The Pattern Matrix Method for Lower Bounds on Quantum Communication

Alexander A. Sherstov

The Univ. of Texas at Austin, Dept. of Computer Sciences, Austin, TX 78712 USA
sherstov@cs.utexas.edu

September 5, 2007

Abstract

In a breakthrough result, Razborov (2003) gave optimal lower bounds on the communication complexity of every function f of the form $f(x, y) = D(|x \wedge y|)$ for some $D : \{0, 1, \ldots, n\} \to \{0, 1\}$, in the bounded-error quantum model with and without prior entanglement. This was proved by the *multidimensional* discrepancy method. We give an entirely different proof of Razborov's result, using the original, *one-dimensional* discrepancy method. This refutes the commonly held intuition (Razborov 2003) that the original discrepancy method fails for functions such as DISJOINTNESS.

More importantly, our communication lower bounds hold for a much broader class of functions for which no methods were available. Namely, fix an arbitrary function $f : \{0, 1\}^{n/2} \to \{0, 1\}$ and let A be the Boolean matrix whose rows are each an application of f to some subset of the variables $x_1, \neg x_1, \ldots, x_n, \neg x_n$. We prove that the communication complexity of A in the bounded-error quantum model with and without entanglement is $\Omega(d)$, where d is the $\frac{1}{3}$-approximate degree of f. From this result, Razborov's lower bounds follow easily.

Our proof technique is novel and has two ingredients. The first is a certain equivalence of approximation and orthogonality in Euclidean n-space, which we establish using linear-programming duality. The second is a new construction of suitably structured matrices with low spectral norm, which we realize using matrix analysis and the Fourier transform over $\mathbb{Z}_2^n$.

1 Introduction

Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be an arbitrary predicate. Consider the communication problem $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ given by

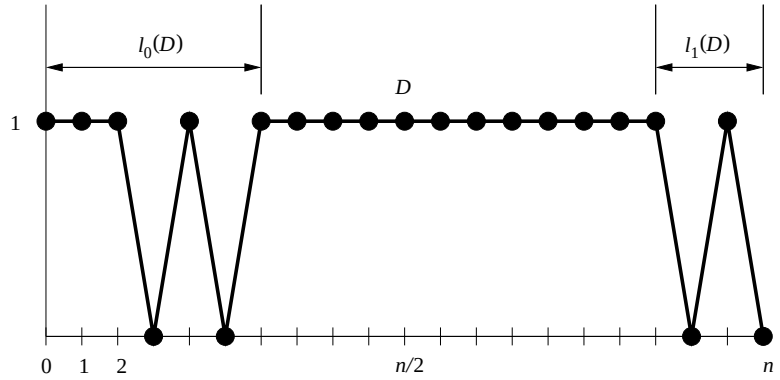
$$f(x, y) \stackrel{\text{def}}{=} D(|x \wedge y|),$$

where $|x \wedge y|$ stands for the number of positions where x and y both have a 1. As usual, the objective is for Alice and Bob to evaluate $f(x, y)$, where Alice holds x and Bob holds y.

As we will see shortly, the hardness of this communication problem depends on whether D changes value close to the middle of the range $\{0, 1, \ldots, n\}$. Specifically, define

$$\ell_0(D) \in \{0, 1, \ldots, \lfloor n/2 \rfloor\},$$
$$\ell_1(D) \in \{0, 1, \ldots, \lceil n/2 \rceil\}$$

to be the smallest integers such that D is constant in the range $[\ell_0(D), n - \ell_1(D)]$. The figure below illustrates this definition for a typical predicate D:



Let $Q^*_{1/3}(f)$ denote the quantum communication complexity of f with error $1/3$ in the model with prior entanglement. Define $Q_{1/3}(f)$ analogously for the model without prior entanglement. In a breakthrough result, Razborov [18] established optimal lower bounds on the quantum communication complexity of every function of the above form:

Theorem 1.1 (Razborov [18]). *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be an arbitrary predicate. Put $f(x, y) \stackrel{\text{def}}{=} D(|x \wedge y|)$. Then*

$$Q_{1/3}(f) \geqslant Q^*_{1/3}(f) \geqslant \Omega\left(\sqrt{n\ell_0(D)} + \ell_1(D)\right).$$

We give an entirely different proof of this theorem. In fact, we give communication lower bounds for a substantially broader class of functions that were beyond the reach of the current techniques. The setting for our work is as follows. Let n and t be integers with $t \mid n$. Fix an arbitrary function $f : \{0,1\}^t \to \{0,1\}$. Consider the communication problem of computing

$$f\Big(x|_V \ \oplus \ w\Big),$$

where:

- the bit string $x \in \{0,1\}^n$ is Alice's input;
- the bit string $w \in \{0,1\}^t$ and the set $V \subset \{1,2,\ldots,n\}$ with $|V| = t$ are Bob's inputs;
- and $x|_V$ denotes the projection of x onto the indices in V. Formally, $x|_V \stackrel{\text{def}}{=} (x_{i_1}, x_{i_2}, \ldots, x_{i_t}) \in \{0,1\}^t$, where $i_1 < i_2 < \cdots < i_t$ are the elements of V.

We prove communication lower bounds for this problem in the bounded-error quantum model with and without prior entanglement. Our lower bound is in terms of the *approximate degree* $\deg_{1/3}(f)$ of f, which is the least degree of a multivariate polynomial p with

$$|f(z) - p(z)| \leqslant \frac{1}{3} \qquad \text{for each } z \in \{0,1\}^t.$$

We prove:

Theorem 1.2 (Main Theorem). *Any quantum protocol, with or without prior entanglement, that solves the above problem with error probability at most* $1/5$ *on each input must exchange at least*

$$\frac{1}{4} \deg_{1/3}(f) \cdot \log\left(\frac{n}{t}\right) - 2$$

qubits.

The lower bound of Theorem 1.2 continues to hold when the sets V are restricted to have a particularly simple form; see Section 4 for details.

The value of Theorem 1.2 is that the notion of approximate degree plays an important role in complexity theory and has been studied in countless works. In particular, tight estimates of the approximate degree are available for many functions, including all symmetric functions (Paturi 1992) and some DNF formulas (Aaronson & Shi, 2004). As a result, our work gives strong lower bounds on communication for *any* of these base functions f with high approximate degree. To illustrate the applicability of Theorem 1.2, we use it to give a short and elementary proof of Razborov's celebrated result, Theorem 1.1.

Motivation. Our work is of interest for several reasons. First, the bounded-error quantum model with prior entanglement is the most powerful model of bounded-error communication. Despite encouraging recent progress [2, 3, 15, 18], proving lower bounds in this model remains difficult, and few general methods are available. We address this state of affairs by establishing, in Theorem 1.2, communication lower bounds for a rather broad class of functions for which no methods were previously known. In particular, Razborov's proof technique depends crucially on the high symmetry of the functions $D(|x \wedge y|)$, in addition to their high approximate degree. Our Theorem 1.2, on the other hand, completely removes the symmetry requirement: the symmetry or nonsymmetry of the base function f is irrelevant.

Second, our proof of Theorem 1.2 contributes considerable technical novelty. One of its ingredients is a new construction of matrices with low spectral norm, the *pattern matrices,* which correspond to hard communication problems. Matrix analysis and the Fourier transform over $\mathbb{Z}_2^n$ are critical to this construction. The pattern matrices reduce our task to finding, in the vicinity of the base function f, a real-valued function that is orthogonal to the low-degree parity functions. This is where the other major ingredient of our proof is needed, the *Approximation/Orthogonality Principle.* It asserts a certain equivalence between approximation and orthogonality, and we prove it using linear-programming duality. We are able to carry out this development in the general setting of Euclidean n-space, making for a cleaner and simpler proof. We describe the technical content of this paper in greater detail under "Techniques" below.

It is noteworthy that our proof of Theorem 1.2 fits in the framework of the original, *one-dimensional* discrepancy method (described in detail in Section 2.4), as opposed to Razborov's approach known as the *multidimensional* discrepancy method. This refutes the commonly held intuition that the original discrepancy method could not have yielded Theorem 1.1. For example, Razborov writes, "even this generalized form of the discrepancy method does not work for the disjointness predicate" [18, p. 155]. In view of our results, that statement was in error [19]. We emphasize that it does not affect the remainder of Razborov's article, which is correct.

Finally, we find it already valuable to offer a new proof of Razborov's result. This result is one of the strongest in communication complexity and, as such, deserves more than one proof. Despite sustained efforts by various authors [2,3,8–10,15], no such alternate proof exists. As a matter of fact, the next-best lower bounds for general predicates are nowhere close to Theorem 1.1. To illustrate, consider the familiar *disjointness predicate* D, given by $D(t) = 1 \Leftrightarrow t = 0$. Theorem 1.1 shows its communication complexity to be $\Omega(\sqrt{n})$, while the next-best lower bound [2, 3] is only $\Omega(\log n)$.

Razborov's proof technique has seen several applications, including direct-product theorems [11], separations for small-bias communication [4], and learning theory [12]. We hope that the ideas in this paper will also find uses beyond quantum communication.

