THE PATTERN MATRIX METHOD*

ALEXANDER A. SHERSTOV[†]

Abstract. We develop a novel technique for communication lower bounds, the *pattern matrix method*. Specifically, fix an arbitrary function $f\colon \{0,1\}^n \to \{0,1\}$ and let A_f be the matrix whose columns are each an application of f to some subset of the variables $x_1, x_2, \dots, x_{4n}$. We prove that A_f has bounded-error communication complexity $\Omega(d)$, where d is the approximate degree of f. This result remains valid in the quantum model, regardless of prior entanglement. In particular, it gives a new and simple proof of Razborov's breakthrough quantum lower bounds for disjointness and other symmetric predicates. We further characterize the discrepancy, approximate rank, and approximate trace norm of A_f in terms of well-studied analytic properties of f, broadly generalizing several recent results on small-bias communication and agnostic learning. The method of this paper has also enabled important progress in multiparty communication complexity.

Key words. Pattern matrix method, bounded-error communication complexity, quantum communication complexity, discrepancy, Degree/Discrepancy Theorem, approximate rank, approximate trace norm, linear programming duality, approximation and sign-representation of Boolean functions by real polynomials.

AMS subject classifications. 03D15, 68Q15, 81P68

1. Introduction.

A central model in communication complexity is the *bounded-error model*. Let $f\colon X \times Y \to \{-1,+1\}$ be a given function, where X and Y are finite sets. Alice receives an input $x \in X$, Bob receives $y \in Y$, and their objective is to compute $f(x,y)$ with minimal communication. To this end, Alice and Bob share an unlimited supply of random bits. Their protocol is said to *compute f* if on every input (x,y), the output is correct with probability at least $1-\epsilon$. The canonical setting is $\epsilon = 1/3$, but any other parameter $\epsilon \in (0,1/2)$ can be considered. The *cost* of a protocol is the worst-case number of bits exchanged on any input. Depending on the physical nature of the communication channel, one studies the *classical model*, in which the messages are classical bits 0 and 1, and the more powerful *quantum model*, in which the messages are quantum bits and arbitrary prior entanglement is allowed. The communication complexity in these models is denoted $R_\epsilon(f)$ and $Q^*_\epsilon(f)$, respectively.

Bounded-error protocols have been the focus of much research in communication complexity since the introduction of the area by Yao [66] three decades ago. A variety of techniques have been developed for proving lower bounds on classical communication, e.g., [27, 56, 21, 55, 13, 42, 20, 61]. There has been consistent progress on quantum communication as well [67, 3, 12, 31, 29, 57, 42], although quantum protocols remain less understood than their classical counterparts.

The main contribution of this paper is a novel method for lower bounds on classical and quantum communication complexity, the *pattern matrix method*. The method converts analytic properties of Boolean functions into lower bounds for the corresponding communication problems. The analytic properties in question pertain to the approximation and sign-representation of a given Boolean function by real polynomials of low degree, which are among the oldest and most studied objects in theoretical computer science. In other words, the pattern matrix method takes

[*] A preliminary version of this article appeared under the title "The Pattern Matrix Method for Lower Bounds on Quantum Communication" in *Proceedings of the 40th Annual ACM Symposium on Theory of Computing* (STOC), pages 85-94, 2008.

[†] Department of Computer Science, The University of Texas at Austin (sherstov@cs.utexas.edu).

the wealth of results available on the representations of Boolean functions by real polynomials and puts them at the disposal of communication complexity.

We consider two ways of representing Boolean functions by real polynomials. Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be a given Boolean function. The *ϵ-approximate degree* of f, denoted $\deg_\epsilon(f)$, is the least degree of a real polynomial p such that $|f(x)-p(x)| \leqslant \epsilon$ for all $x \in \{0,1\}^n$. There is an extensive literature on the ϵ-approximate degree of Boolean functions [49, 51, 26, 8, 1, 2, 59, 65], for the canonical setting $\epsilon = 1/3$ and various other settings. Apart from uniform approximation, the other representation scheme of interest to us is sign-representation. Specifically, the *degree-d threshold weight* $W(f,d)$ of f is the minimum $\sum_{|S|\leqslant d} |\lambda_S|$ over all integers λ_S such that

$$f(x) \equiv \operatorname{sgn}\left(\sum_{S\subseteq\{1,\dots,n\},\, |S|\leqslant d} \lambda_S \chi_S(x) \right),$$

where $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. If no such integers λ_S exist, we write $W(f,d)=\infty$. The threshold weight of Boolean functions has been heavily studied, both when $W(f,d)$ is infinite [46, 4, 37, 38, 33, 32, 50] and when it is finite [46, 47, 7, 7, 64, 34, 36, 53, 54]. The notions of uniform approximation and sign-representation are closely related, as we discuss in Section 2. Roughly speaking, the study of threshold weight corresponds to the study of the ϵ-approximate degree for $\epsilon = 1 - o(1)$.

Having defined uniform approximation and sign-representation for Boolean functions, we now describe how we use them to prove communication lower bounds. The central concept in our work is what we call a *pattern matrix.* Consider the communication problem of computing

$$f(x|_V),$$

where $f\colon \{0,1\}^t \to \{-1,+1\}$ is a fixed Boolean function; the string $x \in \{0,1\}^n$ is Alice's input (n is a multiple of t); and the set $V \subset \{1,2,\dots,n\}$ with $|V| = t$ is Bob's input. In words, this communication problem corresponds to a situation when the function f depends on only t of the inputs $x_1,\dots,x_n$. Alice knows the values of all the inputs $x_1,\dots,x_n$ but does not know which t of them are relevant. Bob, on the other hand, knows which t inputs are relevant but does not know their values. This communication game was introduced and studied in an earlier work by the author [61], in the context of small-bias communication. For the purposes of the introduction, one can think of the *(n,t,f)-pattern matrix* as the matrix $[f(x|_V)]_{x,V}$, where V ranges over the $(n/t)^t$ sets that have exactly one element from each block of the following partition:

$$\{1,\dots,n\} = \left\{1,2,\dots,\frac{n}{t}\right\} \cup \left\{\frac{n}{t}+1,\dots,\frac{2n}{t}\right\} \cup \dots \cup \left\{\frac{(t-1)n}{t}+1,\dots,n\right\}.$$

We defer the precise definition to Section 4. Observe that restricting V to be of special form only makes our results stronger.

1.1. Our results. Our main result is a lower bound on the communication complexity of a pattern matrix in terms of the ϵ-approximate degree of the base function f. The lower bound holds for both classical and quantum protocols, regardless of prior entanglement.

THEOREM 1.1 (communication complexity). *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then for every $\epsilon \in [0,1)$ and every $\delta < \epsilon/2$,*

$$Q^*_\delta(F) \geqslant \frac{1}{4}\deg_\epsilon(f)\log_2\left(\frac{n}{t}\right) - \frac{1}{2}\log_2\left(\frac{3}{\epsilon - 2\delta}\right).$$

In particular,

$$Q^*_{1/7}(F) > \frac{1}{4}\deg_{1/3}(f)\log_2\left(\frac{n}{t}\right) - 3. \tag{1.1}$$

Note that Theorem 1.1 yields lower bounds for communication complexity with error probability δ for any $\delta \in (0,1/2)$. In particular, apart from bounded-error communication (1.1), we obtain lower bounds for communication with small bias, i.e., error probability $\frac{1}{2} - o(1)$. In Section 6, we derive another lower bound for small-bias communication, in terms of threshold weight $W(f,d)$.

As R. de Wolf pointed out to us [17], the lower bound (1.1) for bounded-error communication is within a polynomial of optimal. More precisely, F has a classical deterministic protocol with cost $O(\deg_{1/3}(f)^6\log(n/t))$, by the results of Beals et al. [5]. See Proposition 5.1 for details. In particular, Theorem 1.1 exhibits a large new class of communication problems F whose quantum communication complexity is polynomially related to their classical complexity, *even* if prior entanglement is allowed. Prior to our work, the largest class of problems with polynomially related quantum and classical bounded-error complexities was the class of symmetric functions (see Theorem 1.3 below), which is broadly subsumed by Theorem 1.1. Exhibiting a polynomial relationship between the quantum and classical bounded-error complexities for *all* functions $F\colon X \times Y \to \{-1,+1\}$ is a longstanding open problem.

Pattern matrices are of interest because they occur as submatrices in natural communication problems. For example, Theorem 1.1 can be interpreted in terms of function composition. Setting $n = 4t$ for concreteness, we obtain:

COROLLARY 1.2. *Let $f\colon \{0,1\}^t \to \{-1,+1\}$ be given. Define $F\colon \{0,1\}^{4t} \times \{0,1\}^{4t} \to \{-1,+1\}$ by $F(x,y) = f(\dots,(x_{i,1}y_{i,1} \vee x_{i,2}y_{i,2} \vee x_{i,3}y_{i,3} \vee x_{i,4}y_{i,4}),\dots)$. Then*

$$Q^*_{1/7}(F) > \frac{1}{4}\deg_{1/3}(f) - 3.$$

As another illustration of Theorem 1.1, we revisit the quantum communication complexity of symmetric functions. In this setting Alice has a string $x \in \{0,1\}^n$, Bob has a string $y \in \{0,1\}^n$, and their objective is to compute $D(\sum x_iy_i)$ for some predicate $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$ fixed in advance. This framework encompasses several familiar functions, such as DISJOINTNESS (determining if x and y intersect) and INNER PRODUCT MODULO 2 (determining if x and y intersect in an odd number of positions). In a celebrated result, Razborov [57] established optimal lower bounds on the quantum communication complexity of every function of such form:

THEOREM 1.3 (Razborov). *Let $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$ be a given predicate. Put $f(x,y) = D(\sum x_iy_i)$. Then*

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

where $\ell_0(D) \in \{0,1,\dots,\lfloor n/2\rfloor\}$ *and* $\ell_1(D) \in \{0,1,\dots,\lceil n/2\rceil\}$ *are the smallest integers such that* D *is constant in the range* $[\ell_0(D), n-\ell_1(D)]$.

Using Theorem 1.1, we give a new and simple proof of Razborov's result. No alternate proof was available prior to this work, despite the fact that this problem has drawn the attention of various researchers [3, 12, 31, 29, 24, 42]. Moreover, the next-best lower bounds for general predicates were nowhere close to Theorem 1.3. To illustrate, consider the disjointness predicate D, given by $D(t) = 1 \Leftrightarrow t = 0$. Theorem 1.3 shows that it has communication complexity $\Omega(\sqrt{n})$, while the next-best lower bound [3, 12] was $\Omega(\log n)$.

