^{2N}$ and $V \in \mathcal{V}(2N, n)$, given by $(-1)^{f(x|_V)}$. We restrict ourselves to inputs x where $\forall i \in [N], x_{2i} = 1 - x_{2i-1}$. Then, if $(-1)^f(y|_{V' \oplus w})$ is a generic entry of $A(N, n, (-1)^f)$, we consider x in which the bits in the odd positions are same as y and at the even positions are their complements. If $V'(i)$ is the position chosen in V' in the i th block, we choose $V(i) = 2V'(i) - 1 + w_i$, i.e. we complement the bit of the projected input if w_i is 1. This gives a way to identify rows and columns in $A(N, n, (-1)^f)$ with those in $M(2N, n)$. The problem then reduces to proving a lower bound on $Q_{1/5}^*(A(N, n, (-1)^f))$.

3.1 Reduction to finding a correlated matrix

In this case, the reduction to finding a correlated matrix is immediately given by a result of Razborov [Raz03]. The notion of correlation between two (equal size) matrices K and M is the natural one given by $\sum_{i,j} K_{ij}M_{ij}$ where i, j range over the rows and columns. We denote this quantity by $\langle K, M \rangle$. The following result is implicit in the proof of Razborov [Raz03]

Theorem 4 (Discrepancy Method) *Let $f : X \times Y \rightarrow \{0, 1\}$ be a given function for finite X, Y and let M be the matrix with $M_{x,y} = (-1)^{f(x,y)}$. Let K be any real matrix of size $|X| \times |Y|$ such that $\sum |K_{x,y}| = 1$. Then, for any $\epsilon > 0$*

$$4Q_\epsilon^*(M) \geq \frac{\langle K, M \rangle - 2\epsilon}{3 \|K\| \sqrt{|X||Y|}}$$

Here $\|K\|$ denotes the spectral norm of K . The problem then reduced to finding a matrix K which has good correlation with $A(N, n, (-1)^f)$ and has small singular values. We will first find a function ϕ which correlates with $(-1)^f$ and then take K to be $A(N, n, \phi)$ (suitably scaled so that $\sum |K_{x,y}| = 1$).

3.2 Finding a function correlated with f

Since we want a function which correlates with f and has no low-degree Fourier coefficients, we look at a quantity which captures the error in approximating f by a function with *only* low-degree characters. We say that the ϵ -approximate degree of f is d if d is the least integer such that there exists g with $\hat{g}(S) = 0$ for $|S| > d$, and $\forall x |f(x) - g(x)| \leq \epsilon$. Also, since d is the least such integer, the optimum of the following primal and dual linear programs is at least ϵ

From the dual solution, we get that there exist values β_x such that $\sum_x \beta_x f(x) \geq \epsilon$ and $\sum_x |\beta_x| = 1$. Using the facts that $(-1)^{f(x)} = 1 - 2f(x)$ and $\sum_x \beta_x = \sum_x \beta_x \chi_\emptyset(x) = 0$, we get that

<i>Primal</i>		<i>Dual</i>	
minimize	δ	maximize	$\sum_{x \in X} \beta_x f(x)$
subject to	$\left f(x) - \sum_{ S < d} \alpha_S \chi_S \right \leq \delta \quad \forall x \in X$	subject to	$\sum_{x \in X} \beta_x = 1$
	$\alpha_S, \delta \in \mathbb{R}$		$\sum_{x \in X} \beta_x \chi_S(x) = 0 \quad S < d$
			$\beta_x \in \mathbb{R}$

$\sum_x (-\beta_x)(-1)^{f(x)} > 2\epsilon$. Thus, if f has ϵ -approximate degree d , then $\phi(x) = -\beta_x$ gives a function such that $\sum_x |\phi(x)| = 1$, $\sum_x \phi(x)(-1)^{f(x)} > 2\epsilon$ and all Fourier coefficients of ϕ have support of size at least d .