Our Techniques. Razborov's proof proceeds by the *multidimensional* discrepancy method, a powerful extension of the original discrepancy method. Central to that technique are the so-called combinatorial matrices, which have rare and useful spectral properties. By contrast, we use the simpler, original discrepancy method. Roughly speaking, this original method works as follows. Let $F(x, y)$ be the Boolean function whose communication complexity is of interest. The method says: find a "hard" function $H(x, y)$ in the vicinity of F, thereby proving that F itself must be "somewhat hard." More precisely, the discrepancy method asks for a function $H(x, y)$ and a distribution μ on (x, y)-pairs such that:

- F and H are highly correlated under μ; and
- all low-cost quantum protocols have negligible advantage in computing H under μ.

If such H indeed exists, it follows that no low-cost protocol can compute F to high accuracy (or else it would be a good predictor for the hard function H as well!).

The discrepancy method thus reduces our task to finding the hard function H. It is here that we contribute a new technique. A key ingredient of this technique is a new construction of matrices with low spectral norm and suitable structure, the *pattern matrices.* The idea of a pattern matrix originated in a recent article by the author [22], where we introduced a somewhat different family of matrices and placed an upper bound on their spectral norm (with the end result of separating AC^0 from depth-2 majority circuits). This paper gives an *exact, closed-form* expression for the singular values of a pattern matrix and their multiplicities, substantially improving on the estimates from [22]. These exact calculations are in fact crucial to our main result: the earlier estimates would not be strong enough. As an additional benefit, our exact analysis here improves the main results of [22]; we discuss this in the concluding part of this work, Section 7.

The pattern matrices further reduce our challenge to proving the existence of a function $\psi : \{0, 1\}^t \to \mathbb{R}$ with two properties. First, ψ must be well-correlated with the base function f. Second, ψ must be orthogonal to all low-degree parity functions. To infer the existence of such ψ, we prove the *Approximation/Orthogonality Principle,* which states that the notions of approximation and orthogonality in Euclidean n-space are equivalent in a certain precise sense. Our proof here exploits linear-programming duality.

Once Theorem 1.2 is established, Razborov's lower bounds follow readily as a special case.

Organization. We start with a thorough review of technical preliminaries in Section 2. The two sections that follow are concerned with the two principal ingredients of our technique, the pattern matrices and the Approximation/Orthogonality Principle. Section 5 integrates them into the discrepancy method and establishes our main result, Theorem 1.2. Section 6 deduces from it Razborov's lower bounds. Section 7 uses our exact spectral calculations to strengthen the author's earlier result [22] on classical discrepancy and its applications to AC^0.

2 Preliminaries

This section provides relevant technical background. We describe our notation (Section 2.1) and then briefly review matrix analysis (Section 2.2), the quantum communication model (Section 2.3), and the discrepancy method for communication lower bounds (Section 2.4). Finally, we recall fundamental results on the approximation of Boolean functions by polynomials (Section 2.5).

2.1 General

A *Boolean function* is a mapping $X \to \{0,1\}$, where X is a finite set. Typically, $X = \{0,1\}^n$ or $X = \{0,1\}^n \times \{0,1\}^n$. A *predicate* is a mapping $D : \{0,1,\ldots,n\} \to \{0,1\}$. The notation $[n]$ stands for the set $\{1,2,\ldots,n\}$. For a set $S \subseteq [n]$, its *characteristic vector* $\mathbf{1}_S \in \{0,1\}^n$ is defined by

$$(\mathbf{1}_S)_i = \begin{cases} 1 & i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

For $b \in \{0,1\}$, we put $\neg b \stackrel{\text{def}}{=} 1 - b$. For $x \in \{0,1\}^n$, we write $|x| \stackrel{\text{def}}{=} |\{i : x_i = 1\}|$. For $x, y \in \{0,1\}^n$, the notation $x \wedge y$ refers as usual to the component-wise AND of x and y. In particular, $|x \wedge y|$ stands for the number of positions where x and y both have a 1.

Throughout this manuscript, "log" refers to the logarithm to base 2.

Finally, we recall the Fourier transform over $\mathbb{Z}_2^n$. Consider the vector space of functions $\{0,1\}^n \to \mathbb{R}$, equipped with the inner product

$$\langle f, g \rangle \stackrel{\text{def}}{=} \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) g(x).$$

For $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{-1,+1\}$ by $\chi_S(x) \stackrel{\text{def}}{=} (-1)^{\sum_{i \in S} x_i}$. Then $\{\chi_S\}_{S \subseteq [n]}$ is an orthonormal basis for the inner product space in question. As a result, every function $f : \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form

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

where $\hat{f}(S) \stackrel{\text{def}}{=} \langle f, \chi_S \rangle$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of* f. The following fact is immediate from the definition of $\hat{f}(S)$:

Proposition 2.1. *Let $f : \{0,1\}^n \to \mathbb{R}$ be given. Then*

$$\max_{S \subseteq [n]} |\hat{f}(S)| \leqslant \frac{1}{2^n} \sum_{x \in \{0,1\}^n} |f(x)|.$$

2.2 Matrix Analysis

We draw freely on basic notions from matrix analysis. For example, we assume familiarity with the singular value decomposition; positive semidefinite matrices; matrix similarity; matrix trace and its properties; the Kronecker product and its spectral properties; the relation between singular values and eigenvalues; and eigenvalue computation for matrices of simple form. An excellent reference on the subject is [7]. The review below is limited to notation and the more substantial results.

The symbol $\mathbb{R}^{m \times n}$ refers to the family of all $m \times n$ matrices with real entries. The (i, j)th entry of a matrix A is denoted by A_{ij}. We frequently use "generic-entry" notation to specify a matrix succinctly: we write $A = [F(i,j)]_{i,j}$ to mean that that the (i, j)th entry of A is given by the expression $F(i, j)$. In most matrices that arise in this work, the exact ordering of the columns (and rows) is irrelevant. In such cases we describe a matrix by the notation $[F(i,j)]_{i \in I,\, j \in J}$, where I and J are some index sets.

Let $A \in \mathbb{R}^{m \times n}$. We use the following standard notation:

$$\|A\|_\infty \stackrel{\text{def}}{=} \max_{i,j} \{|A_{ij}|\}, \qquad \|A\|_1 \stackrel{\text{def}}{=} \sum_{i,j} |A_{ij}|.$$

We denote the singular values of A by $\sigma_1(A) \geqslant \sigma_2(A) \geqslant \ldots \geqslant \sigma_{\min\{m,n\}}(A) \geqslant 0$.

Recall that the spectral norm, trace norm, and Frobenius norm of A are given by

$$\begin{aligned} \|A\| &= \max_{x \in \mathbb{R}^n,\, \|x\|=1} \|Ax\| = \sigma_1(A), \\ \|A\|_\Sigma &= \sum \sigma_i(A), \\ \|A\|_{\mathrm{F}} &= \sqrt{\sum A_{ij}^2} = \sqrt{\sum \sigma_i(A)^2}. \end{aligned}$$

Recall that every matrix $A \in \mathbb{R}^{m \times n}$ has a singular value decomposition $A = U\Sigma V^{\mathsf{T}}$, where U and V are both orthogonal matrices and Σ is diagonal with entries $\sigma_1(A), \sigma_2(A), \ldots, \sigma_{\min\{m,n\}}(A)$. For $A, B \in \mathbb{R}^{m \times n}$, we write $\langle A, B \rangle \stackrel{\text{def}}{=} \sum_{i,j} A_{ij} B_{ij}$. A useful consequence of the singular value decomposition is:

$$\langle A, B \rangle \leqslant \|A\| \, \|B\|_\Sigma \qquad\qquad (A, B \in \mathbb{R}^{m \times n}). \tag{2.1}$$

We will need the following well-known bound on the trace norm of a matrix product, which we state with a proof for the reader's convenience.

Proposition 2.2 (Trace norm of the product). $\|AB\|_\Sigma \leqslant \|A\|_{\mathrm{F}} \, \|B\|_{\mathrm{F}}$.

Proof. Write the singular value decomposition $AB = U\Sigma V^{\mathsf{T}}$. Let $u_1, u_2, \ldots$ and $v_1, v_2, \ldots$ stand for the columns of U and V, respectively. By definition, $\|AB\|_\Sigma$ is the sum of the diagonal entries of Σ. We have:

$$\begin{aligned} \|AB\|_\Sigma = \sum (U^{\mathsf{T}} ABV)_{ii} &= \sum (u_i^{\mathsf{T}} A)(Bv_i) \leqslant \sum \|A^{\mathsf{T}} u_i\| \, \|Bv_i\| \\ &\leqslant \sqrt{\sum \|A^{\mathsf{T}} u_i\|^2} \sqrt{\sum \|Bv_i\|^2} = \|U^{\mathsf{T}} A\|_{\mathrm{F}} \, \|BV\|_{\mathrm{F}} = \|A\|_{\mathrm{F}} \, \|B\|_{\mathrm{F}}. \end{aligned}$$

□

2.3 Quantum Communication

This section reviews the bounded-error model of quantum communication. We include this review mainly for completeness; our proofs rely solely on a standard lower-bound technique for such protocols and on no other aspect of quantum communication.

