THE PATTERN MATRIX METHOD

Abstract. In this paper we give a self-contained proof of Sherstov's pattern matrix method. This theorem combines two main ideas—an extension of the discrepancy method known as the generalized discrepancy method and the notion of dual polynomials—with the techniques of matrix norms, Fourier analysis, and approximate degree. All together, this results in lower bounds on the randomized communication complexity of a certain class of communication problems, known as pattern matrices. These results extend to lower bounds for quantum communication complexity with almost no addition work, though we do not prove these results here.

Sherstov's pattern matrix method is interesting not only because of the elegance of its proof, but because the class of pattern matrices appears as subproblems of several natural problems, allowing us to give lower bounds on these communication problems as well. We give a short proof of Razborov's lower bound on the communication complexity of symmetric functions using the pattern matrix method.

Contents

1. Introduction

A communication complexity problem consists of a matrix F which takes values in $\{-1, 1\}$, whose rows are indexed by X and whose columns are indexed by Y. We study the communication complexity necessary to compute the entry $F_{x,y}$ where Alice knows the value of $x \in X$ while Bob knows the value of $y \in Y$. Various models of communication complexity exist such as the deterministic model, the bounded-error randomized model, and the quantum model.

One method to prove lower bounds on the randomized communication complexity of a function involves the discrepancy of the matrix F under some probability distribution μ. Specifically,

$$R_\epsilon(F) \geq \log \frac{1 - 2\epsilon}{\operatorname{disc}_\mu(F)}.$$

Date: May 6, 2016.

One useful property of the discrepancy method is that it gives good bounds even for ϵ close to $1/2$. However, this is also a downside since it means that for some F, the discrepancy method cannot produce good lower bounds on $R_\epsilon(F)$ for any ϵ. For example, DISJ has constant-cost randomized algorithms with error $1/2 - 1/4n$. This implies that $\mathrm{disc}_\mu(\mathsf{DISJ})$ is large, so the discrepancy method will not apply for any values of ϵ.

We can modify this idea as follows. Suppose we have a matrix F that is well-correlated with a matrix H such that H has small discrepancy. We know that H is hard to compute by the discrepancy method, and this implies that F must also be hard to compute. This idea, known as the generalized discrepancy method, theoretically allows one to prove lower bounds on a larger a class of functions. The difficulty, however, lies in finding the correct matrix H that is well-correlated with F and has small discrepancy.

We can nicely solve this problem for a class of matrices known as pattern matrices. For a single-variable function $f\colon \{0,1\}^t \to \{-1,1\}$, we associate it with a matrix known as its pattern matrix. These pattern matrices have a very specific form which preserves the Fourier-analytic properties of f. Now given a function f, there is a dual polynomial ψ that is well-correlated with it. The pattern matrix is defined in such a way that this correlation implies that f and ψ have well-correlated pattern matrices. Furthermore, the pattern matrix of ψ has a small discrepancy when f has a property known as having large ϵ-approximate degree.

In this paper we provide a self-contained introduction to Sherstov's pattern matrix method, following [5]. We introduce the two main ideas of the proof—the generalized discrepancy method and dual polynomials—as well as several useful techniques—namely, matrix norms, Fourier analysis, and approximate degree. The generalized discrepancy method (and thus the pattern matrix method as well) apply not only to randomized communication complexity, but to quantum complexity as well with essentially no additional work. However, for simplicity, we do not introduce quantum communication complexity in this paper.

The importance of the pattern matrix method lies in the fact that pattern matrices appear as submatrices of many communication problems of interest. For example, an important result of Razborov gives lower bounds on the quantum communication complexity of symmetric functions [4]. This result may be easily proven using the pattern matrix method, allowing us to obtain lower bounds on the communication complexity of DISJ, among many other problems.

In Section 2 we prove the generalized discrepancy method. Then in Section 3 we introduce basic ideas of matrix analysis and use them to recast the generalized discrepancy method in a simpler form. In Section 4 we introduce the basics of Fourier analysis and use them to define the approximate degree of a function. We then define dual polynomials and prove their existence. Then in Section 5 we define pattern matrices and prove lower bounds on their communication complexity. Finally, in Section 6 we give a simple proof of Razborov's lower bounds on symmetric functions.

2. Generalized Discrepancy

In this section we first review the ideas in the standard discrepancy method without proof. We then introduce and prove the generalized discrepancy method.