3.3 Putting things together

Let d be the $1/3$ -approximate degree of f and let ϕ be the function as described above. We take K to be the $(N, n, 2^{-N}(n/N)^{-n}\phi)$ pattern matrix, so that $\sum_{x, (V, w)} |K_{x, (V, w)}| = 1$. Also, we have that $\sum_x \phi(x)(-1)^{f(x)} > 2/3$, which implies that $\langle K, A(N, n, (-1)^f) \rangle > 2/3$. Also, since $\sum_x |\phi(x)| = 1$, for all S , $\hat{\phi}(S) = 2^{-n} \sum_x \phi(x) \chi_S(x) \leq 2^{-n}$. Finally, we know from Theorem 3 that

$$\|K\| \leq \sqrt{2^{N+n} \left(\frac{N}{n}\right)^n} \cdot \frac{1}{2^N} \left(\frac{n}{N}\right)^n |\hat{\phi}(S)| \cdot \left(\frac{n}{N}\right)^{d/2} \leq \frac{1}{2^{(n+N)/2}} \left(\frac{n}{N}\right)^{(n+d)/2}$$

Combining the above estimates with Theorem 4, we get the bound that

$$Q_{1/5}^*(M(2N, n)) \geq Q_{1/5}^*(A(N, n, (-1)^f)) \geq \frac{1}{4} d \log \left(\frac{N}{n}\right) - 2$$

4 Lower bound on sign-rank of \mathbf{AC}^0

The sign-rank of a real matrix A is defined as the least rank of an equal size matrix B such that every non-zero entry of A has the same sign as the corresponding entry in B i.e. $A_{ij}B_{ij} > 0$ whenever $A_{ij} \neq 0$. One way to think about the sign-rank is to think of an entry of A as the output of a function $f : \{0, 1\}^n \rightarrow \{-1, 1\}$. Let \mathcal{C} be a class of such functions and let $M(\mathcal{C})$ be a matrix of size $|\mathcal{C}| \times 2^n$ with $M_{f, x} = f(x)$. The sign rank of $M(\mathcal{C})$ is r if and only if each entry of M can be written as $M_{f, x} = \text{sign}(a_1(f)\phi_1(x) + \dots + a_r(f)\phi_r(x))$. Hence, a lower bound on the sign rank of $M(\mathcal{C})$ implies a lower bound on the minimum number of functions required to learn \mathcal{C} as a threshold. This quantity is also known as the dimension complexity of \mathcal{C} .

Razborov and Sherstov [RS08] exhibit an explicit matrix each entry of which is computable by an \mathbf{AC}^0 circuit, but the sign rank of the matrix is exponentially high. Their method of proof also implies a lower bound on the dimension complexity of read-once DNF formulas. Specifically, they prove the following results

Theorem 5 *Let $f_m(x, y) = \bigwedge_{i=1}^m \bigvee_{j=1}^{m^2} (x_{ij} \wedge y_{ij})$, where true is interpreted as -1 and false as 1. Then the matrix M given by $M_{x, y} = f_m(x, y)$ has sign-rank $2^{\Omega(m)}$.*

Corollary 6 *Let \mathcal{C} be the class of all read-once DNF formulas $f : \{0, 1\}^n \rightarrow \{-1, 1\}$. Then the dimension complexity of \mathcal{C} is $2^{\Omega(n^{1/3})}$.*

From the definition of sign-rank, it is immediate that the sign-rank of a submatrix of M is at most the sign rank of M . Hence, it suffices to prove a lower bound on the sign-rank of a submatrix. The submatrix we consider is the pattern matrix of the function $\text{MP}_m : \{0, 1\}^{4m^3} \rightarrow \{-1, 1\}$ defined by Minsky and Papert [MP88] as

$$\text{MP}_m(z) = \bigwedge_{i=1}^m \bigvee_{j=1}^{4m^2} z_{ij}$$

We consider the (N, n, MP_m) -pattern matrix $A(N, n, \text{MP}_m)$ with $n = 4m^3$ and $N = 10^6 n$. We first argue that it is a submatrix of the matrix M corresponding to $f_{cm}(x, y)$ for a large enough constant c . We again focus on inputs x (of size $c^3 m^3$) with half the bits in odd positions being the complements of the ones in even positions, as in the previous section. It is easy to establish a correspondence between $\text{MP}_m(z|_V \oplus w)$ and $f_m(x, y)$ by choosing the value of y to select out only the positions included in V , or their complements depending on the bits of w . The problem again reduces to establishing a lower bound on the sign-rank of $A(N, n, \text{MP}_m)$.