There are several equivalent ways to describe a quantum communication protocol. Our description closely follows Razborov [18]. Let $\mathcal{A}$ and $\mathcal{B}$ be complex finite-dimensional Hilbert spaces. Let C be a Hilbert space of dimension 2, whose orthonormal basis we denote by $|0\rangle$, $|1\rangle$. Consider the tensor product $\mathcal{A} \otimes C \otimes \mathcal{B}$, which is itself a Hilbert space with an inner product inherited from $\mathcal{A}$, $\mathcal{B}$, and C. The *state* of a quantum system is a unit vector in $\mathcal{A} \otimes C \otimes \mathcal{B}$, and conversely any such unit vector corresponds to a distinct quantum state. The quantum system starts in a given state and traverses a sequence of states, each obtained from the previous one

via a unitary transformation chosen according to the protocol. Formally, a *quantum communication protocol* is a finite sequence of unitary transformations

$$U_1 \otimes I_{\mathcal{B}}, \quad I_{\mathcal{A}} \otimes U_2, \quad U_3 \otimes I_{\mathcal{B}}, \quad I_{\mathcal{A}} \otimes U_4, \quad \ldots, \quad U_{2k-1} \otimes I_{\mathcal{B}}, \quad I_{\mathcal{A}} \otimes U_{2k},$$

where: $I_{\mathcal{A}}$ and $I_{\mathcal{B}}$ are the identity transformations in $\mathcal{A}$ and $\mathcal{B}$, respectively; $U_1, U_3, \ldots, U_{2k-1}$ are unitary transformations in $\mathcal{A} \otimes C$; and $U_2, U_4, \ldots, U_{2k}$ are unitary transformations in $C \otimes \mathcal{B}$. The *cost* of the protocol is the length of this sequence, namely, $2k$. On Alice's input $x \in X$ and Bob's input $y \in Y$ (where X, Y are given finite sets), the computation proceeds as follows.

1. The quantum system starts out in an initial state $\mathsf{Initial}(x, y)$.

2. Through successive applications of the above unitary transformations, the system reaches the state

$$\mathsf{Final}(x, y) \stackrel{\text{def}}{=} (I_{\mathcal{A}} \otimes U_{2k})(U_{2k-1} \otimes I_{\mathcal{B}}) \cdots (I_{\mathcal{A}} \otimes U_2)(U_1 \otimes I_{\mathcal{B}})\, \mathsf{Initial}(x, y).$$

3. Let v denote the projection of $\mathsf{Final}(x, y)$ onto $\mathcal{A} \otimes \mathrm{span}(|1\rangle) \otimes \mathcal{B}$. The output of the protocol is 1 with probability $\langle v, v \rangle$, and 0 with probability $1 - \langle v, v \rangle$.

All that remains is to specify how the initial state $\mathsf{Initial}(x, y) \in \mathcal{A} \otimes C \otimes \mathcal{B}$ is constructed from x, y. It is here that the model with prior entanglement differs from the model without prior entanglement.

In the model without prior entanglement, $\mathcal{A}$ and $\mathcal{B}$ have orthonormal bases $\{|x, w\rangle : x \in X,\ w \in W\}$ and $\{|y, w\rangle : y \in Y,\ w \in W\}$, respectively, where W is a finite set corresponding to the private workspace of each of the parties. The initial state is the pure state

$$\mathsf{Initial}(x, y) = |x, 0\rangle\, |0\rangle\, |y, 0\rangle,$$

where $0 \in W$ is a certain fixed element. In the model with prior entanglement, the spaces $\mathcal{A}$ and $\mathcal{B}$ have orthonormal bases $\{|x, w, e\rangle : x \in X,\ w \in W,\ e \in E\}$ and $\{|y, w, e\rangle : y \in Y,\ w \in W,\ e \in E\}$, respectively, where W is as before and E is a finite set corresponding to the prior entanglement. The initial state is now the entangled state

$$\mathsf{Initial}(x, y) = \frac{1}{\sqrt{|E|}} \sum_{e \in E} |x, 0, e\rangle\, |0\rangle\, |y, 0, e\rangle.$$

Apart from finite size, no assumptions are made about W or E. In particular, the model with prior entanglement allows for an unlimited supply of entangled qubits. This mirrors the unlimited supply of shared random bits in the classical public-coin randomized model.

Let $f : X \times Y \to \{0,1\}$ be a given function. A quantum protocol P is said to compute f with error ϵ if

$$\mathbf{Pr}[P(x,y) \neq f(x,y)] \leqslant \epsilon \qquad \text{for all } x, y,$$

where the random variable $P(x,y) \in \{0,1\}$ is the output of the protocol on input (x,y). Let $Q_\epsilon(f)$ denote the least cost of a quantum protocol without prior entanglement that computes f with error ϵ. Define $Q^*_\epsilon(f)$ analogously for protocols with prior entanglement. The precise choice of a constant $\epsilon \in (0,1)$ affects $Q_\epsilon(f)$ and $Q^*_\epsilon(f)$ by at most a constant factor, and thus the setting $\epsilon = 1/3$ entails no loss of generality.

Let $D : \{0,1,\ldots,n\} \to \{0,1\}$ be a predicate. We associate with D the function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ defined by

$$f(x,y) \stackrel{\text{def}}{=} D(|x \wedge y|).$$

We let $Q_\epsilon(D) \stackrel{\text{def}}{=} Q_\epsilon(f)$ and $Q^*_\epsilon(D) \stackrel{\text{def}}{=} Q^*_\epsilon(f)$. More generally, by computing D in the quantum model we mean computing the associated function f. As one last convention, by the communication complexity of a Boolean matrix $F = [F_{ij}]_{i \in I,\, j \in J}$ is meant the communication complexity of the associated function $f : I \times J \to \{0,1\}$, given by

$$f(i,j) = F_{ij}.$$

2.4 The Discrepancy Method

The discrepancy method is an intuitive and elegant technique for proving lower bounds on quantum communication. A starting point in our discussion is the following fact.

Theorem 2.3 (Razborov [18, Thm. 5.5], Linial & Shraibman [15, Lem. 10]). *Let X, Y be finite sets. Let P be a quantum protocol (with or without prior entanglement) with cost C qubits and input sets X and Y. Then*

$$\Big[\mathbf{E}[P(x,y)]\Big]_{x,y} = AB$$

for some real matrices A, B with $\|A\|_{\mathrm{F}} \leqslant 2^C \sqrt{|X|}$ and $\|B\|_{\mathrm{F}} \leqslant 2^C \sqrt{|Y|}$.

Theorem 2.3 states that the matrix of acceptance probabilities, $[\mathbf{E}[P(x,y)]]_{x,y}$, of every low-cost protocol P has some nontrivial factorization. This transition from quantum protocols to matrix factorization is a standard technique and has been applied by various authors in various contexts [6, 9, 13, 15]. We now state the discrepancy method as adapted to the quantum model by Razborov [18]. This is not to be confused with the *multidimensional discrepancy method,* also due to Razborov [18], which we will have no occasion to use or describe.

Theorem 2.4 (Discrepancy method, Razborov [18, Sec. 5.2], implicit). *Let X, Y be finite sets and $f : X \times Y \to \{0,1\}$ a given function. Let $K = [K_{xy}]_{x\in X,\, y\in Y}$ be any real matrix with $\|K\|_1 = 1$. Then for each $\epsilon > 0$,*

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

where $M \stackrel{\text{def}}{=} \left[(-1)^{f(x,y)}\right]_{x\in X,\, y\in Y}$.

Proof. Let P be a quantum protocol with prior entanglement that computes f with error ϵ and cost C. Put

$$\Pi \stackrel{\text{def}}{=} \left[\mathbf{E}[P(x,y)]\right]_{x\in X,\, y\in Y}.$$

Then we can write $M = (J - 2\Pi) + 2E$, where J is the all-ones matrix and E is some matrix with $\|E\|_\infty \leqslant \epsilon$. As a result,

$$\begin{aligned}
\langle K, J - 2\Pi\rangle &= \langle K, M\rangle - 2\,\langle K, E\rangle \\
&\geqslant \langle K, M\rangle - 2\epsilon\,\|K\|_1 \\
&= \langle K, M\rangle - 2\epsilon. \qquad (2.2)
\end{aligned}$$

On the other hand, Theorem 2.3 guarantees the existence of matrices A and B with $AB = \Pi$ and $\|A\|_{\mathrm{F}}\,\|B\|_{\mathrm{F}} \leqslant 4^C\,\sqrt{|X|\,|Y|}$. Therefore,

$$\begin{aligned}
\langle K, J - 2\Pi\rangle &\leqslant \|K\|\,\|J - 2\Pi\|_\Sigma && \text{by (2.1)} \\
&\leqslant \|K\|\,\left(\sqrt{|X|\,|Y|} + 2\,\|\Pi\|_\Sigma\right) && \text{since } \|J\|_\Sigma = \sqrt{|X|\,|Y|} \\
&\leqslant \|K\|\,\left(\sqrt{|X|\,|Y|} + 2\,\|A\|_{\mathrm{F}}\,\|B\|_{\mathrm{F}}\right) && \text{by Prop. 2.2} \\
&\leqslant \|K\|\,\left(2\cdot 4^C + 1\right)\sqrt{|X|\,|Y|}. && (2.3)
\end{aligned}$$

The theorem follows by comparing (2.2) and (2.3). □

Remark 2.5. A quick glance at the proof reveals that Theorem 2.4 is valid with $M = [f(x,y)]_{x,y}$. In fact, this choice of M would slightly simplify the proof of this theorem as well as its primary use in the paper, Theorem 5.1. Nevertheless, we prefer the definition of M as a sign matrix because, as we shall see shortly, this makes it possible to view the above proof in terms of *correlation* and relate Theorem 2.4 to the classical discrepancy method.