Definition 2.1. For a function $f\colon X \times Y \to \{-1,1\}$ and a probability distribution μ over $X \times Y$, define the **discrepancy**, $\mathrm{disc}_\mu(f)$, to be the maximum over all $S \subseteq X$ and $T \subseteq Y$ of

$$\mathbf{E}_{(x,y)\sim\mu}[f(x,y)1_S(x)1_T(y)].$$

Here $1_S(x)$ is the indicator function that takes value 1 if $x \in S$ and 0 otherwise and similarly for $1_T(y)$.

Definition 2.2. For a function $f\colon X \times Y \to \{-1,1\}$, a probability distribution μ over $X \times Y$, and an $\epsilon > 0$, define $D^\mu_\epsilon(f)$ to be the minimum cost over any deterministic protocol P such that

$$\mathbf{E}_{(x,y)\sim\mu}[f(x,y)P(x,y)] \geq 1 - 2\epsilon.$$

In particular, this means that in expectation over all possible inputs, the protocol P computes f correctly at least $1-\epsilon$ of the time.

Proposition 2.3. *For a function $f\colon X \times Y \to \{-1,1\}$, a probability distribution μ over $X \times Y$, and some $\epsilon > 0$, it holds that*

$$R_\epsilon(f) \geq D^\mu_\epsilon(f).$$

Theorem 2.4 (Discrepancy Method)**.** *For a function $f\colon X \times Y \to \{-1,1\}$, a probability distribution μ over $X \times Y$, and any $0 < \epsilon < 1/2$, it holds that*

$$D^\mu_\epsilon(f) \geq \log \frac{1-2\epsilon}{\mathrm{disc}_\mu(f)}.$$

The discrepancy method states that if a function f has small discrepancy under some probability distribution μ, then f is hard to compute. We extend this method to a larger class of functions. In particular, for any function f that is well-correlated to some h under μ and h has small discrepancy under μ, then f is hard to compute.

Similar ideas were was used in both [4, 2]; we formalize it as in [5].

Theorem 2.5 (Generalized Discrepancy Method)**.** *For a function $f\colon X \times Y \to \{-1,1\}$, a probability distribution μ over $X \times Y$, and a function $h\colon X \times Y \to \{-1,1\}$ such that*

$$\boldsymbol{E}_{(x,y)\sim\mu}[f(x,y)h(x,y)] \geq \delta,$$

the following holds:

$$R_\epsilon(f) \geq \log \frac{\delta - 2\epsilon}{\mathrm{disc}_\mu(h)}.$$

Proof. Suppose that P is a deterministic protocol with cost $c = D^\mu_\epsilon(f)$ such that

$$\mathbf{P}_{(x,y)\sim\mu}[f(x,y) \neq P(x,y)] \leq \epsilon.$$

Applying the standard discrepancy method to h and P, we obtain

$$c \geq \log \frac{\mathbf{E}_{(x,y)\sim\mu}[h(x,y)P(x,y)]}{\mathrm{disc}_\mu(h)}.$$

Now since f and P differ with probability bounded by ϵ, we deduce that

$$R_\epsilon(f) \geq D^\mu_\epsilon(f) = c \geq \log \frac{\mathbf{E}_{(x,y)\sim\mu}[f(x,y)h(x,y)] - 2\epsilon}{\mathrm{disc}_\mu(h)} \geq \log \frac{\delta - 2\epsilon}{\mathrm{disc}_\mu(h)}.$$

□

3. Matrix Norms

In this section we introduce some basic notions from matrix analysis, and rephrase Theorem 2.5 in terms of them.

Definition 3.1. For a matrix M, define its norm to be

$$\|M\| = \sup_{x : \|x\|=1} \|Mx\|.$$

Definition 3.2. For two matrices M, N with rows labeled by X and columns by Y, define $\langle M, N\rangle$ to be

$$\sum_{(x,y)\in X\times Y} M_{x,y}N_{x,y}$$

We use the following simple proposition to bound the matrix norm.

Proposition 3.3. *For an arbitrary matrix M,*

$$\|M\|^2 \leq \|M^\top M\|.$$

Proof. Pick some x with $\|x\| = 1$ and $\|Mx\| = \|M\|$. Then

$$\|M\|^2 = \|Mx\|^2 = (Mx)^\top(Mx) = x^\top M^\top Mx \leq \|x\|\|M^\top Mx\| \leq \|M^\top M\|.$$

□

Remark 3.4. The matrix norm has a number of deep properties, connecting it to singular value decomposition and eigenvalue decomposition of matrices. In particular, the above inequality is actually an equality, though we do not use this or any other of these facts in this paper.