Approximate rank and trace norm. We now describe some matrix-analytic consequences of our work. The ϵ-approximate rank of a matrix $F \in \{-1,+1\}^{m\times n}$, denoted $\operatorname{rk}_\epsilon F$, is the least rank of a real matrix A such that $|F_{ij} - A_{ij}| \leqslant \epsilon$ for all i, j. This natural analytic quantity arose in the study of quantum communication [67, 12, 57] and has since found applications to learning theory. In particular, Klivans and Sherstov [35] proved that concept classes (i.e., sign matrices) with high approximate rank are beyond the scope of known techniques for efficient learning, in Kearns' well-studied *agnostic model* [28]. Exponential lower bounds were cited in [35] on the approximate rank of disjunctions, majority functions, and decision lists, with the corresponding implications for agnostic learning. We broadly generalize these results on approximate rank to *any* functions with high approximate degree or high threshold weight:

Theorem 1.4 (approximate rank). *Let* F *be the* (n,t,f)*-pattern matrix, where* $f\colon \{0,1\}^t \to \{-1,+1\}$ *is given. Then for every* $\epsilon \in [0,1)$ *and every* $\delta \in [0,\epsilon]$,

$$\operatorname{rk}_\delta F \geqslant \left(\frac{\epsilon-\delta}{1+\delta}\right)^2 \left(\frac{n}{t}\right)^{\deg_\epsilon(f)}.$$

In addition, for every $\gamma \in (0,1)$ *and every integer* $d \geqslant 1$,

$$\operatorname{rk}_{1-\gamma} F \geqslant \left(\frac{\gamma}{2-\gamma}\right)^2 \min\left\{\left(\frac{n}{t}\right)^d, \frac{W(f,d-1)}{2t}\right\}.$$

We derive analogous results for the *approximate trace norm,* another matrix-analytic notion that has been studied in complexity theory. Theorem 1.4 is close to optimal for a broad range of parameters. See Section 8 for details.

Discrepancy. The discrepancy of a function $F\colon X \times Y \to \{-1,+1\}$, denoted $\operatorname{disc}(F)$, is a combinatorial measure of the complexity of F (small discrepancy corresponds to high complexity). This complexity measure plays a central role in the study of communication. In particular, it fully characterizes membership in PP^{cc}, the class of communication problems with efficient small-bias protocols [30]. Discrepancy is also known [43] be to equivalent to *margin complexity*, a key notion in learning theory. Finally, discrepancy is of interest in circuit complexity [21, 22, 48]. We are able to characterize the discrepancy of every pattern matrix in terms of threshold weight:

Theorem 1.5 (discrepancy). *Let* F *be the* (n,t,f)*-pattern matrix, for a given function* $f\colon \{0,1\}^t \to \{-1,+1\}$. *Then*

$$\operatorname{disc}(F) \leqslant \min_{d=1,\dots,t} \max\left\{\left(\frac{2t}{W(f,d-1)}\right)^{1/2}, \left(\frac{t}{n}\right)^{d/2}\right\}.$$

As we show in Section 7, Theorem 1.5 is close to optimal. It is a substantial improvement on the author's earlier work [61].

As an application of Theorem 1.5, we revisit the discrepancy of AC^0, the class of polynomial-size constant-depth circuits with AND, OR, NOT gates. In an earlier work [61], we obtained the first exponentially small upper bound on the discrepancy of a function in AC^0. We used this result in [61] to prove that depth-2 majority circuits for AC^0 require exponential size, solving an open problem due to Krause and Pudlák [37]. Using Theorem 1.5, we are able to considerably sharpen the bound in [61]. Specifically, we prove:

Theorem 1.6. *Let* $f(x,y)=\bigvee_{i=1}^{m}\bigwedge_{j=1}^{m^2}(x_{ij}\vee y_{ij})$. *Then*

$$\operatorname{disc}(f)=\exp\{-\Omega(m)\}.$$

We defer the new circuit implications and other discussion to Sections 7 and 10. Independently of the work in [61], Buhrman et al. [11] exhibited another function in AC^0 with exponentially small discrepancy:

Theorem (Buhrman et al.). *Let* $f:\{0,1\}^n\times\{0,1\}^n\to\{-1,+1\}$ *be given by* $f(x,y)=\operatorname{sgn}\left(1+\sum_{i=1}^{n}(-2)^i x_i y_i\right)$. *Then*

$$\operatorname{disc}(f)=\exp\{-\Omega(n^{1/3})\}.$$

Using Theorem 1.5, we give a new and simple proof of this result.

1.2. Our techniques. The setting in which to view our work is the *generalized discrepancy method,* a straightforward but very useful principle introduced by Klauck [29] and reformulated in its current form by Razborov [57]. Let $F(x,y)$ be a Boolean function whose bounded-error communication complexity is of interest. The generalized discrepancy method asks for a Boolean function $H(x,y)$ and a distribution μ on (x,y)-pairs such that:

(1) the functions F and H have correlation $\Omega(1)$ under μ; and

(2) all low-cost protocols have negligible advantage in computing H under μ.

If such H and μ indeed exist, it follows that no low-cost protocol can compute F to high accuracy (otherwise it would be a good predictor for the hard function H as well). This method applies broadly to many models of communication, as we discuss in Section 2.4. It generalizes Yao's original discrepancy method [40], in which $H=F$. The advantage of the generalized version is that it makes it possible, in theory, to prove lower bounds for functions such as DISJOINTNESS, to which the traditional method does not apply.

The hard part, of course, is finding H and μ with the desired properties. Except in rather restricted cases [29, Thm. 4], it was not known how to do it. As a result, the generalized discrepancy method was of limited practical use prior to this paper. Here we overcome this difficulty, obtaining H and μ for a broad range of problems, namely, the communication problems of computing $f(x|_V)$.

Pattern matrices are a crucial first ingredient of our solution. We derive an exact, closed-form expression for the singular values of a pattern matrix and their multiplicities. This spectral information reduces our search from H and μ to a much smaller and simpler object, namely, a function $\psi\colon\{0,1\}^t\to\mathbb{R}$ with certain properties. On the one hand, ψ must be well-correlated with the base function f. On the other hand, ψ must be orthogonal to all low-degree polynomials. We establish the existence of such ψ by passing to the *linear programming dual* of the approximate degree of f. Although the approximate degree and its dual are classical notions, we are not aware of any previous use of this duality to prove communication lower bounds.

For the results that feature threshold weight, we combine the above program with the dual characterization of threshold weight. To derive the remaining results on approximate rank, approximate trace norm, and discrepancy, we apply our main technique along with several additional matrix-analytic and combinatorial arguments.

1.3. Recent work on multiparty complexity. We are pleased to report that this paper has enabled important progress in multiparty communication complexity by a number of researchers. Lee and Shraibman [41] and Chattopadhyay and Ada [14] generalized our method to three and more players, thereby obtaining much improved lower bounds on the multiparty communication complexity of DISJOINTNESS. David and Pitassi [15] ingeniously combined this line of work with the probabilistic method, establishing a separation of the communication classes NP_k^{cc} and BPP_k^{cc} for up to $k = (1-\epsilon)\log n$ players. Their construction was derandomized in a follow-up paper by David, Pitassi, and Viola [16], resulting in an explicit separation. See the survey article [62] for a unified guide to these results, complete with all the key proofs. A very recent development is due to Beame and Huynh-Ngoc [6], who continue this line of research with much improved multiparty lower bounds for AC^0 functions.

1.4. 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 dual characterization of the approximate degree and threshold weight. Section 5 integrates them into the generalized discrepancy method and establishes our main result, Theorem 1.1. In Section 6, we prove an additional version of our main result using threshold weight. We characterize the discrepancy of pattern matrices in Section 7. Approximate rank and approximate trace norm are studied next, in Section 8. We illustrate our main result in Section 9 by giving a new proof of Razborov's quantum lower bounds. As another illustration, we study the discrepancy of AC^0 in Section 10. We conclude with some remarks on the well-known log-rank conjecture in Section 11 and a discussion of related work in Section 12.

2. Preliminaries. We view Boolean functions as mappings $X \to \{-1,+1\}$ for a finite set X, where -1 and 1 correspond to "true" and "false," respectively. Typically, the domain will be $X = \{0,1\}^n$ or $X = \{0,1\}^n \times \{0,1\}^n$. A *predicate* is a mapping $D\colon \{0,1,\ldots,n\} \to \{-1,+1\}$. 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

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

For $b \in \{0,1\}$, we put $\neg b = 1 - b$. For $x \in \{0,1\}^n$, we define $|x| = x_1 + \cdots + x_n$. For $x, y \in \{0,1\}^n$, the notation $x \wedge y \in \{0,1\}^n$ refers as usual to the component-wise conjunction of x and y. Analogously, the string $x \vee y$ stands for the component-wise disjunction of x and y. In particular, $|x \wedge y|$ is the number of positions in which the strings x and y both have a 1. Throughout this manuscript, "log" refers to the logarithm to base 2. As usual, we denote the base of the natural logarithm by $\mathrm{e} = 2.718\ldots$. For any mapping $\phi\colon X \to \mathbb{R}$, where X is a finite set, we adopt the standard notation $\|\phi\|_\infty = \max_{x\in X} |\phi(x)|$. We adopt the standard definition of the sign function:

$$\operatorname{sgn} t = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

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

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

For $S \subseteq [n]$, define $\chi_S\colon \{0,1\}^n \to \{-1,+1\}$ 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 $f\colon \{0,1\}^n \to \mathbb{R}$ has a unique representation of the form

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

where $\hat{f}(S) = \langle f, \chi_S\rangle$. The reals $\hat{f}(S)$ are called the *Fourier coefficients of f*. The degree of f, denoted $\deg(f)$, is the quantity $\max\{|S| : \hat{f}(S) \neq 0\}$. The orthonormality of $\{\chi_S\}$ immediately yields *Parseval's identity*:

$$\sum_{S\subseteq[n]} \hat{f}(S)^2 = \langle f, f\rangle = \mathop{\mathbf{E}}_{x}[f(x)^2]. \tag{2.1}$$

The following fact is immediate from the definition of $\hat{f}(S)$.

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

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

A Boolean function $f\colon \{0,1\}^n \to \{-1,+1\}$ is called *symmetric* if $f(x)$ is uniquely determined by $\sum x_i$. 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 $\sigma\colon [n] \to [n]$. 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 $f(x) \equiv D(\sum x_i)$.

2.1. Matrix analysis. We draw freely on basic notions from matrix analysis. In particular, 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 [23]. 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. We specify matrices by their generic entry, e.g., $A = [F(i,j)]_{i,j}$. 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. We denote the rank of $A \in \mathbb{R}^{m\times n}$ by $\operatorname{rk} A$. We also write

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

We denote the singular values of A by $\sigma_1(A) \geqslant \sigma_2(A) \geqslant \cdots \geqslant \sigma_{\min\{m,n\}}(A) \geqslant 0$. Recall that the spectral norm, trace norm, and Frobenius norm of A are given by

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

For a square matrix $A \in \mathbb{R}^{n\times n}$, its trace is given by $\operatorname{tr} A = \sum A_{ii}$.

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 orthogonal matrices and Σ is diagonal with entries $\sigma_1(A), \sigma_2(A), \dots, \sigma_{\min\{m,n\}}(A)$. For $A, B \in \mathbb{R}^{m\times n}$, we write $\langle A, B\rangle = \sum A_{ij}B_{ij} = \operatorname{tr}(AB^{\mathsf{T}})$. A useful consequence of the singular value decomposition is:

$$\langle A, B\rangle \leqslant \|A\|\,\|B\|_\Sigma \qquad (A, B \in \mathbb{R}^{m\times n}). \tag{2.2}$$

Following [57], we define the *ϵ-approximate trace norm* of a matrix $F \in \mathbb{R}^{m\times n}$ by

$$\|F\|_{\Sigma,\epsilon} = \min\{\|A\|_\Sigma : \|F - A\|_\infty \leqslant \epsilon\}.$$

The next proposition is a trivial consequence of (2.2).

