A Discrepancy-Based Proof of Razborov's Quantum Lower Bounds

Alexander A. Sherstov

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

July 3, 2007

Abstract. In a breakthrough result, Razborov (2003) gave optimal lower bounds on the quantum communication complexity $Q^*_{1/3}(f)$ of every function $f(x, y) = D(|x \wedge y|)$, where $D : \{0, 1, \dots, n\} \to \{0, 1\}$. Namely, he showed that

$$Q^*_{1/3}(f) \ = \ \Omega\Big(\ell_1(D) \ + \ \sqrt{n\,\ell_0(D)}\Big),$$

where $\ell_0(D)$, $\ell_1(D) \ \in \ \{0, 1, \dots, \lceil n/2 \rceil\}$ are the smallest integers such that D is constant in the range $[\ell_0(D), n - \ell_1(D)]$. This was proved by the so-called *multidimensional* discrepancy method. We study this problem afresh, using the original, *one-dimensional* discrepancy method. We show:

$$Q^*_{1/3}(f) \ = \ \Omega\Big(\ell_1(D) \ + \ n^{1/3}\,\ell_0(D)^{2/3}\Big).$$

Thus, our lower bound for each f ranges between $\mathsf{opt}^{2/3}$ and opt, where opt is the true answer found by Razborov. Prior to this work, Razborov (2003) conjectured that the original discrepancy method could not yield nontrivial lower bounds for the problem. In addition, our technique gives strong lower bounds for a rather broad class of functions which we call *pattern functions* and for which no methods were previously known.

Our proof technique has two ingredients. The first is a certain equivalence of *approximation* and *orthogonality* in a Euclidean space (the *Approximation/Orthogonality Principle*), which we establish using duality theory. The second is the author's recent construction (Sherstov 2007) of matrices with low spectral norm, which we originally used to separate AC^0 from depth-2 majority circuits.

1 Introduction

Let $D : \{0,1,\ldots,n\} \to \{0,1\}$ be an arbitrary predicate. We study the quantum communication complexity of the associated function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ given by

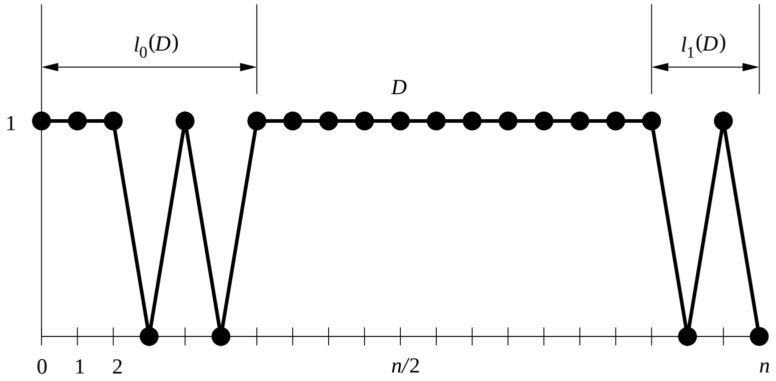
$$f(x,y) \overset{\text{def}}{=} D(|x \wedge y|),$$

where $|x \wedge y|$ stands for the number of positions where x and y both have a 1. As usual, the communication problem 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. In a breakthrough result, Razborov [16] established optimal lower bounds on the quantum communication complexity of every function of the form $f(x,y) = D(|x \wedge y|)$:

Theorem 1.1 (Razborov [16]). *Let $D : \{0,1,\ldots,n\} \to \{0,1\}$ be an arbitrary predicate. Put $f(x,y) \overset{\text{def}}{=} D(|x \wedge y|)$. Then*

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

We look at this problem afresh, with new techniques. We prove:

Theorem 1.2. *Let $D : \{0,1,\dots,n\} \to \{0,1\}$ be an arbitrary predicate. Put $f(x,y) \overset{\text{def}}{=} D(|x \wedge y|)$. Then*

$$Q^*_{1/3}(f) \geqslant \Omega\left(\ell_1(D) + n^{1/3}\,\ell_0(D)^{2/3}\right).$$

Clearly, both theorems remain valid in the model without prior entanglement. A quick glance shows that there are predicates for which we do *not* match Razborov's optimal bounds. However, the gap is not large: our lower bound for each f is at least $\mathsf{opt}^{2/3}$, where opt is the true answer found by Razborov. Theorem 1.2 subsumes *all* previous work except for Razborov's. For some functions, we give an exponential improvement over that previous work, such as $\Omega(n^{1/3})$ vs. $\Omega(\log n)$ for the *disjointness* predicate D defined by $D(i) = 1 \Leftrightarrow i = 0$.