The reason for introducing the matrix norm is because we can replace discrepancy in the generalized discrepancy theorem with it. The following proposition demonstrates this relation.

Proposition 3.5. *For a function $f \colon X\times Y \to \{-1, 1\}$ and a probability distribution μ over $X \times Y$, define the matrix Ψ by $\Psi_{x,y} = f(x, y)\mu(x, y)$. Then the following relation holds:*

$$\mathrm{disc}_\mu(f) \leq \|\Psi\|\sqrt{|X||Y|}.$$

Proof. By the definition of discrepancy, we have sets $S \subseteq X$ and $T \subseteq Y$ such that

$$\mathbf{E}_{(x,y)\sim\mu}[f(x, y)1_S(x)1_T(y)] = \mathrm{disc}_\mu(f).$$

We can rewrite this as

$$\sum_{(x,y)\in X\times Y} 1_S(x)f(x, y)\mu(x, y)1_T(y) = \mathrm{disc}_\mu(f).$$

Equivalently, this says $1_S^\top \Psi 1_T = \mathrm{disc}_\mu(f)$. Now we can bound this quantity as

$$\mathrm{disc}_\mu(f) = 1_S^\top\Psi 1_T \leq \|1_S\|\|\Psi 1_T\| \leq \|1_S\|\|\Psi\|\|1_T\| = \|\Psi\|\sqrt{|S||T|} \leq \|\Psi\|\sqrt{|X||Y|}.$$

□

Theorem 3.6 (Generalized Discrepancy Method, Matrix Form)**.** *For a function $f \colon X \times Y \to \{-1, 1\}$, an $\epsilon > 0$, and any matrix Ψ with rows labeled by X and columns by Y such that $\sum_{(x,y)\in X\times Y} |\Psi_{x,y}| = 1$, we have the following inequality*

$$R_\epsilon(f) \geq \log \frac{\langle F, \Psi\rangle - 2\epsilon}{\|\Psi\|\sqrt{|X||Y|}}.$$

Proof. For a function $h\colon X \times Y \to \{-1, 1\}$ and a probability distribution μ on $X \times Y$, we can define a matrix Ψ by $\Psi_{x,y} = h(x, y)\mu(x, y)$. Furthermore, for any matrix Ψ such that $\sum_{(x,y)\in X\times Y} |\Psi_{x,y}| = 1$, we can reverse this procedure. Simply let $\mu(x, y) = |\Psi_{x,y}|$ and let $h(x, y)$ be the sign of $\Psi(x, y)$.

Now note that

$$\mathbf{E}_{(x,y)\sim\mu}[f(x, y)h(x, y)] = \sum_{(x,y)\in X\times Y} f(x, y)h(x, y)\mu(x, y) = \langle F, \Psi\rangle.$$

Furthermore, applying Proposition 3.5, the conclusion of Theorem 2.5 becomes

$$R_\epsilon(f) \geq \log \frac{\langle F, \Psi\rangle - 2\epsilon}{\operatorname{disc}_\mu(h)} \geq \log \frac{\langle F, \Psi\rangle - 2\epsilon}{\|\Psi\| \sqrt{|X||Y|}}.$$

□

4. Approximate Degree and Dual Polynomials

In this section we first introduce the basics of Fourier analysis on the Boolean hypercube and use this to define the approximate degree. Then we define dual polynomials and show their existence using LP duality.

Definition 4.1. For $S \subseteq [n]$, define the **character** $\chi_S\colon \{0, 1\}^n \to \{-1, 1\}$ by

$$\chi_S(x) = (-1)^{\sum_{i\in S} x_i}.$$

Proposition 4.2. *The characters obey the following two properties:*

(1) Linearity: for $x, y \in \{0, 1\}^n$,

$$\chi_S(x \oplus y) = \chi_S(x)\chi_S(y).$$

(2) Orthogonality: for $S \neq T$,

$$\sum_{x\in\{0,1\}^n} \chi_S(x)\chi_T(x) = 0.$$

Orthogonality of characters states that $\{\chi_S\}_{S\subseteq[n]}$ forms an orthogonal basis of $\mathbb{R}^{\{0,1\}^n}$, the vector space of functions $\{0, 1\}^n \to \mathbb{R}$. We define the Fourier transform of f as the coordinates of f in this basis.