PROPOSITION 2.2. *Let $F \in \mathbb{R}^{m\times n}$ and $\epsilon \geqslant 0$. Then*

$$\|F\|_{\Sigma,\epsilon} \geqslant \sup_{\Psi\neq 0} \frac{\langle F, \Psi\rangle - \epsilon\|\Psi\|_1}{\|\Psi\|}.$$

Proof. Fix any $\Psi \neq 0$ and A such that $\|F - A\|_\infty \leqslant \epsilon$. Then $\langle A, \Psi\rangle \leqslant \|A\|_\Sigma \|\Psi\|$ by (2.2). On the other hand, $\langle A, \Psi\rangle \geqslant \langle F, \Psi\rangle - \|A - F\|_\infty \|\Psi\|_1 \geqslant \langle F, \Psi\rangle - \epsilon\|\Psi\|_1$. Comparing these two estimates of $\langle A, \Psi\rangle$ gives the sought lower bound on $\|A\|_\Sigma$. □

Following [12], we define the *ϵ-approximate rank* of a matrix $F \in \mathbb{R}^{m\times n}$ by

$$\operatorname{rk}_\epsilon F = \min\{\operatorname{rk} A : \|F - A\|_\infty \leqslant \epsilon\}.$$

The approximate rank and approximate trace norm are related by virtue of the singular value decomposition, as follows.

PROPOSITION 2.3. *Let $F \in \mathbb{R}^{m\times n}$ and $\epsilon \geqslant 0$ be given. Then*

$$\operatorname{rk}_\epsilon F \geqslant \frac{(\|F\|_{\Sigma,\epsilon})^2}{\sum_{i,j}(|F_{ij}| + \epsilon)^2}.$$

Proof (adapted from [35]). Fix A with $\|F - A\|_\infty \leqslant \epsilon$. Then

$$\|F\|_{\Sigma,\epsilon} \leqslant \|A\|_\Sigma \leqslant \|A\|_{\mathrm{F}}\sqrt{\operatorname{rk} A} \leqslant \left(\sum_{i,j}(|F_{ij}| + \epsilon)^2\right)^{1/2} \sqrt{\operatorname{rk} A}. \qquad \square$$

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

PROPOSITION 2.4. *For all real matrices A and B of compatible dimensions,*

$$\|AB\|_\Sigma \leqslant \|A\|_{\mathrm{F}}\ \|B\|_{\mathrm{F}}.$$

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:

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

2.2. Approximation and sign-representation. For a function $f\colon \{0,1\}^n \to \mathbb{R}$, we define

$$E(f,d) = \min_p \|f-p\|_\infty,$$

where the minimum is over real polynomials of degree up to d. The *ϵ-approximate degree* of f, denoted $\deg_\epsilon(f)$, is the least d with $E(f,d) \leqslant \epsilon$. In words, the ϵ-approximate degree of f is the least degree of a polynomial that approximates f uniformly within ϵ.

For a Boolean function $f\colon \{0,1\}^n \to \{-1,+1\}$, the ϵ-approximate degree is of particular interest for $\epsilon = 1/3$. The choice of $\epsilon = 1/3$ is a convention and can be replaced by any other constant in $(0,1)$, without affecting $\deg_\epsilon(f)$ by more than a multiplicative constant. Another well-studied notion is the *threshold degree* $\deg_\pm(f)$, defined for a Boolean function $f\colon \{0,1\}^n \to \{-1,+1\}$ as the least degree of a real polynomial p with $f(x) \equiv \operatorname{sgn} p(x)$. In words, $\deg_\pm(f)$ is the least degree of a polynomial that represents f in sign.

So far we have considered representations of Boolean functions by *real* polynomials. Restricting the polynomials to have *integer* coefficients yields another heavily studied representation scheme. The main complexity measure here is the sum of the absolute values of the coefficients. Specifically, for a Boolean function $f\colon \{0,1\}^n \to \{-1,+1\}$, its *degree-$d$ threshold weight* $W(f,d)$ is defined to be the minimum $\sum_{|S|\leqslant d} |\lambda_S|$ over all integers λ_S such that

$$f(x) \equiv \operatorname{sgn}\left(\sum_{S\subseteq\{1,\ldots,n\},\,|S|\leqslant d} \lambda_S \chi_S(x)\right).$$

If no such integers λ_S can be found, we put $W(f,d) = \infty$. It is straightforward to verify that the following three conditions are equivalent: $W(f,d) = \infty$; $E(f,d) = 1$; $d < \deg_\pm(f)$. In all expressions involving $W(f,d)$, we adopt the standard convention that $1/\infty = 0$ and $\min\{t,\infty\} = t$ for any real t.

As one might expect, representations of Boolean functions by real and integer polynomials are closely related. In particular, we have the following relationship between $E(f,d)$ and $W(f,d)$.

THEOREM 2.5. *Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be given. Then for $d = 0, 1, \ldots, n$,*

$$\frac{1}{1-E(f,d)} \leqslant W(f,d) \leqslant \frac{2}{1-E(f,d)} \left\{ \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} \right\}^{3/2},$$

with the convention that $1/0 = \infty$.

Since Theorem 2.5 is not directly used in our derivations, we defer its proof to Appendix A. Similar statements have been noted earlier by several authors [38, 11]. We close this section with Paturi's tight estimate [51] of the approximate degree for each symmetric Boolean function.

THEOREM 2.6 (Paturi). *Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be a given function such that $f(x) \equiv D(\sum x_i)$ for some predicate $D\colon \{0,1,\ldots,n\} \to \{-1,+1\}$. Then*

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

where $\ell_0(D) \in \{0,1,\ldots,\lfloor n/2 \rfloor\}$ and $\ell_1(D) \in \{0,1,\ldots,\lceil n/2 \rceil\}$ are the smallest integers such that D is constant in the range $[\ell_0(D), n - \ell_1(D)]$.

2.3. Quantum communication. This section reviews the quantum model of communication complexity. We include this review mainly for completeness; our proofs rely solely on a basic matrix-analytic property of 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 [57]. 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 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

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

3. Let v denote the projection of $\mathsf{Final}(x,y)$ onto $\mathscr{A} \otimes \mathrm{span}(|1\rangle) \otimes \mathscr{B}$. The output of the protocol is 1 with probability $\langle v, v\rangle$, and 0 with the complementary 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 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 $\mathscr{A}$ 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 $f\colon X \times Y \to \{-1,+1\}$ be a given function. A quantum protocol P is said to compute f with error ϵ if

$$\mathbf{P}\left[f(x,y) = (-1)^{P(x,y)}\right] \geqslant 1-\epsilon$$

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 $0 < \epsilon < 1/2$ 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 $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$ be a predicate. We associate with D the function $f\colon \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ defined by $f(x,y) = D(\sum x_i y_i)$. We let $Q_\epsilon(D) = Q_\epsilon(f)$ and $Q^*_\epsilon(D) = Q^*_\epsilon(f)$. More generally, by computing D in the quantum model we mean computing the associated function f. We write $R_\epsilon(f)$ for the least cost of a classical public-coin protocol for f that errs with probability at most ϵ on any given input. Another classical model that figures in this paper is the *deterministic* model. We let $D(f)$ denote the deterministic communication complexity of f. Throughout this paper, by the communication complexity of a Boolean matrix $F = [F_{ij}]_{i \in I,\, j \in J}$ we will mean the communication complexity of the associated function $f\colon I \times J \to \{-1,+1\}$, given by $f(i,j) = F_{ij}$.

2.4. The generalized discrepancy method. The generalized discrepancy method is an intuitive and elegant technique for proving communication lower bounds. A starting point in our discussion is the following fact due to Linial and Shraibman [42, Lem. 10], with closely analogous statements established earlier by Yao [67], Kremer [39], and Razborov [57].

THEOREM 2.7. *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 2.7 states that the matrix of acceptance probabilities of every low-cost protocol P has a nontrivial factorization. This transition from quantum protocols to matrix factorization is now a standard technique and has been used by various authors in various contexts.

The generalized discrepancy method was first applied by Klauck [29, Thm. 4] and reformulated more broadly by Razborov [57]. The treatment in [57] is informal. In what follows, we propose a precise formulation of the generalized discrepancy method and supply a proof.

Theorem 2.8 (generalized discrepancy method). *Let X, Y be finite sets and $f\colon X \times Y \to \{-1,+1\}$ a given function. Let $\Psi = [\Psi_{xy}]_{x\in X,\, y\in Y}$ be any real matrix with $\|\Psi\|_1 = 1$. Then for each $\epsilon > 0$,*

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

where $F = [f(x,y)]_{x\in X,\, y\in Y}$.

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

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

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

$$\begin{aligned} \langle \Psi, J - 2\Pi\rangle &= \langle \Psi, F\rangle - 2\,\langle \Psi, E\rangle \\ &\geqslant \langle \Psi, F\rangle - 2\epsilon\,\|\Psi\|_1 \\ &= \langle \Psi, F\rangle - 2\epsilon. \end{aligned} \tag{2.3}$$

On the other hand, Theorem 2.7 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,

$$\begin{aligned} \langle \Psi, J - 2\Pi\rangle &\leqslant \|\Psi\|\ \|J - 2\Pi\|_\Sigma && \text{by (2.2)} \\ &\leqslant \|\Psi\|\ \Big(\sqrt{|X|\,|Y|} + 2\,\|\Pi\|_\Sigma\Big) && \text{since } \|J\|_\Sigma = \sqrt{|X|\,|Y|} \\ &\leqslant \|\Psi\|\ \Big(\sqrt{|X|\,|Y|} + 2\,\|A\|_{\mathrm{F}}\,\|B\|_{\mathrm{F}}\Big) && \text{by Prop. 2.4} \\ &\leqslant \|\Psi\|\ \big(2\cdot 4^C + 1\big)\sqrt{|X|\,|Y|}. \end{aligned} \tag{2.4}$$

The theorem follows by comparing (2.3) and (2.4). □

Remark 2.9. Theorem 2.8 is not to be confused with Razborov's *multidimensional* technique, also found in [57], which we will have no occasion to use or describe.

We will now abstract away the particulars of Theorem 2.8 and articulate the fundamental mathematical technique in question. This will clarify the generalized discrepancy method and show that it is simply an extension of Yao's original discrepancy method [40, §3.5]. Let $f\colon X \times Y \to \{-1,+1\}$ 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 $h\colon X \times Y \to \{-1,+1\}$ and a distribution μ on $X \times Y$ that satisfy the following two properties.

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

$$\mathop{\mathbf{E}}_{(x,y)\sim\mu}[f(x,y)h(x,y)] \geqslant \epsilon, \tag{2.5}$$

   where $\epsilon > 0$ is a given constant.
2. *Hardness.* No low-cost protocol P in the given model of communication can compute h to a substantial advantage under μ. Formally, if $P\colon X\times Y \to \{0,1\}$ is a protocol in the given model with cost C bits, then

$$\mathop{\mathbf{E}}_{(x,y)\sim\mu}\left[h(x,y)\,\mathbf{E}\left[(-1)^{P(x,y)}\right]\right] \leqslant 2^{O(C)}\gamma, \tag{2.6}$$

   where $\gamma = o(1)$. The inner expectation in (2.6) 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 $\Omega(\log\{\epsilon/\gamma\})$. Indeed, let P be a protocol with $\mathbf{P}[P(x,y) \neq f(x,y)] \leqslant \epsilon/3$ for all x,y. Then standard manipulations reveal:

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

where the last step uses (2.5). In view of (2.6), this shows that P must have cost $\Omega(\log\{\epsilon/\gamma\})$.