We now reinterpret Theorem 2.4 and its proof in a different terminology, which will clarify it and show that it is simply an extension of the classical discrepancy

method to the quantum model. Let $f : X \times Y \to \{0,1\}$ be a given function whose communication complexity we wish to estimate. The underlying communication model is irrelevant at this point. Suppose we can find a function $h : X \times Y \to \{0,1\}$ and a distribution μ on $X \times Y$ that satisfy the following two properties:

1. **Correlation of f and h.** The functions f and h are well correlated under μ:

$$\mathop{\mathbf{E}}_{(x,y)\sim\mu}\left[(-1)^{f(x,y)+h(x,y)}\right] \geqslant \epsilon, \tag{2.4}$$

   where $\epsilon > 0$ is a given constant.

2. **Hardness of h.** No low-cost protocol P in the given model of communication can compute h to a substantial advantage under μ. Formally, if P is a protocol in the given model with cost C bits, then

$$\mathop{\mathbf{E}}_{(x,y)\sim\mu}\left[(-1)^{h(x,y)}\,\mathbf{E}\left[(-1)^{P(x,y)}\right]\right] \leqslant 2^{O(C)}\gamma, \tag{2.5}$$

   where $\gamma = o(1)$. The inner expectation in (2.5) is over the internal operation of the protocol on the fixed input (x, y).

If the above two conditions hold, we claim that any protocol in the given model that computes f with error at most $\epsilon/3$ on each input must have cost $\Omega\left(\log \frac{\epsilon}{\gamma}\right)$. Indeed, let P be a protocol with $\mathbf{Pr}[P(x,y) \neq f(x,y)] \leqslant \epsilon/3$ for all x, y. Then standard manipulations reveal:

$$\mathop{\mathbf{E}}_{(x,y)\sim\mu}\left[(-1)^{h(x,y)}\,\mathbf{E}\left[(-1)^{P(x,y)}\right]\right] \;\geqslant\; \mathop{\mathbf{E}}_{(x,y)\sim\mu}\left[(-1)^{f(x,y)+h(x,y)}\right] - 2\cdot\frac{\epsilon}{3} \;\overset{(2.4)}{\geqslant}\; \frac{\epsilon}{3}.$$

In view of (2.5), this shows that P must have cost $\Omega\left(\log \frac{\epsilon}{\gamma}\right)$.

We call the described lower-bound technique the *discrepancy method,* following the terminology of Razborov [18]. Some authors, including Kushilevitz and Nisan [14], restrict the term "discrepancy method" to the case when $f = h$ and the communication takes place in the classical randomized model. This restriction apparently reflects the fact that the method originated in the classical setting, before the need to study quantum models arose. Our broad usage of the term is meant to highlight the fundamental mathematical technique in question, which is clearly independent of the communication model.

Indeed, the communication model enters the picture only in the proof of (2.5). It is here that the analysis must exploit the particularities of the model. To place an upper bound on the advantage under μ in the quantum model with entanglement, as we see from (2.3), one considers the quantity $\|K\|\sqrt{|X|\,|Y|}$, where

$K = [h(x,y)\mu(x,y)]_{x,y}$. In the classical randomized model, the quantity to estimate happens to be

$$\max_{\substack{S\subseteq X,\\ T\subseteq Y}} \left|\sum_{x\in S}\sum_{y\in T}\mu(x,y)h(x,y)\right|,$$

which is actually known as the *discrepancy of h under μ*.

2.5 Approximation by Polynomials

Let $f : \{0,1\}^n \to \mathbb{R}$ be given. As we saw in Section 2.1, any such function f has an *exact* representation as a linear combination of χ_S, where $S \subseteq [n]$. A fundamental question to ask is how closely f can be *approximated* by a linear combination of functions χ_S with $|S|$ small.

Definition 2.6 (Approximate degree of functions). Let $f : \{0,1\}^n \to \mathbb{R}$ and $\epsilon \geqslant 0$. The *ϵ-approximate degree* $\deg_\epsilon(f)$ *of* f is the minimum integer $d \in \{0,1,\ldots,n\}$ for which there exists $\phi \in \operatorname{span}(\{\chi_S\}_{|S|\leqslant d})$ with

$$\max_{x\in\{0,1\}^n} |f(x) - \phi(x)| \leqslant \epsilon.$$

We will be primarily interested in the approximate degree of Boolean functions. As a first observation, $\deg_\epsilon(f) = \deg_\epsilon(\neg f)$ for all such functions and all $\epsilon \geqslant 0$. Second, the exact choice of constant $\epsilon \in (0, 1/2)$ affects $\deg_\epsilon(f)$ only by a multiplicative constant. This fact is well-known and follows from basic approximation theory. It is therefore standard practice to work with $\deg_{1/3}(f)$ by default. Finally, the approximate degree does not change much as one switches from the $\{0,1\}$ representation of Boolean functions to the $\{-1,+1\}$ representation. More precisely, fix $f : \{0,1\}^n \to \{0,1\}$ and define $f^*(x) = (-1)^{f(x)}$. Then

$$\deg_\epsilon(f^*) = \deg_{\epsilon/2}(f) \qquad \text{for all } \epsilon \geqslant 0, \tag{2.6}$$

as one can verify from the equation $f^* = 1 - 2f$.

We will take a special interest in *symmetric* Boolean functions, i.e., functions $f : \{0,1\}^n \to \{0,1\}$ whose value is uniquely determined by $x_1 + \cdots + x_n$. Equivalently, a Boolean function f is symmetric if and only if

$$f(x_1, x_2, \ldots, x_n) = f(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$$

for all inputs $x \in \{0,1\}^n$ and all permutations $\sigma : [n] \to [n]$. Note that there is a one-to-one correspondence between predicates and symmetric Boolean functions. Namely, one associates a predicate D with the symmetric function

$$f(x) \stackrel{\text{def}}{=} D(x_1 + \cdots + x_n).$$

Asymptotically tight estimates of the $\frac{1}{3}$-approximate degree are available for every symmetric Boolean function (Paturi 1992). These estimates are in terms of the quantities $\ell_0(f)$ and $\ell_1(f)$, defined next.

Definition 2.7. Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Define

$$\ell_0(D) \in \{0, 1, \ldots, \lfloor n/2 \rfloor\},$$
$$\ell_1(D) \in \{0, 1, \ldots, \lceil n/2 \rceil\}$$

to be the smallest integers such that D is constant in the range $[\ell_0(D), n-\ell_1(D)]$. For a symmetric function $f : \{0,1\}^n \to \{0,1\}$, define $\ell_0(f) = \ell_0(D)$ and $\ell_1(f) = \ell_1(D)$, where D is the predicate for which $f(x) \equiv D(x_1 + \cdots + x_n)$.

See Section 1 for a pictorial illustration of this definition. We are now ready to state Paturi's fundamental theorem.

Theorem 2.8 (Paturi [17]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a symmetric function. Then*

$$\deg_{1/3}(f) = \Theta\left(\sqrt{n(\ell_0(f) + \ell_1(f))}\right).$$

3 The Approximation/Orthogonality Principle

This section marks the beginning of our proof. Here we consider the notions of *approximation* and *orthogonality* in Euclidean space and establish a certain equivalence between them. We will later reinterpret this result in terms of protocols rather than points in Euclidean space.

Let X be a finite set. Consider $\mathbb{R}^X$, the linear space of all functions $X \to \mathbb{R}$. For $\phi \in \mathbb{R}^X$, let

$$\|\phi\|_\infty \stackrel{\text{def}}{=} \max_{x \in X} |\phi(x)|.$$

Then $(\mathbb{R}^X, \|\cdot\|_\infty)$ is a real normed linear space.

Definition 3.1 (Best error). For $f : X \to \mathbb{R}$ and $\Phi \subseteq \mathbb{R}^X$, let

$$\epsilon^*(f, \Phi) \stackrel{\text{def}}{=} \min_{\phi \in \text{span}(\Phi)} \|f - \phi\|_\infty.$$

In words, $\epsilon^*(f, \Phi)$ is the best error in an approximation of f by a linear combination of functions in Φ. Since $\text{span}(\Phi)$ has finite dimension, a best approximation to f out of $\text{span}(\Phi)$ always exists [20, Thm. I.1], justifying our use of "min" instead of "inf" in the above definition.