Definition 4.3. For a function $f\colon \{0, 1\}^n \to \mathbb{R}$, define its **Fourier transform** to be $\hat{f}\colon 2^{[n]} \to \mathbb{R}$ defined by

$$\hat{f}(S) = 2^{-n} \sum_{x\in X} f(x)\chi_S(x).$$

Proposition 4.4. *For a function $f\colon \{0, 1\}^n \to \mathbb{R}$, its Fourier transform satisfies*

$$f(x) = \sum_{S\subseteq[n]} \hat{f}(S)\chi_S(x).$$

Now we introduce the notion of approximate degree.

Definition 4.5. For a function $f\colon \{0, 1\}^n \to \mathbb{R}$, define its **degree** to be the largest integer d such that there exists S with $\hat{f}(S) \neq 0$ and $d = |S|$.

Definition 4.6. For a function $f\colon \{0,1\}^n \to \mathbb{R}$, define its **$\epsilon$-approximate degree**, $\deg_\epsilon(f)$, to be the smallest integer d such that there exists a function $h\colon \{0,1\}^n \to \mathbb{R}$ of degree d with

$$\|f-h\|_\infty \le \epsilon.$$

Here $\|\phi\|_\infty$ is defined to be the maximum absolute value that ϕ obtains.

We can decompose $\mathbb{R}^{\{0,1\}^n}$ into two orthogonal subspaces: the subspace V of all degree $d-1$ functions, spanned by $\{\chi_S\}_{|S|<d}$ and its orthogonal complement W, spanned by $\{\chi_S\}_{|S|\ge d}$.

Now if $d = \deg_\epsilon(f)$, this means that f is not well-approximated by any element of V. Using duality, we show that this implies that we can find a function $\psi \in W$ which is well-correlated with f. First we prove a more general version of this idea, originally proved in [1] in even more generality. We follow the presentation in [5] which proves it only in the case that we use.

Theorem 4.7 ([1]). *Let X be a finite set and Φ be an arbitrary subspace of $\mathbb{R}^X$, the vector space of functions $X \to \mathbb{R}$. For any function $f \in \mathbb{R}^X$,*

$$\min_{\phi\in\Phi} \|f-\phi\|_\infty = \max_{\psi\in\Phi^\perp : \left|\sum_{x\in X}\psi(x)\right|\le 1} \sum_{x\in X} \psi(x) f(x).$$

Proof. This follows by LP duality. Pick a basis $\phi_1,\dots,\phi_k$ for Φ. We construct two linear programming problems: the first has variables $(\epsilon, c_1, \dots, c_k)$ subject to the constraints $\left|f(x) - \sum_{i=1}^k c_i\phi_i(x)\right| \le \epsilon$ for each $x \in X$ and objective to minimize ϵ. The second has variables $(\psi(x))_{x\in X}$ subject to the constraints $\left|\sum_{x\in X}\psi(x)\right| \le 1$ and $\sum_{x\in X}\psi(x)\phi_i(x) = 0$ for $1 \le i \le k$ and objective to maximize $\sum_{x\in X}\psi(x)f(x)$.

Setting $\phi = \sum_{i=1}^k c_i\phi_i$, we see that solutions to the first problem correspond exactly to pairs (ϕ,ϵ) where $\phi \in \Phi$ and $\|f-\phi\|_\infty \le \epsilon$. Similarly, solutions to the second problem correspond exactly to functions ψ such that $\psi \in \Phi^\perp$ and $\left|\sum_{x\in X}\psi(x)\right| \le 1$. It is not hard to check that these two programs are dual to each other, so we conclude that optima of these two programs are equal, which is exactly the desired result.[1] □

Corollary 4.8. *For $f\colon \{0,1\}^n \to \mathbb{R}$, let $d = \deg_\epsilon(f)$. Then there exists a **dual polynomial**, $\psi\colon \{0,1\}^n \to \mathbb{R}$, such that*

- $\hat{\psi}(S) = 0$ *for* $|S| < d$,
- $\sum_{x\in\{0,1\}^n}\psi(x)f(x) > \epsilon$,
- $\left|\sum_{x\in\{0,1\}^n}\psi(x)\right| = 1$.

Proof. Let Φ be the subspace of $\mathbb{R}^{\{0,1\}^n}$ spanned by χ_S for $|S| < d$. By orthogonality of characters, we know that $\Phi^\perp$ consists of all ψ such that $\hat{\psi}(S) = 0$ for all $|S| < d$.