We attach the term *generalized discrepancy method* to this abstract framework. Readers with background in communication complexity will note that the original discrepancy method of Yao [40, §3.5] corresponds to the case when $f = h$ and the communication takes place in the two-party randomized model.

The purpose of our abstract discussion was to expose the fundamental mathematical technique in question, which is independent of the communication model. Indeed, the communication model enters the picture only in the proof of (2.6). 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 (2.4), one considers the quantity $\|\Psi\|\sqrt{|X|\,|Y|}$, where $\Psi = [h(x,y)\mu(x,y)]_{x,y}$. 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 known as the *discrepancy* of h under μ.

3. Duals of approximation and sign-representation. Crucial to our work are the dual characterizations of the uniform approximation and sign-representation of Boolean functions by real polynomials. As a starting point, we recall a classical result from approximation theory due to Ioffe and Tikhomirov [25] on the duality of norms. A more recent treatment is available in the textbook of DeVore and Lorentz [18], p. 61, Thm. 1.3. We provide a short and elementary proof of this result in Euclidean space, which will suffice for our purposes. We let $\mathbb{R}^X$ stand for the linear space of real functions on the set X.

Theorem 3.1 (Ioffe and Tikhomirov). *Let X be a finite set. Fix $\Phi \subseteq \mathbb{R}^X$ and a function $f\colon X \to \mathbb{R}$. Then*

$$\min_{\phi \in \operatorname{span}(\Phi)} \|f - \phi\|_\infty = \max_{\psi} \left\{ \sum_{x \in X} f(x)\psi(x) \right\}, \tag{3.1}$$

where the maximum is over all functions $\psi\colon X \to \mathbb{R}$ such that

$$\sum_{x \in X} |\psi(x)| \leqslant 1$$

and, for each $\phi \in \Phi$,

$$\sum_{x \in X} \phi(x)\psi(x) = 0.$$

Proof. The theorem holds trivially when $\operatorname{span}(\Phi) = \{0\}$. Otherwise, let $\phi_1, \ldots, \phi_k$ be a basis for $\operatorname{span}(\Phi)$. Observe that the left member of (3.1) is the optimum of the following linear program in the variables $\epsilon, \alpha_1, \ldots, \alpha_k$:

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

Standard manipulations reveal the dual:

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

Both programs are clearly feasible and thus have the same finite optimum. We have already observed that the optimum of first program is the left-hand side of (3.1). Since $\phi_1, \ldots, \phi_k$ form a basis for $\operatorname{span}(\Phi)$, the optimum of the second program is by definition the right-hand side of (3.1). □

As a corollary to Theorem 3.1, we obtain a dual characterization of the approximate degree.

THEOREM 3.2 (approximate degree). *Fix $\epsilon \geqslant 0$. Let $f\colon \{0,1\}^n \to \mathbb{R}$ be given, $d = \deg_\epsilon(f) \geqslant 1$. Then there is a function $\psi\colon \{0,1\}^n \to \mathbb{R}$ such that*

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

Proof. Set $X = \{0,1\}^n$ and $\Phi = \{\chi_S : |S| < d\} \subset \mathbb{R}^X$. Since $\deg_\epsilon(f) = d$, we conclude that

$$\min_{\phi\in\operatorname{span}(\Phi)} \|f - \phi\|_\infty > \epsilon.$$

In view of Theorem 3.1, we can take ψ to be any function for which the maximum is achieved in (3.1). □

We now state the dual characterization of the threshold degree, which is better known as Gordan's Transposition Theorem [58, §7.8].

THEOREM 3.3 (threshold degree). *Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be given, $d = \deg_\pm(f)$. Then there is a distribution μ over $\{0,1\}^n$ with*

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

See [61] for a derivation of Theorem 3.3 using linear programming duality. Alternately, it can be derived as a corollary to Theorem 3.1. We close this section with one final dual characterization, corresponding to sign-representation by integer polynomials.

THEOREM 3.4 (threshold weight). *Fix a function $f\colon \{0,1\}^n \to \{-1,+1\}$ and an integer $d \geqslant \deg_\pm(f)$. Then for every distribution μ on $\{0,1\}^n$,*

$$\max_{|S|\leqslant d} \left| \mathop{\mathbf{E}}_{x\sim\mu}[f(x)\chi_S(x)] \right| \geqslant \frac{1}{W(f,d)}. \tag{3.2}$$

Furthermore, there exists a distribution μ such that

$$\max_{|S|\leqslant d} \left| \mathop{\mathbf{E}}_{x\sim\mu}[f(x)\chi_S(x)] \right| \leqslant \left(\frac{2n}{W(f,d)}\right)^{1/2}. \tag{3.3}$$

Inequalities (3.2) and (3.3) are originally due to Hajnal et al. [22] and Freund [19], respectively. For an integrated and simplified treatment of both results, see Goldmann et al. [21], Lem. 4 and Thm. 10.

4. Pattern matrices. We now turn to the second ingredient of our proof, a certain family of real matrices that we introduce. Our goal here is to explicitly calculate their singular values. As we shall see later, this provides a convenient means to generate hard communication problems.

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

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

where $i_1 < i_2 < \cdots < i_t$ are the elements of V. We are ready for a formal definition of our matrix family.

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

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

In words, A is the matrix of size 2^n by $(n/t)^t 2^t$ whose rows are indexed by strings $x \in \{0,1\}^n$, whose columns are indexed by pairs $(V,w) \in \mathscr{V}(n,t) \times \{0,1\}^t$, 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 $\sigma_1(A), \ldots, \sigma_{\operatorname{rk} A}(A), \sigma_1(B), \ldots, \sigma_{\operatorname{rk} B}(B)$.*

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,

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

Fix spectral decompositions

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

Then

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

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 $u_1, \ldots, u_{\mathrm{rk}\,A}, v_1, \ldots, v_{\mathrm{rk}\,B}$ form an orthonormal set. Recalling (4.1), we conclude that the spectral decomposition of $(A + B)(A + B)^{\mathsf{T}}$ is

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

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

We are ready for the main result of this section.

THEOREM 4.3 (singular values of a pattern matrix). *Let $\phi\colon \{0,1\}^t \to \mathbb{R}$ 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,

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

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

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

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 $(n/t)^{|S|}$. 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 claimed. □

5. Pattern matrix method using uniform approximation. The previous two sections examined relevant dual representations and the spectrum of pattern matrices. Having studied 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 complexity, which gives strong lower bounds for every pattern matrix generated by a Boolean function with high approximate degree.

THEOREM 1.1 (restated from p. 3). *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then for every $\epsilon \in [0,1)$ and every $\delta < \epsilon/2$,*

$$Q^*_\delta(F) \geqslant \frac{1}{4}\deg_\epsilon(f)\log\left(\frac{n}{t}\right) - \frac{1}{2}\log\left(\frac{3}{\epsilon-2\delta}\right). \tag{5.1}$$

In particular,

$$Q^*_{1/7}(F) > \frac{1}{4}\deg_{1/3}(f)\log\left(\frac{n}{t}\right) - 3. \tag{5.2}$$

Proof. Since (5.1) immediately implies (5.2), we will focus on the former in the remainder of the proof. Let $d = \deg_\epsilon(f) \geqslant 1$. By Theorem 3.2, there is a function $\psi\colon \{0,1\}^t \to \mathbb{R}$ such that:

$$\hat{\psi}(S) = 0 \qquad (|S| < d), \tag{5.3}$$

$$\sum_{z\in\{0,1\}^t} |\psi(z)| = 1, \tag{5.4}$$

$$\sum_{z\in\{0,1\}^t} \psi(z)f(z) > \epsilon. \tag{5.5}$$

Let Ψ be the $(n,t,2^{-n}(n/t)^{-t}\psi)$-pattern matrix. Then (5.4) and (5.5) show that

$$\|\Psi\|_1 = 1, \qquad \langle F,\Psi\rangle > \epsilon. \tag{5.6}$$

Our last task is to calculate $\|\Psi\|$. By (5.4) and Proposition 2.1,

$$\max_{S\subseteq[t]} |\hat{\psi}(S)| \leqslant 2^{-t}. \tag{5.7}$$

Theorem 4.3 yields, in view of (5.3) and (5.7):

$$\|\Psi\| \leqslant \left(\frac{t}{n}\right)^{d/2} \left(2^{n+t}\left(\frac{n}{t}\right)^{t}\right)^{-1/2}. \tag{5.8}$$

Now (5.1) follows from (5.6), (5.8), and Theorem 2.8. □

Theorem 1.1 gives lower bounds not only for bounded-error communication but also for communication protocols with error probability $\frac{1}{2} - o(1)$. For example, if a function $f\colon \{0,1\}^t \to \{-1,+1\}$ requires a polynomial of degree d for approximation within $1-o(1)$, equation (5.1) gives a lower bound for small-bias communication. We will complement and refine that estimate in the next section, which is dedicated to small-bias communication.

We now prove the corollary to Theorem 1.1 on function composition, stated in the introduction.

Proof of Corollary 1.2. The $(2t, t, f)$-pattern matrix occurs as a submatrix of $[F(x,y)]_{x,y\in\{0,1\}^{4t}}$. □

Finally, we show that the lower bound (5.2) derived above for bounded-error communication complexity is tight up to a polynomial factor, even for deterministic protocols. The proof follows a well-known argument in the literature [9, 5] and was pointed out to us by R. de Wolf [17].

Proposition 5.1 (R. de Wolf, personal communication [17]). *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then*

$$D(F) \leqslant O(\mathrm{dt}(f)\log(n/t)) \leqslant O(\deg_{1/3}(f)^6 \log(n/t)),$$

where $\mathrm{dt}(f)$ *is the least depth of a decision tree for f. In particular,* (5.2) *is tight up to a polynomial factor.*

Proof. Beals al. [5, Cor. 5.6] prove that $\mathrm{dt}(f) \leqslant O(\deg_{1/3}(f)^6)$ for all Boolean functions f. Therefore, it suffices to prove an upper bound of $O(d\log(n/t))$ on the deterministic communication complexity of F, where $d = \mathrm{dt}(f)$.

The needed deterministic protocol is well-known. Fix a depth-d decision tree for f. Let $(x,(V,w))$ be a given input. Alice and Bob start at the root of the decision tree, labeled by some variable $i \in \{1,\ldots,t\}$. By exchanging $\lceil\log(n/t)\rceil + 2$ bits, Alice and Bob determine $(x|_V)_i \oplus w_i \in \{0,1\}$ and take the corresponding branch of the tree. The process repeats until a leaf is reached, at which point both parties learn $f(x|_V \oplus w)$. □

6. Pattern matrix method using threshold weight. As we have already mentioned, Theorem 1.1 of the previous section can be used to obtain lower bounds not only for bounded-error communication but also small-bias communication. In the latter case, one first needs to show that the base function $f\colon \{0,1\}^t \to \{-1,+1\}$ cannot be approximated pointwise within $1-o(1)$ by a real polynomial of a given degree d. In this section, we derive a different lower bound for small-bias communication, this time using the assumption that the threshold weight $W(f,d)$ is high. We will see that this new lower bound is nearly optimal and closely related to the lower bound in Theorem 1.1.