We now introduce a closely related quantity, $\gamma^*(f, \Phi)$, that measures how well f correlates with a real function that is orthogonal to all of Φ.

Definition 3.2 (Modulus of orthogonality). Let X be a finite set, $f : X \to \mathbb{R}$, and $\Phi \subseteq \mathbb{R}^X$. The *modulus of orthogonality* of f with respect to Φ is:

$$\gamma^*(f, \Phi) \stackrel{\text{def}}{=} \max_{\psi} \left\{ \sum_{x \in X} f(x)\psi(x) \right\}, \tag{3.1}$$

where the maximum is taken over all $\psi : X \to \mathbb{R}$ such that $\sum_{x \in X} |\psi(x)| \leqslant 1$ and $\sum_{x \in X} \phi(x)\psi(x) = 0$ for all $\phi \in \Phi$.

The maximization in (3.1) is over a nonempty compact set that contains $\psi = 0$. Also, the use of "max" instead of "sup" is legitimate because (3.1) maximizes a continuous function over a compact set. To summarize, the modulus of orthogonality is a well-defined nonnegative real number for every function $f : X \to \mathbb{R}$.

A key result, which we now prove, is that the best error and the modulus of orthogonality are always equal. We call this the *Approximation/Orthogonality Principle.*

Theorem 3.3 (Approximation/Orthogonality Principle). *Let X be a finite set, $\Phi \subseteq \mathbb{R}^X$, and $f : X \to \mathbb{R}$. Then*

$$\epsilon^*(f, \Phi) = \gamma^*(f, \Phi).$$

Proof. The theorem holds trivially when $\text{span}(\Phi) = \{0\}$. In the contrary case, let $\phi_1, \ldots, \phi_k$ be a basis for $\text{span}(\Phi)$. Our first observation is that $\epsilon^*(f, \Phi)$ is the optimum of the following linear program in the variables $\epsilon, \alpha_1, \ldots, \alpha_k$:

minimize:	ϵ	
subject to:		
	$\left\lvert f(x) - \sum_{i=1}^{k} \alpha_i \phi_i(x) \right\rvert \leqslant \epsilon$	for each $x \in X$,
	$\alpha_i \in \mathbb{R}$	for each i,
	$\epsilon \geqslant 0.$	

Standard manipulations reveal the dual:

maximize:	$\sum_{x \in X} \beta_x f(x)$	
subject to:		
	$\sum_{x \in X} \lvert\beta_x\rvert \leqslant 1,$	
	$\sum_{x \in X} \beta_x \phi_i(x) = 0$	for each i,
	$\beta_x \in \mathbb{R}$	for each $x \in X$.

Both programs are clearly feasible and thus have the same finite optimum. We have already observed that the optimum of first program is $\epsilon^*(f, \Phi)$. Since $\phi_1, \ldots, \phi_k$ form a basis for span(Φ), the optimum of the second program is by definition $\gamma^*(f, \Phi)$. □

A useful consequence of the Approximation/Orthogonality Principle for our purposes is the following result.

Corollary 3.3.1. *Fix $\epsilon \geqslant 0$. Let $f : \{0,1\}^t \to \mathbb{R}$ be given with $d \stackrel{\text{def}}{=} \deg_\epsilon(f) \geqslant 1$. Then there is a function $\psi : \{0,1\}^t \to \mathbb{R}$ such that:*

$$\begin{aligned} &\hat{\psi}(S) = 0 && \text{for } |S| < d, \\ &\sum_{z \in \{0,1\}^t} |\psi(z)| = 1, \\ &\sum_{z \in \{0,1\}^t} \psi(z) f(z) > \epsilon. \end{aligned}$$

Proof. Set $X = \{0,1\}^t$ and $\Phi = \{\chi_S : |S| < d\} \subset \mathbb{R}^X$. Since $\deg_\epsilon(f) = d$, we immediately have that $\epsilon^*(f, \Phi) > \epsilon$. But then $\gamma^*(f, \Phi) > \epsilon$ by the Approximation/Orthogonality Principle (Theorem 3.3). Clearly, we can take ψ to be any function for which the maximum is achieved in (3.1). □

4 Pattern Matrices

We now turn to the second ingredient of our proof, a certain family of real matrices that we call *pattern matrices.* Our goal here is to explicitly calculate their spectral norm. As we shall see later, this provides a convenient means to generate hard communication problems.

Let t and n be positive integers with $t \mid n$. Split $[n]$ into t contiguous blocks, each with n/t elements:

$$[n] = \left\{1, 2, \ldots, \frac{n}{t}\right\} \cup \left\{\frac{n}{t} + 1, \ldots, \frac{2n}{t}\right\} \cup \cdots \cup \left\{\frac{(t-1)n}{t} + 1, \ldots, n\right\}.$$

Let $\mathcal{V}(n, t)$ denote the family of subsets $V \subseteq [n]$ that have exactly one element in each of these blocks (in particular, $|V| = t$). Clearly, $|\mathcal{V}(n, t)| = (n/t)^t$. For a bit string $x \in \{0, 1\}^n$ and a set $V \in \mathcal{V}(n, t)$, define the *projection of x onto V* by

$$x|_V \stackrel{\text{def}}{=} (x_{i_1}, x_{i_2}, \ldots, x_{i_t}) \in \{0, 1\}^t,$$

where $i_1 < i_2 < \cdots < i_t$ are the elements of V.

Definition 4.1 (Pattern matrix). For $\phi : \{0, 1\}^t \to \mathbb{R}$, the (n, t, ϕ)-*pattern matrix* is the real matrix A given by

$$A = \Big[\phi(x|_V \oplus w)\Big]_{x \in \{0,1\}^n,\, (V,w) \in \mathcal{V}(n,t) \times \{0,1\}^t}.$$

In words, A is the matrix of size 2^n by $2^t(n/t)^t$ whose rows are indexed by strings $x \in \{0, 1\}^n$, whose columns are indexed by pairs $(V, w) \in \mathcal{V}(n, t) \times \{0, 1\}^t$, and whose entries are given by $A_{x,(V,w)} = \phi(x|_V \oplus w)$.

The logic behind the term "pattern matrix" is as follows: a mosaic arises from repetitions of a pattern in the same way that A arises from applications of ϕ to various subsets of the variables.

Our approach to analyzing the singular values of a pattern matrix A will be to represent it as the sum of simpler matrices and analyze them instead. For this to work, we should be able to reconstruct the singular values of A from those of the simpler matrices. Just when this can be done is the subject of the following lemma.

Lemma 4.2 (Singular values of a matrix sum). *Let A, B be real matrices with $AB^{\mathsf{T}} = 0$ and $A^{\mathsf{T}}B = 0$. Then the nonzero singular values of $A + B$, counting multiplicities, are $\sigma_1(A), \ldots, \sigma_{\operatorname{rank} A}(A), \sigma_1(B), \ldots, \sigma_{\operatorname{rank} B}(B)$.*

Proof. The claim is trivial when $A = 0$ or $B = 0$, so assume otherwise. Since the singular values of $A + B$ are precisely the square roots of the eigenvalues of $(A + B)(A + B)^{\mathsf{T}}$, it suffices to compute the spectrum of the latter matrix. Now,

$$\begin{aligned}(A + B)(A + B)^{\mathsf{T}} &= AA^{\mathsf{T}} + BB^{\mathsf{T}} + \underbrace{AB^{\mathsf{T}}}_{=0} + \underbrace{BA^{\mathsf{T}}}_{=0} \\ &= AA^{\mathsf{T}} + BB^{\mathsf{T}}. \end{aligned} \tag{4.1}$$

Fix spectral decompositions

$$AA^\mathsf{T} = \sum_{i=1}^{\operatorname{rank} A} \sigma_i(A)^2 u_i u_i^\mathsf{T}, \qquad BB^\mathsf{T} = \sum_{j=1}^{\operatorname{rank} B} \sigma_j(B)^2 v_j v_j^\mathsf{T}.$$

Then

$$\begin{aligned}
\sum_{i=1}^{\operatorname{rank} A} \sum_{j=1}^{\operatorname{rank} B} \sigma_i(A)^2 \sigma_j(B)^2 \langle u_i, v_j \rangle^2 &= \left\langle \sum_{i=1}^{\operatorname{rank} A} \sigma_i(A)^2 u_i u_i^\mathsf{T}, \sum_{j=1}^{\operatorname{rank} B} \sigma_j(B)^2 v_j v_j^\mathsf{T} \right\rangle \\
&= \langle AA^\mathsf{T}, BB^\mathsf{T} \rangle \\
&= \operatorname{trace}(AA^\mathsf{T} BB^\mathsf{T}) \\
&= \operatorname{trace}(A \cdot 0 \cdot B^\mathsf{T}) \\
&= 0. \qquad (4.2)
\end{aligned}$$

Since $\sigma_i(A)\,\sigma_j(B) > 0$ for all i, j, it follows from (4.2) that $\langle u_i, v_j \rangle = 0$ for all i, j. Put differently, the vectors $u_1, \ldots, u_{\operatorname{rank} A}, v_1, \ldots, v_{\operatorname{rank} B}$ form an orthonormal set. Recalling (4.1), we conclude that the spectral decomposition of $(A + B)(A + B)^\mathsf{T}$ is

$$\sum_{i=1}^{\operatorname{rank} A} \sigma_i(A)^2 u_i u_i^\mathsf{T} + \sum_{j=1}^{\operatorname{rank} B} \sigma_j(B)^2 v_j v_j^\mathsf{T},$$

and thus the nonzero eigenvalues of $(A + B)(A + B)^\mathsf{T}$ are as claimed. □

We are ready to analyze the spectral norm of a pattern matrix.