Now since $d = \deg_\epsilon(f)$, we have that for any $\phi \in \Phi$, i.e., ϕ of degree less than d, we must have $\|f-\phi\|_\infty > \epsilon$. By Theorem 4.7, this implies the existence of some

[1] Note that the programs used in this proof are not exactly in the standard form of a linear program. Namely, we do not require that the variables are positive and we use absolute values in our linear constraints. However, there are standard techniques to rectify both of these problems—by replacing each equation, $|a| \le b$, by a pair of equations, $a \le b$ and $-a \le b$, and each variable, x, by a pair of variables, x_+ and x_-, corresponding to the positive and negative parts of x.

$\psi \in \Phi^{\perp}$ such that $\left|\sum_{x\in\{0,1\}^n} \psi(x)\right| \leq 1$ and

$$\sum_{x\in\{0,1\}^n} \psi(x)f(x) = \|f - \phi(x)\|_{\infty} > \epsilon.$$

This implies the desired result, since scaling ψ preserves the first two properties while increasing $\left|\sum_{x\in\{0,1\}^n} \psi(x)\right|$ to 1. □

5. Pattern Matrices

Given a function $f\colon \{0,1\}^t \to \{-1,1\}$, its pattern matrix represents the following communication problem. Alice is given a vector $x \in \{0,1\}^n$. Bob is given a subset $V \subseteq [n]$ of size t and a vector $w \in \{0,1\}^t$. They want to find the vector $x|_V$ which consists of the t components of x specified by V and then compute $f(x|_V \oplus w)$.

Definition 5.1. For a function $f\colon \{0,1\}^t \to \mathbb{R}$ and n a multiple of t, define the (n,t,f)**-pattern matrix** $A_{(n,t,f)}$ as follows. The rows are indexed by $\{0,1\}^n$ while the columns are indexed by $[n/t]^t \times \{0,1\}^t$. For $x \in \{0,1\}^n$ and $V \in [n/t]^t$, let $x|_V \in \{0,1\}^t$ be the vector whose ith component is given by $x_{(n/t)(i-1)+V_i}$. Then let the entry in position $(x,(V,w))$ of the pattern matrix be given by $f(x|_V \oplus w)$.

To apply the generalized discrepancy method to pattern matrices, we first need to compute the matrix norm of pattern matrices. The following lemma makes the calculation feasible.

Lemma 5.2. *For two $n \times m$ matrices A, B, if $AB^{\top} = 0$ and $A^{\top}B = 0$, then $\|A+B\|^2 \leq \max\{\|A^{\top}A\|, \|B^{\top}B\|\}$.*

Proof. First note that $\|A+B\|^2 \leq \|(A+B)^{\top}(A+B)\| = \|A^{\top}A + B^{\top}B\|$. The first equality follows by Proposition 3.3 while the second follows since $A^{\top}B$ and $B^{\top}A$ are both 0. Now we claim $\|A^{\top}A + B^{\top}B\| \leq \max\left\{\|A^{\top}A\|, \|B^{\top}B\|\right\}$.

To see this, let $U \subseteq \mathbb{R}^m$ be the span of the rows of A and $V \subseteq \mathbb{R}^m$ be the span of the rows of B. Since $AB^{\top} = 0$, the subspaces U, V are orthogonal. Therefore we can write $\mathbb{R}^m = U \oplus V \oplus W$ where U, V, W are all orthogonal. Now for any vector $x \in \mathbb{R}^m$, we can uniquely write $x = u + v + w$ where $u \in U$, $v \in V$, $w \in W$.

In particular, pick x with $\|x\| = 1$ and such that $\|(A^{\top}A + B^{\top}B)x\| = \|A^{\top}A + B^{\top}B\|$. Then

$$\begin{aligned}
\|A^{\top}A + B^{\top}B\|^2 &= \|(A^{\top}A + B^{\top}B)x\|^2 \\
&= \|A^{\top}Au + B^{\top}Bv\|^2 \\
&= \|A^{\top}Au\|^2 + \|B^{\top}Bv\|^2 \\
&\leq \|A^{\top}A\|^2\|u\|^2 + \|B^{\top}B\|^2\|v\|^2 \\
&\leq \max\left\{\|A^{\top}A\|, \|B^{\top}B\|\right\}^2 (\|u\|^2 + \|v\|^2) \\
&\leq \max\left\{\|A^{\top}A\|, \|B^{\top}B\|\right\}^2 \|x\|^2.
\end{aligned}$$

