Unbounded-Error Communication Complexity of Symmetric Functions

ALEXANDER A. SHERSTOV

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

September 22, 2007

Abstract

The *sign-rank* of a real matrix M is the least rank of a matrix R in which every entry has the same sign as the corresponding entry of M. We determine the sign-rank of every matrix of the form $M = [\ D(|x \wedge y|)\ ]_{x,y}$, where $D : \{0, 1, \ldots, n\} \to \{-1, +1\}$ is given and x and y range over $\{0, 1\}^n$. Specifically, we prove that the sign-rank of M equals $2^{\Theta(k)}$, where k is the number of times D changes sign in $\{0, 1, \ldots, n\}$.

Put differently, we prove an optimal lower bound on the *unbounded-error communication complexity* of every symmetric function, i.e., a function of the form $f(x, y) = D(|x \wedge y|)$ for some D. The unbounded-error model is essentially the most powerful of all models of communication (both classical and quantum), and proving lower bounds in it is a substantial challenge. The only previous nontrivial lower bounds for this model appear in the groundbreaking work of Forster (2001) and its extensions. As corollaries to our result, we give new lower bounds for PAC learning and for threshold-of-majority circuits.

The technical content of our proof is diverse and features random walks on $\mathbb{Z}_2^n$, discrete approximation theory, the Fourier transform on $\mathbb{Z}_2^n$, linear-programming duality, and matrix analysis.

1 Introduction

The unbounded-error model, due to Paturi and Simon [27], is a rich and elegant model of communication. Fix a function $f : X \times Y \to \{0, 1\}$, where X and Y are some finite sets. Alice receives an input $x \in X$, Bob receives $y \in Y$, and their objective is to compute $f(x, y)$. To this end, they exchange bits through a shared communication channel according to a certain strategy, or *protocol*, that they establish ahead of time. Alice and Bob each have an unlimited private source of random bits which they can use in deciding what messages to send. Eventually, Bob concludes this process by sending Alice a single bit, which is taken to be the output of their joint computation.

Define the random variable $P(x, y) \in \{0, 1\}$ as the output bit when the parties receive inputs $x \in X$ and $y \in Y$. Alice and Bob's protocol is said to *compute* f if

$$\mathbf{Pr}[P(x, y) = f(x, y)] > \frac{1}{2} \qquad \text{for each } x \in X, \ y \in Y.$$

The above probability is, of course, over the private use of random bits by Alice and Bob. The *cost* of a given protocol is the worst-case number of bits exchanged on any input (x, y). The *unbounded-error communication complexity* of f, denoted $U(f)$, is the least cost of a protocol that computes f.

The unbounded-error model occupies a special place in the study of communication because it is more powerful than any other standard model (deterministic, nondeterministic, randomized, quantum with or without entanglement). More precisely, the unbounded-error complexity $U(f)$ can be only negligibly greater than the complexity of f in any other model—and often, $U(f)$ is exponentially smaller. We defer precise quantitative statements to Section 2.2. The power of the unbounded-error model resides in its very liberal success criterion: it suffices to produce the correct output with probability greater than $1/2$ (say, by an exponentially small amount). This contrasts with all other models, where the correct output is expected with probability at least $2/3$.

1.1 Motivation

The additional power of the unbounded-error model has the welcome consequence that proving communication lower bounds in it requires richer and more creative mathematical machinery. Furthermore, the resulting lower bounds will have implications that other communication models could not yield. Before we state our results, we take a moment to thoroughly motivate our work by reviewing these new possibilities unique to the unbounded-error model.

Circuit complexity.   Recall that a *threshold gate g* with Boolean inputs $x_1, \ldots, x_n$ is a function of the form $g(x) = \operatorname{sign}(a_1 x_1 + \cdots + a_n x_n - \theta)$, for some fixed reals $a_1, \ldots, a_n, \theta$. Thus, a threshold gate generalizes the familiar *majority* gate. A major unsolved problem in computational complexity is to exhibit a Boolean function that requires a depth-2 threshold circuit of superpolynomial size.