Motivation. Our work is of interest for three reasons. First, Razborov's result is one of the strongest lower bounds for bounded-error communication and deserves several proofs. Despite considerable work by various authors [1, 2, 6–8, 13], no such alternate proof exists.

Second, our proof contributes a new technique for communication lower bounds, much different from Razborov's methods. We outline it below under "Techniques." The high-level strategy of our proof is known as the *discrepancy method.* This in itself is surprising because prior to this paper, it was believed that the discrepancy method could not be advantageously applied to this problem. In particular, Razborov writes [16, p. 155] that the discrepancy method cannot yield strong lower bounds for functions such as disjointness. In view of our results, that statement was in error [17]. We emphasize that it does not affect the remainder of Razborov's article, which is correct.

Third, our technique yields strong lower bounds for a rather broad family of functions which we call *pattern functions* and for which no other methods are currently known. To keep this section brief and nontechnical, we defer the definition of these functions and our lower bounds against them to Section 7, "Extensions." An intuitive explanation is as follows. Critical to Razborov's technique is the high symmetry of the functions $f(x,y) = D(|x \wedge y|)$. The method in this work, on the other hand, makes limited use of the symmetry and thus applies unchanged to a broader range of problems.

Razborov's proof technique has seen several applications, including direct-product theorems [9], separations for small-bias communication [3], and learning theory [10]. We hope that the ideas in this paper will also find uses beyond quantum communication.

Our Techniques. Given a function $f(x,y) = D(|x \wedge y|)$, Razborov places a lower bound on $Q^*_{1/3}(f)$ using his *multidimensional* discrepancy method, a powerful extension of the original discrepancy method. Razborov's technique exploits the so-called combinatorial matrices, which have rare and useful spectral properties.

By contrast, we work with the simple, original discrepancy method. Roughly speaking, this method says: find a "hard" function h in the vicinity of f, thereby proving that f itself must be "somewhat hard." More precisely, the discrepancy method asks for a function h and a distribution μ 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 actually exists, it easily 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. Its critical part is a certain equivalence of *approximation* and *orthogonality* in the Euclidean space (the *Approximation/Orthogonality Principle*), which we prove using duality theory. The other key ingredient is the author's recent construction of matrices with low spectral norm [19], originally employed in separating AC^0 from depth-2 majority circuits.

Organization. We start with a thorough review of technical preliminaries in Section 2. The two sections that follow are concerned with the development of our technique. Section 5 integrates it into the discrepancy method. Section 6 consolidates the proof and gives the final bound on communication. Section 7 concludes with generalizations of our technique.

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 $D : \{0,1,\dots,n\} \to \{0,1\}$. The notation $[n]$ stands for the set $\{1,2,\dots,n\}$. For a set $S \subseteq [n]$, its *characteristic vector* $1_S \in \{0,1\}^n$ is defined by

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

For $x \in \{0,1\}^n$, put $|x| \overset{\text{def}}{=} |\{i : x_i = 1\}|$. 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.

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

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

For $S \subseteq [n]$, define $\chi_S : \{0,1\}^n \to \{-1,+1\}$ by

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

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 *Fourier representation*

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

where $\hat{f}(S) \overset{\text{def}}{=} \langle f, \chi_S \rangle$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of f*. The orthonormality of $\{\chi_S\}$ immediately yields *Parseval's identity*:

$$\sum_{S \subseteq [n]} \hat{f}(S)^2 = \langle f, f \rangle = \mathop{\mathbf{E}}_x [f(x)^2].$$

2.2 Matrix Analysis

We draw freely on basic notions from matrix analysis. A standard reference on the subject is [5]. 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 adopt the following standard notation:

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

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 $\langle A, B\rangle \overset{\text{def}}{=} \sum_{i,j} A_{ij}B_{ij}$. A useful consequence of the singular value decomposition is:

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

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 2.1 (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, since our proofs rely solely on a 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 [16]. Let $\mathscr{A}$ and $\mathscr{B}$ be complex finite-dimensional Hilbert spaces. Let $\mathscr{C}$ be a Hilbert space of dimension 2, whose orthonormal basis we denote by $|0\rangle,\ |1\rangle$. Consider the tensor product $\mathscr{A} \otimes \mathscr{C} \otimes \mathscr{B}$, which is itself a Hilbert space with an inner product inherited from $\mathscr{A}$, $\mathscr{B}$, and $\mathscr{C}$. The *state* of a quantum system is a unit vector in $\mathscr{A} \otimes \mathscr{C} \otimes \mathscr{B}$, 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