Theorem 6.1 (pattern matrix method using threshold weight). *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then for every integer $d \geqslant 1$ and real $\gamma \in (0,1)$,*

$$Q^*_{1/2-\gamma/2}(F) \geqslant \frac{1}{4}\min\left\{d\log\frac{n}{t},\ \log\frac{W(f,d-1)}{2t}\right\} - \frac{1}{2}\log\frac{3}{\gamma}. \tag{6.1}$$

In particular,

$$Q^*_{1/2-\gamma/2}(F) \geqslant \frac{1}{4}\deg_\pm(f)\log\left(\frac{n}{t}\right) - \frac{1}{2}\log\frac{3}{\gamma}. \tag{6.2}$$

Proof. Letting $d = \deg_\pm(f)$ in (6.1) yields (6.2), since $W(f,d-1) = \infty$ in that case. In the remainder of the proof, we focus on (6.1) alone.

We claim that there exists a distribution μ on $\{0,1\}^t$ such that

$$\max_{|S|<d}\left|\mathop{\mathbf{E}}_{z\sim\mu}[f(z)\chi_S(z)]\right| \leqslant \left(\frac{2t}{W(f,d-1)}\right)^{1/2}. \tag{6.3}$$

For $d \leqslant \deg_\pm(f)$, the claim holds by Theorem 3.3 since $W(f,d-1) = \infty$ in that case. For $d > \deg_\pm(f)$, the claim holds by Theorem 3.4.

Now, define $\psi\colon \{0,1\}^t \to \mathbb{R}$ by $\psi(z) = f(z)\mu(z)$. It follows from (6.3) that

$$|\hat\psi(S)| \leqslant 2^{-t}\left(\frac{2t}{W(f,d-1)}\right)^{1/2} \qquad (|S| < d), \tag{6.4}$$

$$\sum_{z\in\{0,1\}^t} |\psi(z)| = 1, \tag{6.5}$$

$$\sum_{z\in\{0,1\}^t} \psi(z)f(z) = 1. \tag{6.6}$$

Let Ψ be the $(n,t,2^{-n}(n/t)^{-t}\psi)$-pattern matrix. Then (6.5) and (6.6) show that

$$\|\Psi\|_1 = 1, \qquad \langle F,\Psi\rangle = 1. \tag{6.7}$$

It remains to calculate $\|\Psi\|$. By (6.5) and Proposition 2.1,

$$\max_{S\subseteq[t]} |\hat\psi(S)| \leqslant 2^{-t}. \tag{6.8}$$

Theorem 4.3 yields, in view of (6.4) and (6.8):

$$\|\Psi\| \leqslant \max\left\{\left(\frac{t}{n}\right)^{d/2}, \left(\frac{2t}{W(f,d-1)}\right)^{1/2}\right\}\left(2^{n+t}\left(\frac{n}{t}\right)^t\right)^{-1/2}. \tag{6.9}$$

Now (6.1) follows from (6.7), (6.9), and Theorem 2.8. ◻

Recall from Theorem 2.5 that the quantities $E(f,d)$ and $W(f,d)$ are related for all f and d. In particular, the lower bounds for small-bias communication in Theorems 1.1 and 6.1 are quite close, and either one can be approximately deduced from the other. In deriving both results from scratch, as we did, our motivation was to obtain the tightest bounds and to illustrate the pattern matrix method in different contexts. We will now see that the lower bound in Theorem 6.1 is close to optimal, even for classical protocols.

THEOREM 6.2. *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then for every integer $d \geqslant \deg_\pm(f)$,*

$$Q^*_{1/2-\gamma/2}(F) \leqslant R_{1/2-\gamma/2}(F) \leqslant d \log \left(\frac{n}{t}\right) + 3,$$

where $\gamma = 1/W(f,d)$.

Proof. The communication protocol that we will describe is standard and has been used in one form or another in several works, e.g., [52, 21, 60, 61]. Put $W = W(f,d)$ and fix a representation

$$f(z) \equiv \operatorname{sgn}\left(\sum_{S\subseteq[t],\ |S|\leqslant d} \lambda_S \chi_S(z)\right),$$

where the integers λ_S satisfy $\sum |\lambda_S| = W$. On input $(x,(V,w))$, the protocol proceeds as follows. Let $i_1 < i_2 < \cdots < i_t$ be the elements of V. Alice and Bob use their shared randomness to pick a set $S \subseteq [t]$ with $|S| \leqslant d$, according to the probability distribution $|\lambda_S|/W$. Next, Bob sends Alice the indices $\{i_j : j \in S\}$ as well as the bit $\chi_S(w)$. With this information, Alice computes the product $\operatorname{sgn}(\lambda_S)\chi_S(x|_V)\chi_S(w) = \operatorname{sgn}(\lambda_S)\chi_S(x|_V \oplus w)$ and announces the result as the output of the protocol.

Assuming an optimal encoding of the messages, the communication cost of this protocol is bounded by

$$\left\lceil \log \left(\frac{n}{t}\right)^d \right\rceil + 2 \leqslant d \log\left(\frac{n}{t}\right) + 3,$$

as desired. On each input x, V, w, the output of the protocol is a random variable $P(x,V,w) \in \{-1,+1\}$ that obeys

$$\begin{aligned} f(x|_V \oplus w)\,\mathbf{E}[P(x,V,w)] &= f(x|_V \oplus w) \sum_{|S|\leqslant d} \frac{|\lambda_S|}{W} \operatorname{sgn}(\lambda_S)\chi_S(x|_V \oplus w) \\ &= \frac{1}{W}\left|\sum_{|S|\leqslant d} \lambda_S \chi_S(x|_V \oplus w)\right| \\ &\geqslant \frac{1}{W}, \end{aligned}$$

which means that the protocol produces the correct answer with probability $\frac{1}{2} + \frac{1}{2W}$ or greater. ◻

7. Discrepancy of pattern matrices. We now restate some of the results of the previous section in terms of *discrepancy*, a key notion already mentioned in Section 2.4. This quantity figures prominently in the study of small-bias communication as well as various applications, such as learning theory and circuit complexity.

For a Boolean function $f\colon X \times Y \to \{-1,+1\}$ and a probability distribution λ on $X \times Y$, the *discrepancy* of f under λ is defined by

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

We put

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

As usual, we will identify a function $f\colon X \times Y \to \{-1,+1\}$ with its communication matrix $F = [f(x,y)]_{x,y}$ and use 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 a well-known matrix-analytic reformulation; cf. Kushilevitz & Nisan [40, Ex. 3.29]. For matrices $A = [A_{xy}]$ and $B = [B_{xy}]$, recall that their *Hadamard product* is given by $A \circ B = [A_{xy}B_{xy}]$.

PROPOSITION 7.1. *Let X, Y be finite sets, $f\colon X \times Y \to \{-1,+1\}$ a given function. Then*

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

where $F = [f(x,y)]_{x \in X,\, y \in Y}$ and P is any matrix whose entries are nonnegative and sum to 1 (viewed as a probability distribution). In particular,

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

where the minimum is over matrices P whose entries are nonnegative and sum to 1.

Proof. We have

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

as claimed. □

We will need one last ingredient, a well-known lower bound on communication complexity in terms of discrepancy.

PROPOSITION 7.2 (see [40, pp. 36–38]). *For every function $f\colon X \times Y \to \{-1,+1\}$ and every $\gamma \in (0,1)$,*

$$R_{1/2-\gamma/2}(f) \geqslant \log \frac{\gamma}{\operatorname{disc}(f)}.$$

Using Theorems 6.1 and 6.2, we will now characterize the discrepancy of pattern matrices in terms of threshold weight.

THEOREM 7.3 (discrepancy of pattern matrices). *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Then for every integer $d \geqslant 0$,*

$$\operatorname{disc}(F) \geqslant \frac{1}{8W(f,d)} \left(\frac{t}{n}\right)^{d} \tag{7.1}$$

and

$$\operatorname{disc}(F)^2 \leqslant \max\left\{\frac{2t}{W(f,d-1)}, \left(\frac{t}{n}\right)^{d}\right\}. \tag{7.2}$$

In particular,

$$\operatorname{disc}(F) \leqslant \left(\frac{t}{n}\right)^{\deg_\pm(f)/2}. \tag{7.3}$$

Proof. The lower bound (7.1) is immediate from Theorem 6.2 and Proposition 7.2. For the upper bound (7.2), construct the matrix Ψ as in the proof of Theorem 6.1. Then (6.7) shows that $\Psi = F \circ P$ for a nonnegative matrix P whose entries sum to 1. As a result, (7.2) follows from (6.9) and Proposition 7.1. Finally, (7.3) follows by taking $d = \deg_\pm(f)$ in (7.2), since $W(f, d-1) = \infty$ in that case. □

This settles Theorem 1.5 from the introduction. Theorem 7.3 follows up and considerably improves on our earlier result, the *Degree/Discrepancy Theorem* [61]:

Theorem 7.4 (Sherstov). *Let $f\colon \{0,1\}^t \to \{-1,+1\}$ be given. Fix an integer $n \geqslant t$. Let $M = [f(x|_S)]_{x,S}$, where the row index x ranges over $\{0,1\}^n$ and the column index S ranges over all t-element subsets of $\{1, 2, \ldots, n\}$. Then*

$$\operatorname{disc}(M) \leqslant \left(\frac{4\mathrm{e}t^2}{n \deg_\pm(f)}\right)^{\deg_\pm(f)/2}.$$

Note that (7.3) is already stronger than Theorem 7.4. In Section 10, we will see an example when Theorem 7.3 gives an exponential improvement on Theorem 7.4.

Threshold weight is typically easier to analyze than the approximate degree. For completeness, however, we will now supplement Theorem 7.3 with an alternate bound on the discrepancy of a pattern matrix in terms of the approximate degree.

Theorem 7.5. *Let F be the (n,t,f)-pattern matrix, for a given function $f\colon \{0,1\}^t \to \{-1,+1\}$. Then for every $\gamma > 0$,*

$$\operatorname{disc}(F) \leqslant \gamma + \left(\frac{t}{n}\right)^{\deg_{1-\gamma}(f)/2}.$$

Proof. Let $d = \deg_{1-\gamma}(f) \geqslant 1$. Define $\epsilon = 1 - \gamma$ and construct the matrix Ψ as in the proof of Theorem 1.1. Then (5.6) shows that $\Psi = H \circ P$, where H is a sign matrix and P is a nonnegative matrix whose entries sum to 1. Viewing P as a probability distribution, we infer from (5.8) and Proposition 7.1 that

$$\operatorname{disc}_P(H) \leqslant \left(\frac{t}{n}\right)^{d/2}. \tag{7.4}$$

Moreover,

$$\begin{aligned}\operatorname{disc}_P(F) &\leqslant \operatorname{disc}_P(H) + \|(F-H)\circ P\|_1\\ &= \operatorname{disc}_P(H) + 1 - \langle F, H\circ P\rangle\\ &\leqslant \operatorname{disc}_P(H) + \gamma, \qquad (7.5)\end{aligned}$$

where the last step follows because $\langle F, \Psi\rangle > \epsilon = 1-\gamma$ by (5.6). The proof is complete in view of (7.4) and (7.5). □

8. Approximate rank and trace norm of pattern matrices. We will now use the results of the previous sections to analyze the approximate rank and approximate trace norm of pattern matrices. These notions were originally motivated by lower bounds on quantum communication [67, 12, 57]. However, they also arise in learning theory [35] and are natural matrix-analytic quantities in their own right.