The second line follows since U, V, W are mutually orthogonal subspaces. The third line follows by the Pythagorean theorem since $A^{\top}B$ implies that $\operatorname{im} A$ and $\operatorname{im} B$ are orthogonal spaces. □

Proposition 5.3. *For a function $f\colon \{0,1\}^t \to \mathbb{R}$ and n a multiple of t, the matrix norm $\|A_{(n,t,f)}\|$ is bounded above by*

$$\max_{S\subseteq[n]} \sqrt{2^{n+t}(n/t)^{t-|S|}}|\hat{f}(S)|.$$

Proof. For $S \subseteq [t]$, we use A_S to refer to $A_{(n,t,\chi_S)}$. Then since $f = \sum_{S\subseteq[t]} \hat{f}(S)\chi_S$, we conclude that

$$A_{(n,t,f)} = \sum_{S\subseteq[t]} \hat{f}(S)A_S.$$

We show that $\|A_{(n,t,f)}\|^2 \leq \max_{S\subseteq[n]} \hat{f}(S)^2\|A_S^\top A_S\|$ and $\|A_S^\top A_S\| = 2^{n+t}(n/t)^t$. The former follows from Lemma 5.2 once we compute $A_S^\top A_T$ and $A_S A_T^\top$ while the latter follows from a similar calculation.

$$\begin{aligned}
[A_S A_T^\top]_{x,y} &= \sum_{\substack{V\in[n/t]^t \\ w\in\{0,1\}^t}} \chi_S(x|_V \oplus w)\chi_T(y|_V \oplus w) \\
&= \sum_{V\in[n/t]^t} \chi_S(x|_V)\chi_T(y|_V) \sum_{w\in\{0,1\}^t} \chi_S(w)\chi_T(w) \\
&= 0
\end{aligned}$$

Similarly, we see

$$\begin{aligned}
[A_S^\top A_T]_{(V,w),(V',w')} &= \sum_{x\in\{0,1\}^n} \chi_S(x|_V \oplus w)\chi_T(x|_{V'} \oplus w') \\
&= \chi_S(w)\chi_T(w') \sum_{x\in\{0,1\}^n} \chi_S(x|_V)\chi_T(x|_{V'})
\end{aligned}$$

Define $S_V \subseteq [n]$ by $S_V = \{(n/t)(i-1) + V_i : i \in S\}$. Then $\chi_S(x|_V) = \chi_{S_V}(x)$. Therefore, by orthogonality of characters, we have $[A_S^\top A_T]_{(V,w),(V',w')} = 0$ for $S \neq T$, since we have $S_V = T_{V'}$.

Furthermore, for $S = T$, we see that $[A_S^\top A_S]_{(V,w),(V',w')} = 2^n\chi_S(w)\chi_S(w')$ for $S_V = S_{V'}$ and 0 otherwise. Now note that for each V, there are $(n/t)^{t-|S|}$ choices of V' such that $S_V = S_{V'}$. This implies that $A_S^\top A_S$ is a matrix made of blocks of size $2^t(n/t)^{t-|S|}$. Each of these blocks is a rank 1 matrix where each row is a vector of length $2^n\sqrt{2^t(n/t)^{t-|S|}}$. This implies that $\|A^S\top A^S\| = 2^{n+t}(n/t)^{t-|S|}$, as desired. □

Theorem 5.4 (Pattern Matrix Method)**.** *For a function $f\colon \{0,1\}^t \to \{-1,1\}$, n a multiple of t, and $\epsilon > 2\delta$, the pattern matrix $A_{(n,t,f)}$ satisfies the following:*

$$R_\delta(A_{(n,t,f)}) \geq \frac{1}{2}\deg_\epsilon(f)\log(n/t) + \log(\epsilon - 2\delta).$$

Proof. Set $d = \deg_\epsilon(f)$ and let ψ be the dual polynomial to f. By Corollary 4.8, this satisfies the following:

i. $\hat{\psi}(S) = 0$ for $|S| < d$,
ii. $\sum_{x\in\{0,1\}^t} \psi(x)f(x) > \epsilon$,
iii. $\left|\sum_{x\in\{0,1\}^t} \psi(x)\right| = 1$.