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

where: $I_{\mathscr{A}}$ and $I_{\mathscr{B}}$ are the identity transformations in $\mathscr{A}$ and $\mathscr{B}$, respectively; $U_1, U_3, \ldots, U_{2k-1}$ are unitary transformations in $\mathscr{A} \otimes \mathscr{C}$; and $U_2, U_4, \ldots, U_{2k}$ are unitary transformations in $\mathscr{C} \otimes \mathscr{B}$. 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 some 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

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

3. Let v denote the projection of $\mathsf{Final}(x, y)$ onto $\mathscr{A} \otimes \operatorname{span}(|1\rangle) \otimes \mathscr{B}$. 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 $\mathsf{Initial}(x, y) \in \mathscr{A} \otimes \mathscr{C} \otimes \mathscr{B}$ 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, $\mathscr{A}$ and $\mathscr{B}$ 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 space 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 $\mathscr{A}$ and $\mathscr{B}$ 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 cost of the optimal 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

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

We let $Q_\epsilon(D) \overset{\text{def}}{=} Q_\epsilon(f)$ and $Q^*_\epsilon(D) \overset{\text{def}}{=} Q^*_\epsilon(f)$. 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 2.2 (Razborov [16, Thm. 5.5], Linial & Shraibman [13, 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 there are real matrices A, B such that $\|A\|_{\mathrm{F}} \leqslant 2^C \sqrt{|X|}$, $\|B\|_{\mathrm{F}} \leqslant 2^C \sqrt{|Y|}$, and*

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

Theorem 2.2 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 [4, 7, 11, 13]. We now state the discrepancy method as adapted to the quantum model by Razborov [16]. This is not to be confused with the *multidimensional discrepancy method,* also due to Razborov [16], which we will have no occasion to use or describe.

Theorem 2.3 (Discrepancy method, Razborov [16, 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 $M \overset{\mathrm{def}}{=} \Big[(-1)^{f(x,y)}\Big]_{x\in X,\, y\in Y}$.

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

$$\Pi \overset{\mathrm{def}}{=} \Big[\mathbf{E}[P(x,y)]\Big]_{x\in X,\, y\in Y}.$$

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,

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

On the other hand, Theorem 2.2 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,

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

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

We now reinterpret Theorem 2.3 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 [16]. Some authors, including Kushilevitz and Nisan [12], restrict the term "discrepancy method" to the case when $f = h$ and the communication takes place in the classical randomized model. This restriction 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 commutation 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 saw in the proof of Theorem 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

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

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

2.5 Approximation by Polynomials

Let $f : \{0,1\}^n \to \mathbb{R}$. 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 2.4 (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, $0 \leqslant d \leqslant n$, 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, $\deg_\epsilon(f)$ is not substantially affected by the choice of a constant $\epsilon \in (0, 1/2)$. More precisely, we have:

Proposition 2.5 (Folklore). *Let $f : \{0,1\}^n \to \{0,1\}$ be arbitrary, ϵ a constant with $0 < \epsilon < 1/2$. Then*

$$\deg_\epsilon(f) = \Theta(\deg_{1/3}(f)).$$

Proof (folklore). Assume that $\epsilon \leqslant 1/3$; the case $\epsilon \in (1/3, 1/2)$ has a closely analogous proof, and we omit it. Let $d \overset{\text{def}}{=} \deg_{1/3}(f)$. We have to show that $\deg_\epsilon(f) = O(d)$. For this, fix $\phi \in \operatorname{span}(\{\chi_S\}_{|S| \leqslant d})$ with $\max_{x \in \{0,1\}^n} |f(x) - \phi(x)| \leqslant 1/3$. By basic approximation theory (see Rivlin [18, Cor. 1.4.1]), there exists a univariate polynomial p of degree $O(1/\epsilon)$ with

$$p([-1/3, 1/3]) \subseteq [-\epsilon, \epsilon], \qquad p([1 - 1/3, 1 + 1/3]) \subseteq [1 - \epsilon, 1 + \epsilon].$$

Then clearly $p(\phi(x))$ is the sought approximator of f. □

In view of Proposition 2.5, it is standard practice to work with $\deg_{1/3}(f)$ by default. Another easy observation is that 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

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

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

Determining $\deg_{1/3}(f)$ for a given Boolean function f can be difficult. There is, however, a family of Boolean functions whose approximate degree is analytically manageable. This is the family of *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

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

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

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

To carry our discussion further, we extend the notion of approximation to predicates.