THEOREM 8.1. *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Let $s = 2^{n+t}(n/t)^t$ be the number of entries in F. Then for every $\epsilon \in [0,1)$ and every $\delta \in [0,\epsilon]$,*

$$\|F\|_{\Sigma,\delta} \geqslant (\epsilon-\delta)\left(\frac{n}{t}\right)^{\deg_\epsilon(f)/2}\sqrt{s} \tag{8.1}$$

and

$$\operatorname{rk}_\delta F \geqslant \left(\frac{\epsilon-\delta}{1+\delta}\right)^2\left(\frac{n}{t}\right)^{\deg_\epsilon(f)}. \tag{8.2}$$

Proof. We may assume that $\deg_\epsilon(f) \geqslant 1$, since otherwise f is a constant function and the claims hold trivially by taking $\Psi = F$ in Proposition 2.2. Construct Ψ as in the proof of Theorem 1.1. Then the claimed lower bound on $\|F\|_{\Sigma,\delta}$ follows from (5.6), (5.8), and Proposition 2.2. Finally, (8.2) follows immediately from (8.1) and Proposition 2.3. □

We prove an additional lower bound in the case of small-bias approximation.

THEOREM 8.2. *Let F be the (n,t,f)-pattern matrix, where $f\colon \{0,1\}^t \to \{-1,+1\}$ is given. Let $s = 2^{n+t}(n/t)^t$ be the number of entries in F. Then for every $\gamma \in (0,1)$ and every integer $d \geqslant 1$,*

$$\|F\|_{\Sigma,1-\gamma} \geqslant \gamma \min\left\{\left(\frac{n}{t}\right)^{d/2}, \left(\frac{W(f,d-1)}{2t}\right)^{1/2}\right\}\sqrt{s} \tag{8.3}$$

and

$$\operatorname{rk}_{1-\gamma} F \geqslant \left(\frac{\gamma}{2-\gamma}\right)^2 \min\left\{\left(\frac{n}{t}\right)^{d}, \frac{W(f,d-1)}{2t}\right\}. \tag{8.4}$$

In particular,

$$\|F\|_{\Sigma,1-\gamma} \geqslant \gamma\left(\frac{n}{t}\right)^{\deg_\pm(f)/2}\sqrt{s} \tag{8.5}$$

and

$$\operatorname{rk}_{1-\gamma} F \geqslant \left(\frac{\gamma}{2-\gamma}\right)^2\left(\frac{n}{t}\right)^{\deg_\pm(f)}. \tag{8.6}$$

Proof. Construct Ψ as in the proof of Theorem 6.1. Then the claimed lower bound on $\|F\|_{\Sigma,\delta}$ follows from (6.7), (6.9), and Proposition 2.2. Now (8.4) follows from (8.3) and Proposition 2.3. Finally, (8.5) and (8.6) follow by taking $d = \deg_\pm(f)$ in (8.3) and (8.4), respectively, since $W(f,d-1) = \infty$ in that case. □

Theorems 8.1 and 8.2 settle Theorem 1.4 from the introduction.

Recall that Theorem 4.3 gives an easy way to calculate the trace norm and rank of a pattern matrix. In particular, it is straightforward to verify that the lower bounds in (8.2) and (8.4) are close to optimal for various choices of ϵ, δ, γ. For example, one has $\|F - A\|_\infty \leqslant 1/3$ by taking F and A to be the (n, t, f)- and (n, t, ϕ)-pattern matrices, where $\phi\colon \{0,1\}^t \to \mathbb{R}$ is any polynomial of degree $\deg_{1/3}(f)$ with $\|f - \phi\|_\infty \leqslant 1/3$.

9. Application: quantum complexity of symmetric functions.

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 $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$. We first solve the problem for all predicates D that change value close to 0. Extension to the general case will require an additional step.

THEOREM 9.1. *Let $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$ be a given predicate. Suppose that $D(\ell) \neq D(\ell-1)$ for some $\ell \leqslant \frac{1}{8}n$. Then*

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

Proof. It suffices to show that $Q^*_{1/7}(D) \geqslant \Omega(\sqrt{n\ell})$. Define $f\colon \{0,1\}^{\lfloor n/4 \rfloor} \to \{-1,+1\}$ by $f(z) = D(|z|)$. Then $\deg_{1/3}(f) \geqslant \Omega(\sqrt{n\ell})$ by Theorem 2.6. Theorem 1.1 implies that

$$Q^*_{1/7}(F) \geqslant \Omega(\sqrt{n\ell}),$$

where F is the $(2\lfloor n/4 \rfloor, \lfloor n/4 \rfloor, f)$-pattern matrix. Since F occurs as a submatrix of $[D(|x \wedge y|)]_{x,y}$, the proof is complete. □

The remainder of this section is a simple if tedious exercise in shifting and padding. We note that Razborov's proof concludes in a similar way (see [57], beginning of Section 5).

THEOREM 9.2. *Let $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$ be a given predicate. Suppose that $D(\ell) \neq D(\ell-1)$ for some $\ell > \frac{1}{8}n$. Then*

$$Q^*_{1/3}(D) \geqslant c(n-\ell) \tag{9.1}$$

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 $D_k\colon \{0,1,\dots,n-k\} \to \{-1,+1\}$ is given by $D_k(i) \equiv D(k+i)$. Since the new problem is a restricted version of the original, we have

$$Q^*_{1/3}(D) \geqslant Q^*_{1/3}(D_k). \tag{9.2}$$

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

$$k = \ell - \left\lfloor \frac{\alpha}{1-\alpha} \cdot (n-\ell) \right\rfloor,$$

where $\alpha = \frac{1}{8}$. Note that k is an integer between 1 and ℓ (because $\ell > \alpha n$). The equality $k = \ell$ occurs if and only if $\left\lfloor \frac{\alpha}{1-\alpha}(n-\ell) \right\rfloor = 0$, in which case (9.1) holds trivially for c suitably small. Thus, we can assume that $1 \leqslant k \leqslant \ell - 1$, in which case $D_k(\ell-k) \neq D_k(\ell-k-1)$ and $\ell - k \leqslant \alpha(n-k)$. Therefore, Theorem 9.1 is applicable to D_k and yields:

$$Q^*_{1/3}(D_k) \geqslant C\sqrt{(n-k)(\ell-k)}, \tag{9.3}$$

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

$$n - k = \left\lfloor \frac{1}{1-\alpha} \cdot (n-\ell) \right\rfloor, \qquad \ell - k = \left\lfloor \frac{\alpha}{1-\alpha} \cdot (n-\ell) \right\rfloor. \tag{9.4}$$

The theorem is now immediate from (9.2)–(9.4). □

Together, Theorems 9.1 and 9.2 give the main result of this section:

Theorem 1.3 (restated from p. 4). *Let $D\colon \{0,1,\dots,n\} \to \{-1,+1\}$. Then*

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

where $\ell_0(D) \in \{0,1,\dots,\lfloor n/2\rfloor\}$ and $\ell_1(D) \in \{0,1,\dots,\lceil n/2\rceil\}$ are the smallest integers such that D is constant in the range $[\ell_0(D), n-\ell_1(D)]$.

Proof. If $\ell_0(D) \neq 0$, set $\ell = \ell_0(D)$ and note that $D(\ell) \neq D(\ell-1)$ by definition. One of Theorems 9.1 and 9.2 must be applicable, and therefore $Q^*_{1/3}(D) \geqslant \min\{\Omega(\sqrt{n\ell}),\ \Omega(n-\ell)\}$. Since $\ell \leqslant n/2$, this simplifies to

$$Q^*_{1/3}(D) \geqslant \Omega(\sqrt{n\ell_0(D)}). \tag{9.5}$$

If $\ell_1(D) \neq 0$, set $\ell = n - \ell_1(D) + 1 \geqslant n/2$ and note that $D(\ell) \neq D(\ell-1)$ as before. By Theorem 9.2,

$$Q^*_{1/3}(D) \geqslant \Omega\left(\ell_1(D)\right). \tag{9.6}$$

The theorem follows from (9.5) and (9.6). □

10. Application: discrepancy of constant-depth circuits. As another application of the pattern matrix method, we revisit the discrepancy of AC^0, the class of polynomial-size constant-depth circuits with AND, OR, NOT gates. In an earlier work [61], we obtained the first exponentially small upper bound on the discrepancy of a function in AC^0, with applications to threshold circuits. Independently, Buhrman et al. [11] exhibited another function in AC^0 with exponentially small discrepancy. We revisit these two discrepancy bounds below, considerably sharpening the bound in [61] and giving a new and simple proof of the bound in [11].

Consider the function $\mathrm{MP}_m\colon \{0,1\}^{4m^3} \to \{-1,+1\}$ given by

$$\mathrm{MP}_m(x) = \bigvee_{i=1}^{m} \bigwedge_{j=1}^{4m^2} x_{ij}.$$

This function was originally defined and studied by Minsky and Papert [46] in their seminal monograph on perceptrons. Using this function and the Degree/Discrepancy Theorem (Theorem 7.4), an upper bound of $\exp\{-\Omega(n^{1/5})\}$ was derived in [61] on the discrepancy of an explicit AC^0 circuit $f\colon \{0,1\}^n \times \{0,1\}^n \to \{-1,+1\}$ of depth 3. We will now sharpen that bound to $\exp\{-\Omega(n^{1/3})\}$.

Theorem 1.6 (restated from p. 5). *Let $f(x,y) = \mathrm{MP}_m(x \vee y)$. Then*

$$\mathrm{disc}(f) = \exp\{-\Omega(m)\}.$$

Proof. Put $d = \lfloor m/2 \rfloor$. A well-known result of Minsky and Papert [46] states that $\deg_\pm(\mathrm{MP}_d) \geqslant d$. Since the $(8d^3, 4d^3, \mathrm{MP}_d)$-pattern matrix is a submatrix of $[f(x,y)]_{x,y}$, the proof is complete in view of equation (7.3) of Theorem 7.3. □

We now turn to the result of Buhrman et al. The ODD-MAX-BIT function $\mathrm{OMB}_n \colon \{0,1\}^n \to \{-1,+1\}$, due to Beigel [7], is given by

$$\mathrm{OMB}_n(x) = \operatorname{sgn}\left(1 + \sum_{i=1}^{n} (-2)^i x_i\right). \tag{10.1}$$

It is straightforward to compute OMB_n by a linear-size DNF formula and even a decision list. In particular, OMB_n belongs to the class AC^0. Buhrman et al. [11, §3.2] proved the following result.

THEOREM 10.1 (Buhrman et al.). *Let $f(x,y) = \mathrm{OMB}_n(x \wedge y)$. Then*

$$\operatorname{disc}(f) = \exp\{-\Omega(n^{1/3})\}.$$

Using the results of this paper, we can give a short alternate proof of this theorem.