Theorem 4.3 (Singular values of a pattern matrix). *Let $\phi : \{0,1\}^t \to \mathbb{R}$ be given. Let A be the (n, t, ϕ)-pattern matrix. Then the nonzero singular values of A, counting multiplicities, are:*

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

In particular,

$$\|A\| = \sqrt{2^{n+t} \left(\frac{n}{t}\right)^t} \max_{S \subseteq [t]} \left\{ |\hat{\phi}(S)| \left(\frac{t}{n}\right)^{|S|/2} \right\}.$$

Proof. For each $S \subseteq [t]$, let A_S be the (n, t, χ_S)-pattern matrix. Thus,

$$A = \sum_{S \subseteq [t]} \hat{\phi}(S) A_S. \qquad (4.3)$$

Fix arbitrary $S, T \subseteq [t]$ with $S \neq T$. Then

$$\begin{aligned} A_S A_T^\mathsf{T} &= \left[\sum_{V \in \mathcal{V}(n,t)} \sum_{w \in \{0,1\}^t} \chi_S(x|_V \oplus w)\, \chi_T(y|_V \oplus w) \right]_{x,y} \\ &= \left[\sum_{V \in \mathcal{V}(n,t)} \chi_S(x|_V)\, \chi_T(y|_V) \underbrace{\sum_{w \in \{0,1\}^t} \chi_S(w)\, \chi_T(w)}_{=0} \right]_{x,y} \\ &= 0. \end{aligned} \tag{4.4}$$

Similarly,

$$A_S^\mathsf{T} A_T = \left[\chi_S(w)\, \chi_T(w') \underbrace{\sum_{x \in \{0,1\}^n} \chi_S(x|_V)\, \chi_T(x|_{V'})}_{=0} \right]_{(V,w),(V',w')} = 0. \tag{4.5}$$

By (4.3)–(4.5) and Lemma 4.2, the nonzero singular values of A are the union of the nonzero singular values of all $\hat{\phi}(S) A_S$, counting multiplicities. Therefore, the proof will be complete once we show that the only nonzero singular value of $A_S^\mathsf{T} A_S$ is $2^{n+t}(n/t)^{t-|S|}$, with multiplicity $(n/t)^{|S|}$.

We proceed to analyze the spectrum of $A_S^\mathsf{T} A_S$. It is convenient to write this matrix as the Kronecker product

$$A_S^\mathsf{T} A_S \;=\; [\chi_S(w)\chi_S(w')]_{w,w'} \;\otimes\; \left[\sum_{x \in \{0,1\}^n} \chi_S(x|_V)\, \chi_S(x|_{V'}) \right]_{V,V'}.$$

The first matrix in this factorization has rank 1 and entries ± 1, which means that its only nonzero singular value is 2^t with multiplicity 1. The other matrix, call it M, is permutation-similar to

$$2^n \begin{bmatrix} J & & & \\ & J & & \\ & & \ddots & \\ & & & J \end{bmatrix},$$

where J is the all-ones square matrix of order $(n/t)^{t-|S|}$. This means that the only nonzero singular value of M is $2^n(n/t)^{t-|S|}$ with multiplicity $(n/t)^{|S|}$. It follows from elementary properties of the Kronecker product that the spectrum of $A_S^\mathsf{T} A_S$ is as desired. □

5 The Pattern Matrix Method

The previous two sections studied the spectrum of pattern matrices and the relationship between approximation and orthogonality. Having examined these notions in their pure and basic form, we now apply our findings to communication complexity. Specifically, we establish the *pattern matrix method* for communication lower bounds, which gives strong lower bounds for every pattern matrix generated by a Boolean function with high approximate degree. The theorem we are about to prove is the main result of this paper, stated in the Introduction as Theorem 1.2.

Theorem 5.1 (The Pattern Matrix Method). *Let F be the (n, t, f)-pattern matrix, where $f : \{0,1\}^t \to \{0,1\}$ is given. Put $d \stackrel{\text{def}}{=} \deg_{1/3}(f)$. Then*

$$Q_{1/5}(F) \;\geqslant\; Q^*_{1/5}(F) \;>\; \frac{1}{4}\, d \log\left(\frac{n}{t}\right) - 2.$$

Proof. Define $f^* : \{0,1\}^t \to \{-1,+1\}$ by $f^*(z) = (-1)^{f(z)}$. Then (2.6) shows that $\deg_{2/3}(f^*) = d$. By Corollary 3.3.1, there is a function $\psi : \{0,1\}^t \to \mathbb{R}$ such that:

$$\hat{\psi}(S) = 0 \qquad \text{for } |S| < d, \tag{5.1}$$

$$\sum_{z \in \{0,1\}^t} |\psi(z)| = 1, \tag{5.2}$$

$$\sum_{z \in \{0,1\}^t} \psi(z) f^*(z) > \frac{2}{3}. \tag{5.3}$$

Let M be the (n, t, f^*)-pattern matrix. Let K be the $(n, t, 2^{-n}(n/t)^{-t}\psi)$-pattern matrix. Immediate consequences of (5.2) and (5.3) are:

$$\|K\|_1 = 1, \qquad \langle K, M \rangle > \frac{2}{3}. \tag{5.4}$$

Our last task is to calculate $\|K\|$. By (5.2) and Proposition 2.1,

$$\max_{S \subseteq [t]} |\hat{\psi}(S)| \;\leqslant\; \frac{1}{2^t}. \tag{5.5}$$

Theorem 4.3 yields, in view of (5.1) and (5.5):

$$\|K\| \leqslant \left(\frac{t}{n}\right)^{d/2} \left(2^{n+t} \left(\frac{n}{t}\right)^t\right)^{-1/2}. \tag{5.6}$$

The desired lower bounds on quantum communication now follow directly from (5.4) and (5.6) by the Discrepancy Method (Theorem 2.4). □

6 Optimal Lower Bounds for Every Symmetric Function

As an illustrative application of the pattern matrix method, we now give a short and elementary proof of Razborov's optimal lower bounds for every predicate $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ (Theorem 1.1).

As a first step, we show that a pattern matrix occurs as a submatrix of $[D(|x \wedge y|)]_{x,y}$, the communication matrix of D. This will immediately put the pattern matrix machinery at our disposal.

Lemma 6.1. *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be a given predicate. Let F be the $(2\lfloor n/4 \rfloor, \lfloor n/4 \rfloor, f)$-pattern matrix, where $f(z) \stackrel{\text{def}}{=} D(|z|)$. Then F is a submatrix of*

$$\Big[D(|x \wedge y|)\Big]_{x \in \{0,1\}^n,\, y \in \{0,1\}^n}. \tag{6.1}$$

Proof. Put $m \stackrel{\text{def}}{=} \lfloor n/4 \rfloor$. By definition,

$$F = \Big[D(|\ x|_V \ \oplus \ w\ |)\Big]_{x \in \{0,1\}^{2m},\ (V,w) \in \mathcal{V}(2m,m) \times \{0,1\}^m}.$$

We will define one-to-one maps

$$\begin{aligned} \alpha :&\quad \{0,1\}^{2m} \to \{0,1\}^n, \\ \beta :&\quad \mathcal{V}(2m, m) \times \{0,1\}^m \to \{0,1\}^n \end{aligned}$$

such that

$$|\ x|_V \ \oplus \ w\ | \ = \ |\ \alpha(x) \wedge \beta(V, w)\ | \qquad \text{for all } x, V, w. \tag{6.2}$$

Obviously, this will mean that F is a submatrix of (6.1).