Communication complexity has been crucial to the progress on this problem. Using randomized communication complexity, many explicit functions have been found [9, 24, 33, 34] that require depth-2 majority circuits of exponential size. Via the reductions due to Goldman et al. [8], these lower bounds remain valid for the broader class of *majority-of-threshold circuits.* This solves an important special case of the general problem. The unbounded-error model solves another important special case [6]: it supplies exponential lower bounds against *threshold-of-majority circuits,* i.e., circuits with a threshold gate at the top that receives inputs from majority gates. To our knowledge, the unbounded-error model is currently the only means to prove lower bounds against threshold-of-majority circuits.

Sign-rank and rigidity.   Unlike other models of communication, the unbounded-error model has a particularly natural matrix-analytic formulation. Fix a real matrix $M = [M_{ij}]$ without zero entries. The *sign-rank* of M, denoted $dc(M)$, is defined as the least rank of a matrix $A = [A_{ij}]$ with $M_{ij} A_{ij} > 0$ for all i, j. In other words, sign-rank measures the sensitivity of the rank of M when its entries undergo sign-preserving perturbations. The sensitivity of rank is an important and difficult subject in complexity theory. For example, much work has focused on the closely related concept of *matrix rigidity* [12, 21].

On the surface, unbounded-error complexity and sign-rank seem unrelated. In reality, they are equivalent notions! More specifically, let $f : X \times Y \to \{0, 1\}$ be a given function. Consider its communication matrix $M = [(-1)^{f(x,y)}]_{x \in X, y \in Y}$. Paturi and Simon [27] showed that

$$U(f) = \log dc(M) \pm O(1).$$

Thus, unbounded-error complexity embodies a fundamental question from matrix analysis, with close ties to complexity theory.

PAC learning.   In a seminal paper [35], Valiant formulated the *probably approximately correct* (PAC) model of learning, now the primary model in computational learning theory. Let C be a given *concept class*, i.e., a set of functions $\{0, 1\}^n \to \{0, 1\}$. The learner in this model receives training examples

$$(x^1, f(x^{(1)})), \quad (x^2, f(x^{(2)})), \quad \ldots, \quad (x^m, f(x^{(m)})),$$

where $f \in C$ is an unknown function and $x^{(1)}, x^{(2)}, \ldots, x^{(m)} \in \{0, 1\}^n$ are sampled independently from some unknown distribution μ. For every choice of f and μ, the learner must produce a hypothesis $h : \{0, 1\}^n \to \{0, 1\}$ that closely approximates the unknown function: $\mathbf{E}_{x \sim \mu}[h(x) \neq f(x)] \leqslant \epsilon$. The objective is to find h efficiently.

Research has shown that PAC learning is surprisingly difficult. Indeed, the problem remains unsolved for such natural concept classes as DNF formulas of polynomial size and intersections of two halfspaces, whereas hardness results and lower bounds are abundant [4, 13, 14, 16–18]. There is, however, an important case when efficient PAC learning is straightforward. Specifically, let C be a given concept class. For notational convenience, view the functions in C as mappings $\{0, 1\}^n \to \{-1, +1\}$ rather than $\{0, 1\}^n \to \{0, 1\}$. The *dimension complexity* of C, denoted $\mathrm{dc}(C)$, is the least r for which there are functions $\phi_1, \ldots, \phi_r : \{0, 1\}^n \to \mathbb{R}$ such that every $f \in C$ is expressible in the form

$$f(x) \equiv \mathrm{sign}(a_1 \phi_1(x) + \cdots + a_r \phi_r(x))$$

for some reals $a_1, \ldots, a_r$. There is a simple and well-known algorithm [15], based on linear programming, that PAC learns C in time polynomial in $\mathrm{dc}(C)$. To relate this discussion to sign-rank (or equivalently, to unbounded-error complexity), let $M_C \stackrel{\text{def}}{=} [f(x)]_{f \in C, \, x \in \{0,1\}^n}$ be the characteristic matrix of C. A moment's reflection reveals that

$$\mathrm{dc}(C) = \mathrm{dc}(M_C),$$

i.e., the dimension complexity of a concept class is precisely the sign-rank of its characteristic matrix.