Proof. Put $m = \lfloor n/4 \rfloor$. A well-known result due to Beigel [7] shows that $W(\mathrm{OMB}_m, cm^{1/3}) \geqslant \exp(cm^{1/3})$ for some absolute constant $c > 0$. Since the $(2m, m, \mathrm{OMB}_m)$-pattern matrix is a submatrix of $[f(x,y)]_{x,y}$, the proof is complete by Theorem 7.3. □

REMARK 10.2. The above proofs illustrate that the characterization of the discrepancy of pattern matrices in this paper (Theorem 7.3) is a substantial improvement on our earlier result (Theorem 7.4). In particular, the representation (10.1) makes it clear that $\deg_{\pm}(\mathrm{OMB}_n) = 1$ and therefore Theorem 7.4 cannot yield an upper bound better than $n^{-\Omega(1)}$ on the discrepancy of $\mathrm{OMB}_n(x \wedge y)$. Theorem 7.3, on the other hand, gives an exponentially better upper bound.

It is well-known [21, 22, 48] that the discrepancy of a function f implies a lower bound on the size of depth-2 majority circuits that compute f. Following [61], we record the consequences of Theorems 1.6 and 10.1 in this regard.

THEOREM 10.3. *Any majority vote of threshold gates that computes the function*

$$f(x,y) = \mathrm{MP}_m(x \vee y)$$

has size $\exp\{\Omega(m)\}$. *Analogously, any majority vote of threshold gates that computes the function*

$$f(x,y) = \mathrm{OMB}_n(x \wedge y)$$

has size $\exp\{\Omega(n^{1/3})\}$.

Proof. Analogous to the proof given in [61, §7]. □

11. Pattern matrices and the log-rank conjecture. In previous sections, we characterized various matrix-analytic and combinatorial properties of pattern matrices, including their classical and quantum communication complexity, discrepancy, approximate rank, and approximate trace norm. We conclude this study with another fact about pattern matrices. Specifically, we show that they satisfy the *log-rank conjecture* due to Lovász and Saks [44].

In a seminal paper, Mehlhorn and Schmidt [45] observed that the deterministic communication complexity of a sign matrix F satisfies $D(F) \geqslant \log \operatorname{rk} F$. The log-rank conjecture is that this lower bound is always tight up to a polynomial factor, i.e., $D(F) \leqslant (\log \operatorname{rk} F)^{O(1)} + O(1)$. Using the results of the previous sections, we can give a short proof of this hypothesis in the case of pattern matrices.

Theorem 11.1 (on the log-rank conjecture). *Let $f\colon \{0,1\}^t \to \{-1,+1\}$ be a given function, $d = \deg(f)$. Let F be the (n,t,f)-pattern matrix. Then*

$$\operatorname{rk} F \geqslant \left(\frac{n}{t}\right)^d \geqslant \exp\{\Omega(D(F)^{1/4})\}. \tag{11.1}$$

In particular, F satisfies the log-rank conjecture.

Proof. Since $\hat{f}(S) \neq 0$ for some set S with $|S| = d$, Theorem 4.3 implies that F has at least $(n/t)^d$ nonzero singular values. This settles the first inequality in (11.1).

Proposition 5.1 implies that $D(F) \leqslant O(\operatorname{dt}(f)\log(n/t))$, where $\operatorname{dt}(f)$ denotes the least depth of a decision tree for f. Nisan and Smolensky [10, Thm. 12] prove that $\operatorname{dt}(f) \leqslant 2\deg(f)^4$ for all f. Combining these two observations establishes the second inequality in (11.1). ☐

12. Related work. Shi and Zhu [63] independently obtained a result related to our lower bound (5.2) on bounded-error communication. Fix functions $f\colon \{0,1\}^n \to \{-1,+1\}$ and $g\colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$. Let $f \circ g^n$ denote the composition of f with n independent copies of g. More formally, the function $f \circ g^n\colon \{0,1\}^{nk} \times \{0,1\}^{nk} \to \{-1,+1\}$ is given by

$$(f \circ g^n)(x,y) = f\Big(\ g(x^{(1)},y^{(1)}),\ \ \ldots,\ \ g(x^{(n)},y^{(n)})\ \Big),$$

where $x = (x^{(1)},\ldots,x^{(n)}) \in \{0,1\}^{nk}$ and $y = (y^{(1)},\ldots,y^{(n)}) \in \{0,1\}^{nk}$. Shi and Zhu study the communication complexity of $f \circ g^n$. Their main result [63, Lem. 3.5] is that

$$Q^*_{1/3}(f \circ g^n) \geqslant \Omega(\deg_{1/3}(f)) \qquad \text{provided that} \qquad \rho(g) \leqslant \frac{\deg_{1/3}(f)}{2en},$$

where $\rho(g)$ is a new variant of discrepancy that the authors introduce. As an illustration, they re-prove a weaker version of Razborov's lower bounds in Theorem 1.3. In our terminology (Section 2.4), their proof also fits in the framework of the Klauck-Razborov generalized discrepancy method.

Shi and Zhu's result revolves around the quantity $\rho(g)$, which needs to be small. This poses two complications. First, the function g will generally need to depend on many variables, from $k = \Theta(\log n)$ to $k = n^{\Theta(1)}$, which weakens the final lower bounds on communication. For example, the lower bounds obtained in [63] for symmetric functions are polynomially weaker than optimal (Theorem 1.3).

A second complication, as the authors note, is that "estimating $\rho(g)$ is unfortunately difficult in general" [63, §4.1]. For example, re-proving Razborov's lower bounds reduces to estimating $\rho(g)$ for $g(x,y) = x_1y_1 \vee \cdots \vee x_ky_k$. Shi and Zhu accomplish this using Hahn matrices, an advanced tool that is the centerpiece of Razborov's own proof (Razborov's use of Hahn matrices is somewhat more demanding).

Our method avoids these complications altogether. For example, we prove (by taking $n = 2t$ in the pattern matrix method, Theorem 1.1) that

$$Q^*_{1/3}(f \circ g^n) \geqslant \Omega(\deg_{1/3}(f))$$

for any function $g\colon \{0,1\}^k \times \{0,1\}^k \to \{0,1\}$ such that the matrix $[g(x,y)]_{x,y}$ contains the following submatrix, up to permutations of rows and columns:

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

To illustrate, one can take g to be

$$g(x,y) \;=\; x_1y_1 \;\vee\; x_2y_2 \;\vee\; x_3y_3 \;\vee\; x_4y_4$$

or

$$g(x,y) \;=\; x_1y_1y_2 \;\vee\; \overline{x_1}\,y_1\overline{y_2} \;\vee\; x_2\,\overline{y_1}\,y_2 \;\vee\; \overline{x_2}\,\overline{y_1}\,\overline{y_2}.$$

In summary, there is a simple function g on $k = 2$ variables that works universally for all f. This means no technical conditions to check, such as $\rho(g)$, and no blow-up in the number of variables. As a result, we are able to re-prove Razborov's optimal lower bounds exactly. Moreover, the technical machinery of this paper is self-contained and disjoint from Razborov's proof.

A further advantage of the pattern matrix method is that it extends in a straightforward way to the multiparty model [41, 14, 15, 16, 6]. This extension depends on the fact that the rows of a pattern matrix are applications of the same function to different subsets of the variables. In the general context of block composition, it is unclear how to carry out this extension. Further details can be found in the survey [62].

These considerations do not diminish the technical merit of Shi and Zhu's method, which is of much interest. The proofs in [63] and this paper start out with the same duality transformation (Theorem 3.2) but diverge substantially from then on, which explains the differences in our results. Specifically, we introduce and analyze pattern matrices, while Shi and Zhu construct a much different family of matrices.

Acknowledgments. I would like to thank Adam Klivans, James Lee, Sasha Razborov, Yaoyun Shi, Avi Wigderson, and Ronald de Wolf for their feedback on an earlier version of this manuscript. I am thankful to Ronald de Wolf for his permission to include his remark on Theorem 1.1. It is also to Ronald that I owe my interest in quantum communication. This research was supported by Adam Klivans' NSF CAREER Award and NSF Grant CCF-0728536.

REFERENCES

[1] Scott Aaronson and Yaoyun Shi, *Quantum lower bounds for the collision and the element distinctness problems*, J. ACM, 51 (2004), pp. 595–605.

[2] Andris Ambainis, *Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range*, Theory of Computing, 1 (2005), pp. 37–46.

[3] Andris Ambainis, Leonard J. Schulman, Amnon Ta-Shma, Umesh V. Vazirani, and Avi Wigderson, *The quantum communication complexity of sampling*, SIAM J. Comput., 32 (2003), pp. 1570–1585.

[4] James Aspnes, Richard Beigel, Merrick L. Furst, and Steven Rudich, *The expressive power of voting polynomials*, Combinatorica, 14 (1994), pp. 135–148.

[5] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf, *Quantum lower bounds by polynomials*, J. ACM, 48 (2001), pp. 778–797.

[6] Paul Beame and Dang-Trinh Huynh-Ngoc, *Multiparty communication complexity and threshold circuit complexity of* AC^0, in Proc. of the 50th Symposium on Foundations of Computer Science (FOCS), 2009. To appear.

[7] Richard Beigel, *Perceptrons,* PP*, and the polynomial hierarchy*, Computational Complexity, 4 (1994), pp. 339–349.

[8] Harry Buhrman, Richard Cleve, Ronald de Wolf, and Christof Zalka, *Bounds for small-error and zero-error quantum algorithms*, in Proc. of the 40th Symposium on Foundations of Computer Science (FOCS), 1999, pp. 358–368.

[9] Harry Buhrman, Richard Cleve, and Avi Wigderson, *Quantum vs. classical communication and computation*, in Proc. of the 13th Symposium on Theory of Computing (STOC), 1998, pp. 63–68.

[10] Harry Buhrman and Ronald de Wolf, *Complexity measures and decision tree complexity: A survey*, Theor. Comput. Sci., 288 (2002), pp. 21–43.

[11] Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf, *On computation and communication with small bias*, in Proc. of the 22nd Conf. on Computational Complexity (CCC), 2007, pp. 24–32.

[12] Harry Buhrman and Ronald de Wolf, *Communication complexity lower bounds by polynomials*, in Proc. of the 16th Conf. on Computational Complexity (CCC), 2001, pp. 120–130.

[13] Amit Chakrabarti, Yaoyun Shi, Anthony Wirth, and Andrew Chi-Chih Yao, *Informational complexity and the direct sum problem for simultaneous message complexity*, in Proc. of the 42nd Symposium on Foundations of Computer Science (FOCS), 2001, pp. 270–278.

[14] Arkadev Chattopadhyay and Anil Ada, *Multiparty communication complexity of disjointness*, in Electronic Colloquium on Computational Complexity (ECCC), January 2008. Report TR08-002.

[15] Matei David and Toniann Pitassi, *Separating NOF communication complexity classes* RP *and* NP, in Electronic Colloquium on Computational Complexity (ECCC), February 2008. Report TR08-014.

[16] Matei David, Toniann Pitassi, and Emanuele Viola, *Improved separations between nondeterministic and randomized multiparty communication*, in Proc. of the 12th Intl. Workshop on Randomization and Computation (RANDOM), 2008, pp. 371–384.

[17] Ronald de Wolf. Personal communication, October 2007.

[18] Ronald A. DeVore and George G. Lorentz, *Constructive Approximation*, vol. 303, Springer-Verlag, Berlin, 1993.

[19] Yoav Freund, *Boosting a weak learning algorithm by majority*, Inf. Comput., 121 (1995), pp. 256–285.