Definition 2.6 (Approximate degree of predicates). For a predicate $D:\{0,1,\dots,n\}\to\{0,1\}$, define its *$\epsilon$-approximate degree* $\deg_\epsilon(D)$ to be the minimum degree of a univariate real polynomial p with

$$\max_{i=0,1,\dots,n}|D(i)-p(i)|\leqslant\epsilon.$$

Analyzing the approximate degree of predicates is a much simpler task and, indeed, a basic question in approximation theory. It is therefore fortunate that the ϵ-approximate degree of a symmetric function is the same as the ϵ-approximate degree of its associated predicate. This equivalence is known as the *symmetrization argument* of Minsky and Papert [14].

Proposition 2.7 (Symmetrization argument; Minsky and Papert [14]). *Let $f:\{0,1\}^n\to\{0,1\}$ be a symmetric Boolean function. Let D be the predicate with $f(x)\equiv D(x_1+\cdots+x_n)$. Then*

$$\deg_\epsilon(f)=\deg_\epsilon(D) \qquad\qquad \textit{for all } \epsilon\geqslant 0.$$

Proof sketch (Minsky and Papert [14]). Since $f(x) = D(x_1 + \cdots + x_n)$, it is clear that $\deg_\epsilon(f) \leqslant \deg_\epsilon(D)$. For the reverse direction, let $d \overset{\text{def}}{=} \deg_\epsilon(f)$ and fix $\phi = \sum_{|S| \leqslant d} a_S \chi_S$ with $\max_{x \in \{0,1\}^n} |f(x) - \phi(x)| \leqslant \epsilon$. Consider the function

$$\phi'(x) \overset{\text{def}}{=} \frac{1}{n!} \sum_{\sigma \in S_n} \phi(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}).$$

On the one hand, $\max_{x \in \{0,1\}^n} |f(x) - \phi'(x)| \leqslant \epsilon$. On the other hand, one can show that

$$\phi'(x) = p(x_1 + \cdots + x_n)$$

for some polynomial p of degree d. Thus, p approximates D pointwise within ϵ, and $\deg_\epsilon(D) \leqslant d = \deg_\epsilon(f)$. □

Using Proposition 2.7 and tools from approximation theory, Paturi [15] gave an asymptotically tight estimate of $\deg_{1/3}(f)$ for every symmetric Boolean function f. The estimates are in terms of the quantities $\ell_0(f)$ and $\ell_1(f)$, defined next.

Definition 2.8 (Razborov [16]). 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 2.9 (Paturi [15]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a symmetric function. Then*

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

In words, Theorem 2.9 states that the $\frac{1}{3}$-approximate degree is $\Omega(\sqrt{n})$ for every nonconstant predicate, and is higher for those predicates that change value near the middle of the range $\{0, 1, \ldots, n\}$.

3 The Approximation/Orthogonality Principle

This section marks the beginning of our proof. Here we consider the notions of *approximation* and *orthogonality* in a Euclidean space and establish a certain equivalence between them. We will later reinterpret this result in terms of protocols rather than points in a 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 span(Φ) has finite dimension, a best approximation to f out of span(Φ) always exists [18, 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:

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

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 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. Let $\phi_1, \ldots, \phi_k : X \to \mathbb{R}$ be a basis for span(Φ). Our first observation is that $\epsilon^*(f, \Phi)$ is the optimum of the following linear program in the variables $\epsilon, \alpha_1, \ldots, \alpha_k$:

$$\boxed{\begin{array}{lrl}
\text{minimize:} & \epsilon & \\
\text{subject to:} & & \\
& \left| f(x) - \sum_{i=1}^{k} \alpha_i \phi_i(x) \right| \leqslant \epsilon & \text{for each } x \in X, \\
& \alpha_i \in \mathbb{R} & \text{for each } i, \\
& \epsilon \geqslant 0. &
\end{array}}$$

Standard manipulations reveal the dual:

$$\boxed{\begin{array}{lrl}
\text{maximize:} & \sum_{x \in X} \beta_x f(x) & \\
\text{subject to:} & & \\
& \sum_{x \in X} |\beta_x| \leqslant 1, & \\
& \sum_{x \in X} \beta_x \phi_i(x) = 0 & \text{for each } i, \\
& \beta_x \in \mathbb{R} & \text{for each } x \in X.
\end{array}}$$

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 $\phi_1, \phi_2, \ldots, \phi_k$ 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. *Let $f : \{0,1\}^t \to \mathbb{R}$ have $\deg_{1/3}(f) \geqslant d$. Then there is a function $\psi : \{0,1\}^t \to \mathbb{R}$ such that:*