We wish to apply the generalized discrepancy method to the pattern matrix of ψ under the uniform distribution. Namely let $\Psi = (n/t)^{-t}2^{-n}A_{(n,t,\psi)}$. Simply combining the above properties of ψ with the definition dual polynomials, we bound the desired quantities. By iii.,

$$\begin{aligned}\sum_{\substack{x\in\{0,1\}^n\\(V,w)\in[n/t]^t\times\{0,1\}^t}} |\Psi_{x,(V,w)}| &= (n/t)^{-t}2^{-n}\sum_{\substack{x\in\{0,1\}^n\\V\in[n/t]^t}}\left(\sum_{w\in\{0,1\}^t}\psi(x|_V\oplus w)\right)\\ &= (n/t)^{-t}2^{-n}\sum_{\substack{x\in\{0,1\}^n\\V\in[n/t]^t}} 1\\ &= 1.\end{aligned}$$

By ii.,

$$\begin{aligned}\langle A_{(n,t,f)},\Psi\rangle &= (n/t)^{-t}2^{-n}\sum_{\substack{x\in\{0,1\}^n\\V\in[n/t]^t}}\left(\sum_{w\in\{0,1\}^t} f(x|_V\oplus w)\psi(x|_V\oplus w)\right)\\ &> (n/t)^{-t}2^{-n}\sum_{\substack{x\in\{0,1\}^n\\V\in[n/t]^t}}\epsilon\\ &= \epsilon.\end{aligned}$$

By i. and Proposition 5.3,

$$\|\Psi\| \le (n/t)^{-t}2^{-n}\sqrt{2^{n+t}(n/t)^{t-d}}\max_{S\subseteq[n]}|\hat{\psi}(S)|.$$

Furthermore, by the triangle inequality applied to the definition of $\hat{\psi}(S)$ and iii, we see

$$|\hat{\psi}(S)| \le 2^{-t}\sum_{x\in\{0,1\}^t}|\psi(x)| = 2^{-t}.$$

Therefore we conclude that $\|\Psi\| \le \left(2^{n+t}(n/t)^{t+d}\right)^{-1/2}$, so by Theorem 3.6

$$\begin{aligned}R_\epsilon(A_{(n,t,f)}) &\ge \log\frac{\epsilon-2\delta}{\|\Psi\|\sqrt{2^n(n/t)^t2^t}}\\ &\ge \frac{1}{2}\deg_\epsilon(f)\log(n/t)+\log(\epsilon-2\delta).\end{aligned}$$

□

6. Application: Communication Complexity of Symmetric Functions

We give an alternative proof of a result of Razborov using the pattern matrix method.

Definition 6.1. A **predicate** is a function $D\colon\{0,1,\dots,n\}\to\{-1,1\}$. Its associated communication problem is the symmetric function $f_D\colon\{0,1\}^n\times\{0,1\}^n\to\{-1,1\}$ given by $f_D(x,y)=D\left(\sum_{i=1}^n x_iy_i\right)$.

Define $0\le l_0(D)\le\lfloor n/2\rfloor$ and $0\le l_1(D)\le\lceil n/2\rceil$ to be the smallest integers such that D is constant on the range $\{l_0(D),l_0(D)+1,\dots,n-l_1(D)\}$.

Razborov's lower bound on the communication complexity of symmetric functions states the following.

Theorem 6.2 ([4]). *Given a predicate $D\colon \{0,1\}^n \to \{-1,1\}$,*

$$R_{1/3}(f_D) \geq \Omega\left(\sqrt{nl_0(D)} + l_1(D)\right).$$

Remark 6.3. Razborov actually proved the above result for quantum communication complexity. The proof we give also proves this result once one extends the generalized discrepancy method (and thus the pattern matrix method as well) from randomized to quantum communication complexity.

We only prove half of this result here. The other half follows by similar techniques, though it has more details to check. First note that the approximate degree of f_D is a well-studied problem. Namely we have the following theorem.