As usual, let juxtaposition of bit strings stand for their concatenation, e.g., $(0, 1)(1, 0, 1) = (0, 1, 1, 0, 1)$. With this convention, define α by

$$\alpha(x_1, x_2, \ldots, x_{2m}) \ \stackrel{\text{def}}{=} \ (x_1, \neg x_1, x_2, \neg x_2, \ldots, x_{2m}, \neg x_{2m})\, 0^{n-4m}.$$

Define β by

$$\beta(V, w) \ \stackrel{\text{def}}{=} \ \gamma(i_1, w_1)\, \gamma(i_2, w_2) \cdots \gamma(i_m, w_m)\, 0^{n-4m},$$

where $i_1 < i_2 < \cdots < i_m$ are the elements of V, and $\gamma : \mathbb{Z} \times \mathbb{Z} \to \{0, 1\}^4$ is given by

$$\gamma(a, b) \ \stackrel{\text{def}}{=} \ \begin{cases} (1,0,0,0) & \text{if } a \text{ is odd, } b \text{ is even,} \\ (0,1,0,0) & \text{if } a \text{ is odd, } b \text{ is odd,} \\ (0,0,1,0) & \text{if } a \text{ is even, } b \text{ is even,} \\ (0,0,0,1) & \text{if } a \text{ is even, } b \text{ is odd.} \end{cases}$$

It is now straightforward to verify (6.2). □

Using the previous lemma, we can now easily solve the problem for all predicates $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ that change value reasonably close to 0. Extension to the general case will require an additional step.

Theorem 6.2. *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Suppose that $D(\ell) \neq D(\ell - 1)$ for some $\ell \leqslant \frac{1}{8}n$. Then*

$$Q^*_{1/3}(D) \geqslant \Omega\left(\sqrt{n\ell}\right).$$

Proof. It suffices to show that $Q^*_{1/5}(D) \geqslant \Omega\left(\sqrt{n\ell}\right)$. Define $f : \{0, 1\}^{\lfloor n/4 \rfloor} \to \{0, 1\}$ by $f(z) = D(|z|)$. Then $\ell_0(f) \geqslant \ell$ since $\ell \leqslant \frac{1}{8}n$. As a result,

$$\deg_{1/3}(f) \geqslant \Omega\left(\sqrt{n\ell}\right)$$

by Paturi's lower bound (Theorem 2.8). Now Theorem 5.1 implies that

$$Q^*_{1/5}(F) \geqslant \Omega\left(\sqrt{n\ell}\right),$$

where F is the $(2\lfloor n/4 \rfloor, \lfloor n/4 \rfloor, f)$-pattern matrix. But Lemma 6.1 states that F is a submatrix of the communication matrix of D, namely, $[D(|x \wedge y|)]_{x,y}$. It follows that $Q^*_{1/5}(D) \geqslant \Omega\left(\sqrt{n\ell}\right)$. □

We have proved the desired lower bounds for all predicates D that change value close to 0. What remains is to extend the result to arbitrary predicates, which is going to be a simple if tedious exercise in shifting and padding. We note that Razborov's proof concludes in a similar way (see [18], beginning of Section 5).

Theorem 6.3. *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Suppose that $D(\ell) \neq D(\ell - 1)$ for some $\ell > \frac{1}{8}n$. Then*

$$Q^*_{1/3}(D) \geqslant c(n - \ell) \tag{6.3}$$

for some absolute constant $c > 0$.

Proof. Consider the communication problem of computing $D(|x \wedge y|)$ when the last k bits in x and y are fixed to 1. In other words, the new problem is to compute $D_k(|x' \wedge y'|)$, where $x', y' \in \{0, 1\}^{n-k}$ and the predicate $D_k : \{0, 1, \ldots, n-k\} \to \{0, 1\}$ is given by

$$D_k(i) \equiv D(k + i).$$

Since the new problem is a restricted version of the original, we have

$$Q^*_{1/3}(D) \geqslant Q^*_{1/3}(D_k) \qquad \text{for all } k. \tag{6.4}$$

We complete the proof by placing a lower bound on $Q^*_{1/3}(D_k)$ for some k.

Put $\alpha \stackrel{\text{def}}{=} \frac{1}{8}$. The quantity

$$k_0 \stackrel{\text{def}}{=} \ell - \left\lfloor \frac{\alpha}{1-\alpha} \cdot (n-\ell) \right\rfloor$$

is an integer between 1 and ℓ (because $\ell > \alpha n$). The equality $k_0 = \ell$ occurs if and only if $\left\lfloor \frac{\alpha}{1-\alpha}(n-\ell) \right\rfloor = 0$, in which case the claimed conclusion (6.3) holds trivially for c suitably small, such as $c = \alpha/(1-\alpha)$. Thus, we can assume that $1 \leqslant k_0 \leqslant \ell - 1$, in which case $D_{k_0}(\ell - k_0) \neq D_{k_0}(\ell - k_0 - 1)$ and $\ell - k_0 \leqslant \alpha(n - k_0)$. Therefore, Theorem 6.2 is applicable to D_{k_0} and yields:

$$Q^*_{1/3}(D_{k_0}) \geqslant C\sqrt{(n-k_0)(\ell-k_0)}, \tag{6.5}$$

where $C > 0$ is an absolute constant. Calculations reveal:

$$n - k_0 = \left\lfloor \frac{1}{1-\alpha} \cdot (n-\ell) \right\rfloor, \qquad \ell - k_0 = \left\lfloor \frac{\alpha}{1-\alpha} \cdot (n-\ell) \right\rfloor. \tag{6.6}$$

The theorem is now immediate from (6.4)–(6.6). □

Together, Theorems 6.2 and 6.3 give the main result of this section:

Theorem 1.1 (Restated from p. 2). *Let* $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. *Then*

$$Q_{1/3}(D) \;\geqslant\; Q^*_{1/3}(D) \;\geqslant\; \Omega\left(\sqrt{n\ell_0(D)} + \ell_1(D)\right).$$

Proof. If $\ell_0(D) \neq 0$, set $\ell \stackrel{\text{def}}{=} \ell_0(D)$ and note that $D(\ell) \neq D(\ell - 1)$ by definition. One of Theorems 6.2 and 6.3 must be applicable, and therefore $Q^*_{1/3}(D) \geqslant \min\left\{\Omega(\sqrt{n\ell}),\ \Omega(n-\ell)\right\}$. Since $\ell \leqslant n/2$, this simplifies to

$$Q^*_{1/3}(D) \geqslant \Omega\left(\sqrt{n\ell_0(D)}\right). \tag{6.7}$$

If $\ell_1(D) \neq 0$, set $\ell \stackrel{\text{def}}{=} n - \ell_1(D) + 1 \geqslant n/2$ and note that $D(\ell) \neq D(\ell - 1)$ as before. By Theorem 6.3,

$$Q^*_{1/3}(D) \geqslant \Omega\left(\ell_1(D)\right). \tag{6.8}$$

The theorem follows from (6.7) and (6.8). □

7 Additional Results

As we have already stated, the concept of a pattern matrix originates in the author's earlier article [22], where a somewhat different family of matrices is introduced and its spectral norm studied. As a result of that study, we constructed the first AC^0 circuit with discrepancy $2^{-\Omega(n^{1/5})}$ and thereby separated AC^0 from depth-2 majority circuits [22, Thms. 1.1–1.3]. The spectral norm calculation in that work is not exact; only a suitable upper bound is obtained. This manuscript, on the other hand, derives an exact, closed-form expression for the singular values of a pattern matrix. As a consequence, we are able to considerably improve the results in [22].

Namely, we improve the discrepancy upper bound for AC^0 from $2^{-\Omega(n^{1/5})}$ to $2^{-\Omega(n^{1/3})}$, with corresponding circuit implications. This is currently the best upper bound on the discrepancy of a function in AC^0. It matches the result of Buhrman et al. [4] who, independently and simultaneously with the author's work [22], proved a $2^{-\Omega(n^{1/3})}$ upper bound with different techniques and for a different function. In addition, we strengthen the *Degree/Discrepancy Theorem* from [22], needed to arrive at the discrepancy result and independently interesting.

7.1 Background and Definitions

We start with a few definitions. Throughout this section, it will be convenient to view Boolean functions as mappings into $\{-1, +1\}$, as opposed to the usual range $\{0, 1\}$. Fix finite sets X, Y and let $f : X \times Y \to \{-1, +1\}$ be given. Let λ be a probability distribution over $X \times Y$. The *discrepancy* of f under λ is defined by

$$\operatorname{disc}_\lambda(f) \stackrel{\text{def}}{=} \max_{\substack{S \subseteq X, \\ T \subseteq Y}} \left| \sum_{x \in S} \sum_{y \in T} \lambda(x, y) f(x, y) \right|.$$

Define

$$\operatorname{disc}(f) \stackrel{\text{def}}{=} \min_{\lambda} \{\operatorname{disc}_\lambda(f)\}.$$

We identify a function $f : X \times Y \to \{-1, +1\}$ with its communication matrix $F = [f(x, y)]_{x,y}$. In particular, we follow the conventions $\operatorname{disc}_\lambda(F) = \operatorname{disc}_\lambda(f)$ and $\operatorname{disc}(F) = \operatorname{disc}(f)$.

The above definition of discrepancy is not convenient to work with, and we will use the following well-known matrix-analytic reformulation; cf. Kushilevitz & Nisan [14, Example 3.29]. For matrices $A = [A_{xy}]$ and $B = [B_{xy}]$, recall that $A \circ B \stackrel{\text{def}}{=} [A_{xy} B_{xy}]$ is their Hadamard product.