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

$$\sum_{z\in\{0,1\}^t} \psi(z)\chi_S(z) = 0 \qquad \textit{for all } S \textit{ with } |S| < d,$$

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

Proof. Set $X = \{0,1\}^t$ and $\Phi = \{\chi_S : |S| < d\} \subset \mathbb{R}^X$. Since $\deg_{1/3}(f) \geqslant d$, we immediately have that $\epsilon^*(f,\Phi) > 1/3$. But then $\gamma^*(f,\Phi) > 1/3$ 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 place an upper bound on their spectral norm. As we shall see later, this provides a convenient means to generate hard communication problems.

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

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

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

$$x|_V \overset{\text{def}}{=} (x_{i_1}, x_{i_2}, \ldots, x_{i_{|V|}}) \in \{0,1\}^{|V|},$$

where $i_1 < i_2 < \cdots < i_{|V|}$ are the elements of V.

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

$$A = \Big[\phi(x|_V)\Big]_{x\in\{0,1\}^m,\, V\in\mathscr{V}(m,t)}.$$

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.

We are now ready to analyze the spectral norm of a pattern matrix. The theorem we are about to prove refines an earlier result due to the author [19, Thm. 1.2].

Theorem 4.2 (Spectrum of pattern matrices). *Let $\phi : \{0,1\}^t \to \mathbb{R}$ be any function with $\hat{\phi}(S) = 0$ for all S with $|S| < d$, where d is a given integer. Let A be the (m,t,ϕ)-pattern matrix. If $m \geqslant 5\mathrm{e}t^2/d$, then*

$$\|A\| \leqslant \frac{1}{2^d} \sqrt{2^m \left(\frac{m}{t}\right)^t} \left(\frac{1}{2^t} \sum_{z \in \{0,1\}^t} |\phi(z)| \right). \tag{4.1}$$

Proof. The assumption about the Fourier spectrum of ϕ can be restated as follows:

$$\phi(z) = \sum_{S \subseteq [t],\ |S| \geqslant d} \hat{\phi}(S) \chi_S(z) \qquad (z \in \{0,1\}^t). \tag{4.2}$$

Define the pattern matrix

$$A_S \stackrel{\text{def}}{=} [\chi_S(x|_V)]_{x,V} \qquad (S \subseteq [t]). \tag{4.3}$$

Using (4.2) and (4.3),

$$A = \sum_{S \subseteq [t],\ |S| \geqslant d} \hat{\phi}(S) A_S.$$

The orthogonality of the χ_S implies that $A_S^{\mathsf{T}} A_T = 0$ for all $S \neq T$. As a result,

$$A^{\mathsf{T}} A = \sum_{S \subseteq [t],\ |S| \geqslant d} \hat{\phi}(S)^2 A_S^{\mathsf{T}} A_S. \tag{4.4}$$

It is not hard to see that $A_S^{\mathsf{T}} A_S$ is permutation-similar to the matrix

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

where J is the all-ones square matrix of order $(m/t)^{t-|S|}$. Therefore,

$$\|A_S^{\mathsf{T}} A_S\| = 2^m \left(\frac{m}{t}\right)^{t-|S|}. \tag{4.5}$$

We need one final observation: for each $S \subseteq [t]$,

$$|\hat{\phi}(S)| = \left|\mathop{\mathbf{E}}_{z}[\phi(z)\chi_S(z)]\right| \leqslant \mathop{\mathbf{E}}_{z}[|\phi(z)|] = \frac{1}{2^t}\sum_{z\in\{0,1\}^t}|\phi(z)|. \tag{4.6}$$

It remains to combine the above ingredients:

$$\begin{aligned}
\|A\|^2 &= \|A^{\mathsf{T}}A\| \\
&\leqslant \sum_{S\subseteq[t],\ |S|\geqslant d} \hat{\phi}(S)^2\,\|A_S^{\mathsf{T}}A_S\| && \text{by (4.4)} \\
&\leqslant 2^m\left(\frac{m}{t}\right)^t\left(\frac{1}{2^t}\sum_{z\in\{0,1\}^t}|\phi(z)|\right)^2\sum_{S\subseteq[t],\ |S|\geqslant d}\left(\frac{t}{m}\right)^{|S|} && \text{by (4.5), (4.6)} \\
&= 2^m\left(\frac{m}{t}\right)^t\left(\frac{1}{2^t}\sum_{z\in\{0,1\}^t}|\phi(z)|\right)^2\sum_{i=d}^{t}\binom{t}{i}\left(\frac{t}{m}\right)^i.
\end{aligned}$$