Theorem 6.4 ([3]). *Given a predicate $D\colon \{0,1\}^n \to \{-1,1\}$,*

$$\deg_{1/3}(f_D) = \Omega\left(\sqrt{nl_0(D)} + \sqrt{nl_1(D)}\right).$$

Proposition 6.5. *Given a predicate $D\colon \{0,\ldots,n\} \to \{-1,1\}$, if there exists $l \leq n/8$ with $D(l) \neq D(l-1)$, then*

$$R_{1/3}(f_D) \geq \Omega(\sqrt{nl}).$$

Proof. It suffices to prove this result for $R_{1/7}(f_D)$. Let $f\colon \{0,1\}^{\lfloor n/4\rfloor} \to \{0,1\}$ be the function given by $f(x) = D(\sum_{i=1}^{\lfloor n/4\rfloor} x_i)$ and let F refer to the pattern matrix $A_{2\lfloor n/4\rfloor,\lfloor n/4\rfloor,f}$. Applying Theorem 5.4 with $\epsilon = 1/3$ and $\delta = 1/7$, we see by Theorem 6.4 that $D_{1/7}(F) = \Omega(\sqrt{nl})$. All that remains is to show that $D_{1/7}(f_D) \geq D_{1/7}(F)$. To prove this, we show that F is a submatrix of f_D.

A row of F corresponds is given by some $x = ((x_i, x_i'))_{i=1}^{\lfloor n/4\rfloor}$ while a column of F is given by some $y = ((y_i, w_i))_{i=1}^{\lfloor n/4\rfloor}$. Then

$$F_{x,y} = D\left(\sum_{i=1}^{\lfloor n/4\rfloor} (x_i, x_i')|_{y_i} \oplus w_i\right).$$

Here we use $(x_i, x_i')|_{y_i} = x_i$ if $y_i = 0$ and x_i' if $y_i = 1$. Below we write out the matrix of $[(x_i, x_i')|_{y_i} \oplus w_i]$.

$$\begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

To show that F is a submatrix of f_D, we define a pair of maps $((x_i, x_i'))_{i=1}^{\lfloor n/4\rfloor} \to ((\mathbf{x}_i, \mathbf{x}_i', \mathbf{x}_i'', \mathbf{x}_i'''))_{i=1}^{\lfloor n/4\rfloor}$ and $((y_i, w_i))_{i=1}^{\lfloor n/4\rfloor} \to ((\mathbf{y}_i, \mathbf{y}_i', \mathbf{y}_i'', \mathbf{y}_i'''))_{i=1}^{\lfloor n/4\rfloor}$ defined below.

(x_i, x_i')	$(\mathbf{x}_i, \mathbf{x}_i', \mathbf{x}_i'', \mathbf{x}_i''')$
00	1000
01	0100
10	0010
11	0001

(y_i, w_i)	$(\mathbf{y}_i, \mathbf{y}_i', \mathbf{y}_i'', \mathbf{y}_i''')$
00	0011
01	1100
10	0101
11	1010

It is easy to check that under this pair of maps, a pair $((x_i, x_i'), (y_i, w_i))$ is sent to a pair $(\mathbf{x}_i, \mathbf{x}_i', \mathbf{x}_i'', \mathbf{x}_i'''), (\mathbf{y}_i, \mathbf{y}_i', \mathbf{y}_i'', \mathbf{y}_i''')$ such that

$$(x_i, x_i')|_{y_i} \oplus w_i = \mathbf{x}_i\mathbf{y}_i + \mathbf{x}_i'\mathbf{y}_i' + \mathbf{x}_i''\mathbf{y}_i'' + \mathbf{x}_i'''\mathbf{y}_i'''.$$

This implies that the map from F to f_D preserves the matrix entries, as desired. □

Example 6.6. DISJ is the communication problem f_D associated to the predicate

$$D(l) = \begin{cases} 1 & l = 0, \\ 0 & \text{otherwise.} \end{cases}.$$

Then $D(1) \neq D(0)$, so applying the above proposition with $l = 1$, we conclude

$$R_{1/3}(\mathsf{DISJ}) \geq \Omega(\sqrt{n}).$$

Remark 6.7. In this example, the bounds we obtain are tight in the quantum case, but not in the randomized case. This is one failing of the pattern matrix method—it cannot prove gaps between quantum and randomized computation.

References

[1] A. Ioffe, V. Tikhomirov. Duality of convex functions and extremum problems, Russian Mathematical Surveys, 23(6):53-124, 1968.

[2] H. Klauck. Lower bounds for quantum communication complexity, Proceedings of the 42nd Symposium on Foundations of Computer Science, 288-297, 2001.

[3] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions, Proceedings of the 35th Symposium on Theory of Computing, 325-334, 2003.

[4] A. Razborov. Quantum communication complexity of symmetric predicates, Izvestiya of the Russian Academy of Science, 67(1):159-176, 2003.

[5] A. Sherstov. The pattern matrix method, SIAM Journal on Computing, 40(6):1969-2000, 2011.