[20] Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf, *Exponential separations for one-way quantum communication complexity, with applications to cryptography*, in Proc. of the 39th Symposium on Theory of Computing (STOC), 2007, pp. 516–525.

[21] Mikael Goldmann, Johan Håstad, and Alexander A. Razborov, *Majority gates vs. general weighted threshold gates*, Computational Complexity, 2 (1992), pp. 277–300.

[22] András Hajnal, Wolfgang Maass, Pavel Pudlák, Mario Szegedy, and György Turán, *Threshold circuits of bounded depth*, J. Comput. Syst. Sci., 46 (1993), pp. 129–154.

[23] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, New York, 1986.

[24] Peter Høyer and Ronald 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), 2002, pp. 299–310.

[25] Aleksandr Davidovich Ioffe and Vladimir Mikhailovich Tikhomirov, *Duality of convex functions and extremum problems*, Russ. Math. Surv., 23 (1968), pp. 53–124.

[26] Jeff Kahn, Nathan Linial, and Alex Samorodnitsky, *Inclusion-exclusion: Exact and approximate*, Combinatorica, 16 (1996), pp. 465–477.

[27] Bala Kalyanasundaram and Georg Schnitger, *The probabilistic communication complexity of set intersection*, SIAM J. Discrete Math., 5 (1992), pp. 545–557.

[28] Michael J. Kearns, Robert E. Schapire, and Linda Sellie, *Toward efficient agnostic learning*, Machine Learning, 17 (1994), pp. 115–141.

[29] Hartmut Klauck, *Lower bounds for quantum communication complexity*, in Proc. of the 42nd Symposium on Foundations of Computer Science (FOCS), 2001, pp. 288–297.

[30] Hartmut Klauck, *Lower bounds for quantum communication complexity*, SIAM J. Comput., 37 (2007), pp. 20–46.

[31] Hartmut Klauck, Ashwin Nayak, Amnon Ta-Shma, and David Zuckerman, *Interaction in quantum communication and the complexity of set disjointness*, in Proc. of the 33rd Symposium on Theory of Computing (STOC), 2001, pp. 124–133.

[32] Adam R. Klivans, Ryan O'Donnell, and Rocco A. Servedio, *Learning intersections and thresholds of halfspaces*, J. Comput. Syst. Sci., 68 (2004), pp. 808–840.

[33] Adam R. Klivans and Rocco A. Servedio, *Learning DNF in time* $2^{\tilde{O}(n^{1/3})}$, J. Comput. Syst. Sci., 68 (2004), pp. 303–318.

[34] Adam R. Klivans and Rocco A. Servedio, *Toward attribute efficient learning of decision lists and parities*, J. Machine Learning Research, 7 (2006), pp. 587–602.

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

[36] Adam R. Klivans and Alexander A. Sherstov, *Unconditional lower bounds for learning*

intersections of halfspaces, Machine Learning, 69 (2007), pp. 97–114.

[37] Matthias Krause and Pavel Pudlák, *On the computational power of depth-2 circuits with threshold and modulo gates*, Theor. Comput. Sci., 174 (1997), pp. 137–156.

[38] Matthias Krause and Pavel Pudlák, *Computing Boolean functions by polynomials and threshold circuits*, Comput. Complex., 7 (1998), pp. 346–370.

[39] I. Kremer, *Quantum communication*, master's thesis, Hebrew University, Computer Science Department, 1995.

[40] Eyal Kushilevitz and Noam Nisan, *Communication complexity*, Cambridge University Press, New York, 1997.

[41] Troy Lee and Adi Shraibman, *Disjointness is hard in the multi-party number-on-the-forehead model*, in Proc. of the 23rd Conf. on Computational Complexity (CCC), 2008, pp. 81–91.

[42] Nati Linial and Adi Shraibman, *Lower bounds in communication complexity based on factorization norms*, in Proc. of the 39th Symposium on Theory of Computing (STOC), 2007, pp. 699–708.

[43] Nathan Linial and Adi Shraibman, *Learning complexity vs. communication complexity*, in Proc. of the 23rd Conf. on Computational Complexity (CCC), 2008, pp. 53–63.

[44] László Lovász and Michael E. Saks, *Lattices, möbius functions and communication complexity*, in Proc. of the 29th Symposium on Foundations of Computer Science (FOCS), 1988, pp. 81–90.

[45] Kurt Mehlhorn and Erik Meineche Schmidt, *Las Vegas is better than determinism in VLSI and distributed computing*, in Proc. of the 14th Symposium on Theory of Computing (STOC), 1982, pp. 330–337.

[46] Marvin L. Minsky and Seymour A. Papert, *Perceptrons: An Introduction to Computational Geometry*, MIT Press, Cambridge, Mass., 1969.

[47] J. Myhill and W. H. Kautz, *On the size of weights required for linear-input switching functions*, IRE Trans. on Electronic Computers, 10 (1961), pp. 288–290.

[48] Noam Nisan, *The communication complexity of threshold gates*, in Combinatorics, Paul Erdős is Eighty, 1993, pp. 301–315.

[49] Noam Nisan and Mario Szegedy, *On the degree of Boolean functions as real polynomials*, Computational Complexity, 4 (1994), pp. 301–313.

[50] Ryan O'Donnell and Rocco A. Servedio, *New degree bounds for polynomial threshold functions*, in Proc. of the 35th Symposium on Theory of Computing (STOC), 2003, pp. 325–334.

[51] Ramamohan Paturi, *On the degree of polynomials that approximate symmetric Boolean functions*, in Proc. of the 24th Symposium on Theory of Computing (STOC), 1992, pp. 468–474.

[52] Ramamohan Paturi and Janos Simon, *Probabilistic communication complexity*, J. Comput. Syst. Sci., 33 (1986), pp. 106–123.

[53] Vladimir V. Podolskii, *Perceptrons of large weight*, in Proc. of the Second International Computer Science Symposium in Russia (CSR), 2007, pp. 328–336.

[54] Vladimir V. Podolskii, *A uniform lower bound on weights of perceptrons*, in Proc. of the Third International Computer Science Symposium in Russia (CSR), 2008, pp. 261–272.

[55] Ran Raz, *Fourier analysis for probabilistic communication complexity*, Comput. Complex., 5 (1995), pp. 205–221.

[56] Alexander A. Razborov, *On the distributional complexity of disjointness*, Theor. Comput. Sci., 106 (1992), pp. 385–390.

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

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

[59] Alexander A. Sherstov, *Approximate inclusion-exclusion for arbitrary symmetric functions*, in Proc. of the 23rd Conf. on Computational Complexity (CCC), 2008, pp. 112–123.

[60] Alexander A. Sherstov, *Halfspace matrices*, Comput. Complex., 17 (2008), pp. 149–178. Preliminary version in 22nd CCC, 2007.

[61] Alexander A. Sherstov, *Separating* AC^0 *from depth-2 majority circuits*, SIAM J. Comput., 38 (2009), pp. 2113–2129. Preliminary version in 39th STOC, 2007.

[62] Alexander A. Sherstov, *Communication lower bounds using dual polynomials*, Bulletin of the EATCS, 95 (2008), pp. 59–93.

[63] Yaoyun Shi and Yufan Zhu, *Quantum communication complexity of block-composed functions*. Available at `http://arxiv.org/abs/0710.0095v4`, 2008.

[64] Nikolai K. Vereshchagin, *Lower bounds for perceptrons solving some separation problems and oracle separation of* AM *from* PP, in Proc. of the Third Israel Symposium on Theory

of Computing and Systems (ISTCS), 1995, pp. 46–51.

[65] Ronald de Wolf, *A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions*, Quantum Information and Computation, 8 (2008), pp. 943–950.

[66] Andrew Chi-Chih Yao, *Some complexity questions related to distributive computing*, in Proc. of the 11th Symposium on Theory of Computing (STOC), 1979, pp. 209–213.

[67] Andrew Chi-Chih Yao, *Quantum circuit complexity*, in Proc. of the 34th Symposium on Foundations of Computer Science (FOCS), 1993, pp. 352–361.

Appendix A. On uniform approximation and sign-representation. The purpose of this appendix is to prove Theorem 2.5 on the representation of a Boolean function by real versus integer polynomials. Similar statements have been noted earlier by several authors [38, 11]. We derive our result by modifying a recent analysis due to Buhrman et al. [11, Cor. 1].

Theorem 2.5 (restated from p. 10). *Let $f\colon \{0,1\}^n \to \{-1,+1\}$ be given. Then for $d = 0, 1, \ldots, n$,*

$$\frac{1}{1-E(f,d)} \leqslant W(f,d) \leqslant \frac{2}{1-E(f,d)} \left\{ \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{d} \right\}^{3/2},$$

with the convention that $1/0 = \infty$.

Proof. One readily verifies that $W(f,d) = \infty$ if and only if $E(f,d) = 1$. In what follows, we focus on the complementary case when $W(f,d) < \infty$ and $E(f,d) < 1$.

For the lower bound on $W(f,d)$, fix integers λ_S with $\sum_{|S|\leqslant d} |\lambda_S| = W(f,d)$ such that the polynomial $p(x) = \sum_{|S|\leqslant d} \lambda_S \chi_S(x)$ satisfies $f(x) \equiv \operatorname{sgn} p(x)$. Then $1 \leqslant f(x)p(x) \leqslant W(f,d)$ and therefore

$$E(f,d) \leqslant \left\| f - \frac{1}{W(f,d)}\, p \right\|_\infty \leqslant 1 - \frac{1}{W(f,d)}.$$

To prove the upper bound on $W(f,d)$, fix any degree-d polynomial p such that $\|f - p\|_\infty = E(f,d)$. Define $\delta = 1 - E(f,d) > 0$ and $N = \sum_{i=0}^{d} \binom{n}{i}$. For a real t, let $\operatorname{rnd} t$ be the result of rounding t to the closest integer, so that $|t - \operatorname{rnd} t| \leqslant 1/2$. We claim that the polynomial

$$q(x) = \sum_{|S|\leqslant d} \operatorname{rnd}(M\hat{p}(S))\chi_S(x),$$

where $M = 3N/(4\delta)$, satisfies $f(x) \equiv \operatorname{sgn} q(x)$. Indeed,

$$\begin{aligned} \left| f(x) - \frac{1}{M} q(x) \right| &\leqslant |f(x) - p(x)| + \frac{1}{M} |Mp(x) - q(x)| \\ &\leqslant 1 - \delta + \frac{1}{M} \sum_{|S|\leqslant d} |M\hat{p}(S) - \operatorname{rnd}(M\hat{p}(S))| \\ &\leqslant 1 - \delta + \frac{N}{2M} \\ &< 1. \end{aligned}$$

It remains to examine the sum of the coefficients of q. We have:

$$\begin{aligned}\sum_{|S|\leqslant d} |\operatorname{rnd}(M\hat{p}(S))| &\leqslant \frac{1}{2}N + M \sum_{|S|\leqslant d} |\hat{p}(S)| \\ &\leqslant \frac{1}{2}N + M\left(N \mathop{\mathbf{E}}_{x}\left[p(x)^2\right]\right)^{1/2} \\ &\leqslant \frac{2N\sqrt{N}}{\delta},\end{aligned}$$

where the second step follows by an application of the Cauchy-Schwarz inequality and Parseval's identity (2.1). ☐