In particular, (4.1) holds as soon as $m \geqslant 5\mathrm{e}t^2/d$. □

5 Application of the Discrepancy 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.

For a predicate $D : \{0,1,\dots,n\} \to \{0,1\}$ and an integer t, $1 \leqslant t \leqslant n$, define a new predicate $D|_t : \{0,1,\dots,t\} \to \{0,1\}$ by

$$D|_t(i) = D(i) \qquad (i = 0,1,\dots,t).$$

In words, $D|_t$ is the restriction of D to $\{0,1,\dots,t\}$. We start with a technical lemma.

Lemma 5.1. *Let $D : \{0,1,\dots,n\} \to \{0,1\}$. Suppose that $D(\ell) \neq D(\ell-1)$ for some $\ell \leqslant \alpha n$, where $\alpha \in (0,1)$ is a suitably small absolute constant. Then there is an integer t, $1 \leqslant t \leqslant n$, such that*

$$\deg_{1/3}(D|_t) = \Omega(n^{1/3}\ell^{2/3}) \qquad \text{and} \qquad \frac{n}{2} \geqslant \frac{5\mathrm{e}t^2}{\deg_{1/3}(D|_t)}. \tag{5.1}$$

Proof. Let t be any integer with $2\ell \leqslant t \leqslant n$. Then $\ell_0(D|_t) \geqslant \ell$ by definition, and

$$\deg_{1/3}(D|_t) \geqslant \beta\sqrt{t\ell}$$

by Theorem 2.9, where $\beta \in (0,1)$ is a certain absolute constant. Therefore, the proof will be complete if we can find an integer $t = \Omega(n^{2/3}\ell^{1/3})$ with

$$2\ell \leqslant t \leqslant n \qquad \text{and} \qquad \frac{n}{2} \geqslant \frac{5\mathrm{e}t^2}{\beta\sqrt{t\ell}}.$$

Calculations reveal that such t exists whenever $\ell \leqslant \beta n/(10 \cdot 2^{3/2}\,\mathrm{e})$. □

The crux of our proof is the following lemma.

Lemma 5.2. *Let $D : \{0,1,\ldots,n\} \to \{0,1\}$. Suppose that $D(\ell) \neq D(\ell-1)$ for some $\ell \leqslant \alpha n$, where $\alpha \in (0,1)$ is the absolute constant from Lemma 5.1. Then there exist sets $X, Y \subseteq \{0,1\}^n$ and a real matrix $K = [K_{xy}]_{x\in X,\, y\in Y}$ such that:*

$$\|K\|_1 = 1, \tag{5.2}$$

$$\|K\| \leqslant \frac{1}{\sqrt{|X|\,|Y|}} \cdot \frac{1}{2^{\Omega(n^{1/3}\ell^{2/3})}}, \tag{5.3}$$

$$\langle K, M\rangle > \frac{1}{3}, \tag{5.4}$$

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

Proof. Lemma 5.1 supplies an integer t, $1 \leqslant t \leqslant n$, that satisfies (5.1). In particular,

$$\deg_{1/3}(D|_t) = \Omega(n^{1/3}\ell^{2/3}). \tag{5.5}$$

Moreover, (5.1) implies that $t \leqslant n/(10\mathrm{e})$. As a result, we can pick an integer $m \in \{n/2, \ldots, n\}$ such that

$$t \mid m \qquad \text{and} \qquad m \geqslant \frac{5\mathrm{e}t^2}{\deg_{1/3}(D|_t)}. \tag{5.6}$$

Define $f : \{0,1\}^t \to \{-1,+1\}$ by

$$f(z) \overset{\text{def}}{=} (-1)^{D(|z|)} \qquad\qquad (z \in \{0,1\}^t).$$

Clearly, $\deg_{1/3}(f) \geqslant \deg_{1/3}(D|_t)$. Then by Corollary 3.3.1, there exists $\psi : \{0,1\}^t \to \mathbb{R}$ such that

$$\sum_{z\in\{0,1\}^t} |\psi(z)| = 1, \tag{5.7}$$

$$\sum_{z\in\{0,1\}^t} \psi(z)\chi_S(z) = 0 \qquad\qquad \text{for all } |S| < \deg_{1/3}(D|_t), \tag{5.8}$$

$$\sum_{z\in\{0,1\}^t} \psi(z)f(z) > \frac{1}{3}. \tag{5.9}$$

Define $X, Y \subseteq \{0,1\}^n$ by

$$\begin{aligned} X &\stackrel{\text{def}}{=} \{x \in \{0,1\}^n : x_{m+1} = \cdots = x_n = 0\}, \\ Y &\stackrel{\text{def}}{=} \{1_V : V \in \mathscr{V}(m,t)\}. \end{aligned}$$