Proposition 7.1. *Let X, Y be finite sets, $f : X \times Y \to \{-1, +1\}$. Then*

$$\operatorname{disc}(f) \leqslant \sqrt{|X|\,|Y|} \min_K \|K \circ F\|,$$

where $F \stackrel{\text{def}}{=} [f(x,y)]_{x\in X,\, y\in Y}$ *and the minimum is over matrices* K *whose entries are nonnegative and sum to* 1.

Proof. Fix K and define a distribution λ on $X \times Y$ by $\lambda(x,y) = K_{xy}$. Then

$$\begin{aligned}\operatorname{disc}_\lambda(f) &= \max_{S,T} \left|\mathbf{1}_S^{\mathsf{T}} (K \circ F)\, \mathbf{1}_T\right| \leqslant \max_{S,T} \left\{\|\mathbf{1}_S\| \cdot \|K \circ F\| \cdot \|\mathbf{1}_T\|\right\} \\ &= \|K \circ F\| \sqrt{|X|\,|Y|}. \end{aligned}$$

□

For a function $f : \{0,1\}^n \to \{-1,+1\}$, its *threshold degree* $\deg(f)$ is the least degree of a multivariate polynomial $p(x_1, \ldots, x_n)$ with

$$p(x)f(x) > 0 \qquad \text{for all } x \in \{0,1\}^n.$$

The following well-known result follows from Gordan's Transposition Theorem [21, Sec. 7.8]; for a detailed proof, see [22].

Theorem 7.2 (Criterion for high threshold degree). *Let* $f : \{0,1\}^n \to \{-1,+1\}$ *be given. Then* $\deg(f) \geqslant d$ *if and only if there is a distribution* μ *over* $\{0,1\}^n$ *with*

$$\mathop{\mathbf{E}}_{x\sim\mu} [f(x)\, \chi_S(x)] = 0 \qquad \textit{whenever } |S| < d.$$

7.2 New Results

We are prepared to state the first improvement on [22].

Theorem 7.3 (Degree/Discrepancy Theorem; cf. Sherstov [22, Thm. 1.2]). *Let* F *be the* (n,t,f)*-pattern matrix, where* $f : \{0,1\}^t \to \{-1,+1\}$ *has* $\deg(f) = d$. *Then*

$$\operatorname{disc}(F) \leqslant \left(\frac{t}{n}\right)^{d/2}.$$

Proof. By Theorem 7.2, there is a probability distribution μ over $\{0,1\}^t$ such that $\mathbf{E}_{z\sim\mu}[f(z)\, \chi_S(z)] = 0$ for $|S| < d$. Putting $\phi(z) \stackrel{\text{def}}{=} \mu(z) f(z)$, we obtain:

$$\hat{\phi}(S) = 0 \qquad \text{for } |S| < d. \tag{7.1}$$

Furthermore, Proposition 2.1 reveals that

$$\max_{S\subseteq[t]} |\hat{\phi}(S)| \leqslant \frac{1}{2^t}. \tag{7.2}$$

In view (7.1) and (7.2), Theorem 4.3 implies that

$$\|A\| \leqslant \left(\frac{t}{n}\right)^{d/2} \left(2^{n+t}\left(\frac{n}{t}\right)^{t}\right)^{-1/2},$$

where A is the $(n, t, 2^{-n}(n/t)^{-t}\phi)$-pattern matrix. But $A = K \circ F$, where K is the $(n, t, 2^{-n}(n/t)^{-t}\mu)$-pattern matrix. Since the entries of K are nonnegative and sum to 1, Proposition 7.1 implies that

$$\operatorname{disc}(F) \leqslant \|A\| \sqrt{2^{n+t}\left(\frac{n}{t}\right)^{t}} \leqslant \left(\frac{t}{n}\right)^{d/2}. \qquad \square$$

Theorem 7.3 states that pattern matrices generated by functions with high threshold degree have low discrepancy. A well-known function in AC^0 with high threshold degree is the *Minsky-Papert function,* given by

$$\mathrm{MP}(x) \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} x_{i,j}.$$

(This is the same function we used in [22].) We have:

Theorem 7.4 (Minsky-Papert [16]). *The function* MP *on* $4m^3$ *variables has* $\deg(\mathrm{MP}) = m$.

As an application of our strengthened Degree/Discrepancy Theorem, we obtain an improved upper bound on the discrepancy of AC^0.

Theorem 7.5 (Discrepancy of $\mathsf{AC}^{0,3}$; cf. Sherstov [22, Thm. 1.3]). *There is an (explicitly given) AND/OR/NOT circuit* $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ *of depth* 3 *and size* $2n$ *such that*

$$\operatorname{disc}_\lambda(f) \leqslant 2^{-\Omega(n^{1/3})}$$

for an explicitly given distribution λ.

Proof. Put $n = 16m^3$ and define

$$f(x,y) \stackrel{\text{def}}{=} \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} \Big((y_{i,j,1} \vee x_{i,j,1}) \wedge (y_{i,j,2} \vee \neg x_{i,j,1}) \wedge (y_{i,j,3} \vee x_{i,j,2}) \wedge (y_{i,j,4} \vee \neg x_{i,j,2})\Big).$$

It is straightforward to verify that the $(8m^3, 4m^3, \mathrm{MP})$-pattern matrix is a submatrix of $[f(x,y)]_{x,y}$. The discrepancy result now follows by Theorems 7.3 and 7.4.

To specify the corresponding distribution explicitly, it suffices to specify the distribution μ on $\{0,1\}^{4m^3}$ with $\mathbf{E}_{z\sim\mu}[\mathrm{MP}(z)\,\chi_S(z)] = 0$ for $|S| < m$. This is because once μ is known, the distribution over the entries of $F = [f(x,y)]$ can be easily reconstructed from the proof of Theorem 7.3. The author has already constructed just such a distribution μ in earlier work [22, Thm. 4.2]. $\square$

This discrepancy upper bound has the following circuit implications.

Theorem 7.6 (cf. Sherstov [22, Thm. 1.3]). *There is an (explicitly given) AND/OR/NOT circuit $f : \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ of depth 3 and size $2n$ such that any majority vote of threshold gates that computes f has size $2^{\Omega(n^{1/3})}$.*

Proof. Identical to the proof given in [22]. □

As a final remark, we note that our improved Degree/Discrepancy Theorem can be used to strengthen the results of Chattopadhyay [5], who extended the author's original work [22] to the multi-party model. However, the necessary manipulations would not offer much technical novelty, and we omit them.

References

[1] S. Aaronson and Y. Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004.

[2] A. Ambainis, L. J. Schulman, A. Ta-Shma, U. V. Vazirani, and A. Wigderson. The quantum communication complexity of sampling. *SIAM J. Comput.*, 32(6):1570–1585, 2003.

[3] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proc. of the 16th Conf. on Computational Complexity (CCC)*, pages 120–130, 2001.

[4] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proc. of the 22nd Conf. on Computational Complexity (CCC)*, pages 24–32, 2007.

[5] A. Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *Proc. of the 48th Symposium on Foundations of Computer Science (FOCS)*, 2007. To appear.

[6] D. Gavinsky, J. Kempe, and R. de Wolf. Strengths and weaknesses of quantum fingerprinting. In *Proc. of the 21st Conf. on Computational Complexity (CCC)*, pages 288–298, 2006.

[7] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge University Press, New York, 1986.

[8] P. Høyer and R. de Wolf. Improved quantum communication complexity bounds for disjointness and equality. In *Proc. of the 19th Annual Symposium on Theoretical Aspects of Computer Science (STACS)*, pages 299–310, 2002.

[9] H. Klauck. Lower bounds for quantum communication complexity. In *Proc. of the 42nd Symposium on Foundations of Computer Science (FOCS)*, pages 288–297, 2001.

[10] H. Klauck, A. Nayak, A. Ta-Shma, and D. Zuckerman. Interaction in quantum communication and the complexity of set disjointness. In *Proc. of the 33rd Symposium on Theory of Computing (STOC)*, pages 124–133, 2001.

[11] H. Klauck, R. Spalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. *SIAM J. Comput.*, 36(5):1472–1493, 2007.

[12] A. R. Klivans and A. A. Sherstov. A lower bound for agnostically learning disjunctions. In *Proc. of the 20th Conf. on Learning Theory (COLT)*, pages 409–423, 2007.

[13] I. Kremer. Quantum communication. Master's thesis, Hebrew University, Computer Science Department, 1995.

[14] E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, New York, 1997.

[15] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proc. of the 39th Symposium on Theory of Computing (STOC)*, pages 699–708, 2007.

[16] M. L. Minsky and S. A. Papert. *Perceptrons: expanded edition*. MIT Press, Cambridge, MA, USA, 1988.

[17] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions. In *Proc. of the 24th Symposium on Theory of Computing*, pages 468–474, 1992.

[18] A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[19] A. A. Razborov. Personal communication, June 2007.

[20] T. J. Rivlin. *An Introduction to the Approximation of Functions*. Dover Publications, New York, 1981.

[21] A. Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, Inc., New York, 1998.

[22] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proc. of the 39th Symposium on Theory of Computing (STOC)*, pages 294–301, 2007.