Thus, the study of sign-rank yields nontrivial PAC learning algorithms. In particular, the best known algorithm for learning polynomial-size DNF formulas (Klivans & Servedio, 2001) was obtained precisely by placing a $2^{\tilde{O}(n^{1/3})}$ upper bound on the dimension complexity of that concept class. Furthermore, this dimension-complexity method actually represents the state of the art in computational learning theory: whatever is known to be efficiently PAC learnable has low dimension complexity—with the only exception of low-degree polynomials over a finite field, which are trivial to learn but have high dimension complexity (Forster 2001). In summary, dimension complexity is an important notion in computational learning theory.

1.2 Our Result

As we have discussed, the unbounded-error model has immediate applications to circuit complexity, matrix analysis, and learning theory, in addition to its intrinsic appeal as a model of communication. Despite this motivation, progress in understanding unbounded-error complexity has been slow and difficult. It is only recently that the first nontrivial lower bound was proved (Forster 2001) on the unbounded-error complexity of an explicit function. Forster's proof has since seen several extensions and refinements [6, 7, 20]. We are not aware of any other progress on unbounded-error complexity.

In this paper, we determine the unbounded-error complexity of a natural class of functions that was beyond the reach of the existing techniques. Specifically, we study functions $f : \{0, 1\}^n \to \{0, 1\}$ of the form

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

where $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ is a given predicate. Abbreviate $U(D) \stackrel{\text{def}}{=} U(f)$. Prior to our work, Forster showed that $U(D) = \Theta(n)$ for the *parity predicate* $D(t) \equiv (t \bmod 2)$. The unbounded-error complexity of general D, however, remained unsettled.

We settle the unbounded-error complexity of *every D*. Let $\deg(D)$ stand for the number of times D changes sign in $\{0, 1, \ldots, n\}$, i.e.,

$$\deg(D) \stackrel{\text{def}}{=} |\{i = 1, 2, \ldots, n : D(i) \neq D(i-1)\}|.$$

We prove:

Theorem 1.1 (Main Result). *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be given. Then*

$$U(D) = \tilde{\Theta}(\deg(D)),$$

where the $\tilde{\Theta}$ notation suppresses $\log n$ factors.

As explained in Section 1.1, this result implies lower bounds for PAC learning and for threshold-of-majority circuits. Since they follow from Theorem 1.1 as immediate corollaries, we defer their statements and proofs to the final version of the paper.

The upper bound in Theorem 1.1 has a short, first-principles proof. The lower bound, on the other hand, is rather nontrivial and has required us to use a variety of techniques (random walks on $\mathbb{Z}_2^n$, discrete approximation theory, the Fourier transform on $\mathbb{Z}_2^n$, linear-programming duality, and matrix analysis). We discuss our proof in greater detail next.
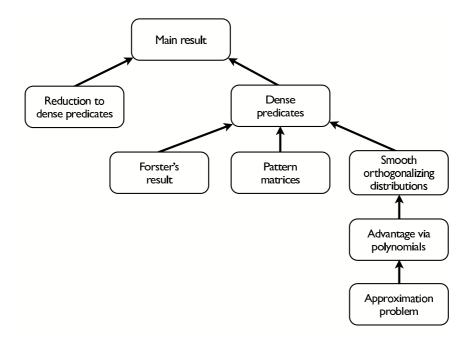
Figure 1: Proof outline.

1.3 Our Techniques

Figure 1 schematically illustrates our proof. As a first step, we reduce the original problem to one that is much smaller and more structured. Specifically, we reduce the overall task to analyzing what we call *dense* predicates. These are predicates that change value $\tilde{\Theta}(n)$ times and at roughly regular intervals. Such predicates behave more predictably and are amenable to our methods, whereas arbitrary predicates are not. The reduction works as follows. Under the assumption that a given predicate D has complexity $\tilde{O}(\deg(D))$, we use random walks on $\mathbb{Z}_2^n$ to infer the existence of some *dense* predicate with low complexity. The remaining part of the paper proves that this is an impossibility, i.e., every dense predicate has unbounded-error complexity $\tilde{\Theta}(n)$.