4.1 Reduction to finding a correlated matrix

For two matrices M and P of equal dimensions, we denote by $M \circ P$ the matrix with $(M \circ P)_{ij} = M_{ij}P_{ij}$. We start with the simple observation that if P is a matrix with non-negative entries, then $\text{sign-rank}(M) \geq \text{sign-rank}(M \circ P)$. Coupled with the following generalization of a result by Forster [For02], this gives our notion of correlation

Theorem 7 *Let A be an $n_1 \times n_2$ real matrix with $n_1 n_2 = s$. If all but h entries of A satisfy $|A_{ij}| \geq \gamma$, then*

$$\text{sign-rank}(A) \geq \frac{\gamma s}{\|A\| \sqrt{s} + \gamma h}$$

Hence, if M is a matrix of size $n_1 \times n_2 = s$ with entries ± 1 and P is a non-negative matrix with all but h entries at least γ , then

$$\text{sign-rank}(M) \geq \text{sign-rank}(M \circ P) \geq \frac{\gamma s}{\|M \circ P\| \sqrt{s} + \gamma h}$$

The problem then reduces to finding a non-negative matrix P such that most entries of P are greater than some γ and $M \circ P$ has small spectral norm. Here, we think of $M \circ P$ as the ‘‘correlated’’ matrix which is obtained by scaling M according to P , such that most entries are scaled by at least γ . We construct P by finding a suitable probability distribution μ such that the function $g(x) = \mu(x)\text{MP}_m(x)$ has no low-degree Fourier coefficients, and then taking P to be the (N, n, μ) pattern matrix (or equivalently, $M \circ P$ to be the $(N, n, \mu \cdot \text{MP}_m)$ pattern matrix).

4.2 Finding a correlated function

Razborov and Sherstov prove the following result which gives the required correlated function

Theorem 8 *There is a distribution μ on $\{0, 1\}^{4m^3}$ such that $\mu(x) \geq \frac{1}{2} 8^{-m} 2^{-4m^3}$ for all $x \in \{0, 1\}^{4m^3}$ with $\text{MP}_m(x) = -1$, and $\mathbb{E}_{x \sim \mu} [\text{MP}_m(x) \chi_S(x)] = 0$ for all S with $|S| < m/3$.*

The proof proceeds by showing that the following linear program³ has optimum at least $\frac{1}{2}8^{-m}$.

$$\begin{aligned}
& \text{maximize} && \epsilon \\
& \text{subject to} && \sum_{x \in \{0,1\}^{4m^3}} \mu(x) = 1 \\
& && \sum_{x \in \{0,1\}^{4m^3}} \mu(x) \text{MP}_m(x) \chi_S(x) = 0 && \text{for } |S| < m/3 \\
& && \mu(x) \geq \epsilon 2^{-4m^3} && \text{for } \text{MP}_m(x) = -1 \\
& && \mu(x), \epsilon \geq 0
\end{aligned}$$

Proving the bound on the optimum requires analyzing the dual of the linear program and some results about interpolation of univariate polynomials which we do not present here. Note that since $\text{MP}_m(x)$ is equal to -1 on all but a $2^{-\Omega(m^2)}$ fraction of the inputs, the pattern matrix for μ provides a good candidate for the required matrix P .

4.3 Putting things together

We take $P = A(N, n, \mu)$ and $M = A(N, n, \text{MP}_m)$, with $n = 4m^3$ and $N = 10^6 n$. Theorem 7 gives that

$$\text{sign-rank}(M) \geq \text{sign-rank}(M \circ P) \geq \min \left\{ \frac{8^{-m} 2^{-n} \sqrt{s}}{4 \|M \circ P\|}, 2^{\Omega(m^2)} \right\}$$

with $s = 2^{N+n} (N/n)^n$. Also, from Theorem 3 (and the fact that $\forall S |\hat{\mu}(S)| \leq 2^{-n}$), we have the bound

$$\|M \circ P\| \leq \sqrt{s} 2^{-n} (N/n)^{-m/6} = 10^{-m} 2^{-n} \sqrt{s}$$

Combining the two bounds, we get $\text{sign-rank}(A(N, n, \text{MP}_m)) \geq 2^{\Omega(m)}$, which proves Theorem 5.

5 Concluding remarks

We show that both the proofs can be cast into a single framework with the same basic steps. The key part in both cases is defining an appropriate notion of correlation, such that the correlation of functions can be translated into correlation of pattern matrices. It may be interesting to investigate other kinds of correlations that can be translated using the method of pattern matrices.

References

- [For02] Jürgen Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.
- [MP88] M. L. Minsky and S. A. Papert. *Perceptrons: Expanded Edition*. MIT Press, 1988.
- [Raz03] A A Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

³It is interesting to compare the primal formulation if this LP with the dual of the LP in case of the quantum communication lower bound, replacing $\mu(x) \text{MP}_m(x)$ by β_x .

- [RS08] A A Razborov and A A Sherstov. The sign-rank of AC^0 . Technical Report TR08-016, Electronic Colloquium on Computational Complexity, 2008.
- [She07] Alexander A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th ACM Symposium on Theory of Computing*, 2007. To appear.