These definitions set up obvious bijections $X \leftrightarrow \{0,1\}^m$ and $Y \leftrightarrow \mathscr{V}(m,t)$. For $y \in Y$, let $V(y) \in \mathscr{V}(m,t)$ denote the set whose characteristic vector is y. With this notation, $|x \wedge y| = \big|x|_{V(y)}\big|$ for all $x \in X$, $y \in Y$. This, along with the definition of f, shows that

$$(-1)^{D(|x\wedge y|)} = f\big(x|_{V(y)}\big) \qquad\qquad (x \in X,\ y \in Y). \tag{5.10}$$

We are now ready to exhibit a real matrix K with the desired criteria. Namely, let

$$K \stackrel{\text{def}}{=} \left[\frac{1}{2^{m-t}(m/t)^t} \cdot \psi(x|_V)\right]_{x\in\{0,1\}^m,\, V\in\mathscr{V}(m,t)} = \left[\frac{1}{2^{m-t}(m/t)^t} \cdot \psi\big(x|_{V(y)}\big)\right]_{x\in X,\, y\in Y}.$$

The first equality defines K as a pattern matrix. The second provides the desired representation $K = [K_{xy}]_{x\in X, y\in Y}$.

It remains to verify that K has properties (5.2)–(5.4). Property (5.2) is immediate from (5.7). Property (5.3) follows by Theorem 4.2 in view of (5.5)–(5.8). Finally, property (5.4) can be seen as follows:

$$\begin{aligned} \langle K, M\rangle &= \frac{1}{2^{m-t}(m/t)^t} \sum_{y\in Y}\sum_{x\in X} \psi\big(x|_{V(y)}\big)\ (-1)^{D(|x\wedge y|)} \\ &= \frac{1}{2^{m-t}(m/t)^t} \sum_{y\in Y}\sum_{x\in X} \psi\big(x|_{V(y)}\big)\ f\big(x|_{V(y)}\big) && \text{by (5.10)} \\ &= \frac{1}{2^{m-t}(m/t)^t} \sum_{y\in Y} 2^{m-t} \sum_{z\in\{0,1\}^t} \psi(z)f(z) \\ &> \frac{1}{3} && \text{by (5.9).} \qquad \square \end{aligned}$$

We are now in a position to prove the desired lower bounds for predicates $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ that change value reasonably close to 0. Extension to the general case is the subject of the next section.

Theorem 5.3. *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Suppose that $D(\ell) \neq D(\ell - 1)$ for some $\ell \leqslant \alpha n$, where $\alpha \in (0, 1)$ is the absolute constant from Lemma* 5.1. *Then*

$$Q^*_{1/3}(D) = \Omega(n^{1/3}\ell^{2/3}).$$

Proof. It suffices to show that $Q^*_{1/9}(D) = \Omega(n^{1/3}\ell^{2/3})$, since the accuracy of a quantum protocol can be amplified from 1/3 to any other constant $\epsilon > 0$ at the expense of a constant multiplicative increase in communication. Consider the sets $X, Y \subseteq \{0, 1\}^n$ and the matrix $K = [K_{xy}]_{x\in X,\, y\in Y}$ from Lemma 5.2. Define $f : X \times Y \to \{0, 1\}$ by

$$f(x, y) = D(|x \wedge y|).$$

By the discrepancy method (Theorem 2.3),

$$Q^*_{1/9}(f) = \Omega(n^{1/3}\ell^{2/3}).$$

Since $Q^*_{1/9}(D) \geqslant Q^*_{1/9}(f)$, the theorem follows. □

6 Final Lower Bound on Communication

In the previous section, we proved strong lower bounds for all predicates D that change value reasonably close to 0. This hard work is now behind us. What remains is to extend the result to arbitrary predicates, which is a straightforward if tedious exercise in shifting and padding. We note that Razborov's proof concludes in a similar way (see [16], beginning of Section 5).

Theorem 6.1. *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Suppose that $D(\ell) \neq D(\ell - 1)$ for some $\ell > \alpha n$, where $\alpha \in (0, 1)$ is the absolute constant from Lemma* 5.1. *Then*

$$Q^*_{1/3}(D) \geqslant c(n - \ell) \tag{6.1}$$

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

$$D_k(i) = D(k + i) \qquad (i = 0, 1, \ldots, n - k).$$

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

$$Q^*_{1/3}(D) \geqslant Q^*_{1/3}(D_k) \qquad\qquad \text{for all } k. \tag{6.2}$$

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

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 (6.1) 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 5.3 is applicable to D_{k_0} and yields:

$$Q^*_{1/3}(D_{k_0}) \geqslant C \cdot (n-k_0)^{1/3}(\ell - k_0)^{2/3}, \tag{6.3}$$

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

$$n - k_0 = \left\lfloor \frac{1}{1-\alpha} \cdot (n-\ell) \right\rfloor \qquad \text{and} \qquad \ell - k_0 = \left\lfloor \frac{\alpha}{1-\alpha} \cdot (n-\ell) \right\rfloor. \tag{6.4}$$

The theorem is now immediate from (6.2)–(6.4). □

Together, Theorems 5.3 and 6.1 yield the main result of this paper.

Theorem 6.2 (Restatement of Theorem 1.2). *Let* $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. *Then*

$$Q^*_{1/3}(D) = \Omega\left(n^{1/3}\ell_0(D)^{2/3} + \ell_1(D)\right).$$

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 5.3 and 6.1 must be applicable, and therefore $Q^*_{1/3}(D) \geqslant \min\left\{\Omega(n^{1/3}\ell^{2/3}),\ \Omega(n-\ell)\right\}$. Since $\ell \leqslant n/2$, this simplifies to

$$Q^*_{1/3}(D) \geqslant \Omega\left(n^{1/3}\ell_0(D)^{2/3}\right). \tag{6.5}$$

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. One of Theorems 5.3 and 6.1 must be applicable, and therefore $Q^*_{1/3}(D) \geqslant \min\{\Omega(n),\ \Omega(\ell_1(D) - 1)\}$. This simplifies to

$$Q^*_{1/3}(D) \geqslant \Omega\left(\ell_1(D)\right). \tag{6.6}$$

The theorem follows from (6.5) and (6.6). □

7 Extensions

Critical to Razborov's proof [16] is the high symmetry of the functions $f(x, y) = D(|x \wedge y|)$. By contrast, the method of this paper applies unchanged to a broader class of functions, for which no lower bounds were previously known. Specifically, we have:

Theorem 7.1. *Let* $f : \{0,1\}^t \to \{0,1\}$ *be arbitrary, where* $t \mid n$. *Put*

$$\begin{aligned} d &\overset{\text{def}}{=} \deg_{1/3}(f), \\ F &\overset{\text{def}}{=} \big[f(x|_V)\big]_{x\in\{0,1\}^n,\, V\in\mathscr{V}(n,t)}. \end{aligned}$$

If $n \geqslant 5\mathrm{e}t^2/d$, *then*

$$Q_{1/3}(F) \geqslant Q^*_{1/3}(F) \geqslant \Omega(d).$$

Proof. Argue as in the proof of Lemma 5.2, with obvious changes, to obtain (5.2)–(5.4). Then apply the discrepancy method (Theorem 2.3). □

In words, Theorem 7.1 delivers the same lower bound on communication while relaxing assumptions about the symmetry of the communication matrix. In particular, the function f in Theorem 7.1 need not be symmetric. The theorem does assume, however, a pattern structure to the communication matrix. There are several ways to work around this assumption:

- It is not necessary to use a single function $f : \{0,1\}^t \to \{0,1\}$ for the entire matrix F. In other words, one can work with distinct functions $f_1, f_2, \ldots,$ one for each column, as long as they all have high approximate degree. The proof of Theorem 4.2 generalizes easily to accommodate this new setup; the remaining machinery needs only cosmetic changes.

- The columns of F need not be indexed by sets in $\mathscr{V}(n, t)$. One could use a different system of sets, e.g., one that arises from a partition of $[n]$ into blocks of varying size. Also, not all sets from the resulting family need to figure as column indices. More generally, one can attempt to work with any set system in which two random sets have small expected intersection.

We designate the functions in Theorem 7.1, as well as their generalizations, by the term *pattern functions*. We defer precise theorems for these generalizations to the full version of the paper, the fundamental mathematical technique being already evident in Theorem 7.1.

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] G. H. Golub and C. F. V. Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, MD, USA, 3rd edition, 1996.

[6] 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.

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

[8] 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.

[9] 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.

[10] 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.

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

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

[13] 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.

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

[15] 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.

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

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

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

[19] 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.