This leaves us with the challenge of analyzing the sign-rank of $F = [(-1)^{D(|x \wedge y|)}]$, the communication matrix of a given dense predicate. To this end, we combine several distinct ideas. The first of these is Forster's generalized result [6]. Applied to our setting, it states that the sign-rank of F is proportional to the quantity

$$\frac{\min_{x,y} |P_{xy}|}{\|P \circ F\|},$$

5

where ∘ denotes entrywise multiplication and $P = [P_{xy}]$ is any matrix whose entries are positive and sum to 1. In other words, we need to prove the existence of a matrix P with large entries that leads to a small spectral norm $\|P \circ F\|$.

To exhibit P with such properties, we use *pattern matrices*. These matrices arose in two earlier works by the author [32, 34], where they proved useful in obtaining strong lower bounds on communication. Their purpose in this paper is to reduce the search for P to a search for a *smooth orthogonalizing distribution* for the predicate D. This informal term refers to a distribution on $\{0, 1\}^n$ that does not put too little weight on any point (the *smooth* part) and under which $(-1)^{D(x_1+\cdots+x_n)}$ is approximately orthogonal to all low-degree parity functions (the *orthogonalizing* part).

To find a smooth orthogonalizing distribution, we apply linear-programming duality and work in the dual space instead. The dual problem turns out to be that of bounding the advantage to which a low-degree univariate polynomial can compute D, in a certain technical sense. We reformulate this new question as a discrete approximation problem and solve the latter from scratch, using fundamentals of approximation theory. Consolidating these various ingredients establishes our main result.

Organization. Section 2 reviews the necessary technical background. Section 3 opens the proof with the reduction to dense predicates. Section 4 solves a certain problem in discrete approximation. Section 5 translates this approximation result, via linear-programming duality and the Fourier transform, into an existence proof of smooth orthogonalizing distributions for every dense predicate. Section 6 combines the above ingredients to give the final lower bounds on unbounded-error complexity.

2 Preliminaries

This section provides the necessary technical background. We start by describing our notation and reviewing some standard preliminaries in Section 2.1. A detailed review of the unbounded-error model of communication is offered in Section 2.2, along with relevant previous work. Finally, Section 2.3 examines an essential ingredient of our proof, the pattern matrices.

2.1 Notation and Standard Preliminaries

A *Boolean function* is a mapping $X \rightarrow \{0, 1\}$, where X is a finite set. Typical cases are $X = \{0, 1\}^n$ and $X = \{0, 1\}^n \times \{0, 1\}^n$. The notation $[n]$ stands for the set

$\{1, 2, \ldots, n\}$. Throughout this manuscript, "log" refers to the logarithm to base 2. The symbol

$$P_k$$

refers to the family of all univariate polynomials of degree at most k.

For $x \in \{0, 1\}^n$, we write

$$|x| \quad \overset{\text{def}}{=} \quad |\{i : x_i = 1\}| \;=\; x_1 + x_2 + \cdots + x_n.$$

For $x, y \in \{0, 1\}^n$, the notation $x \wedge y$ refers as usual to the component-wise AND of x and y. In particular, $|x \wedge y|$ stands for the number of positions where x and y both have a 1. At several places in this manuscript, it will be important to distinguish between addition over the reals and addition over GF(2). To avoid any confusion, we reserve the operator $+$ for the former and $\oplus$ for the latter.

Random walks on $\mathbb{Z}_2^n$ play an important role in this work. In particular, it will be helpful to recall the following fact.

Proposition 2.1 (Folklore). *For an integer $T \geqslant 1$, let $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_T \in \{0, 1\}$ be independent random variables, each taking on 1 with probability p. Then*

$$\mathbf{E}\left[\mathbf{b}_1 \oplus \mathbf{b}_2 \oplus \cdots \oplus \mathbf{b}_T\right] = \frac{1}{2} - \frac{1}{2}(1 - 2p)^T.$$

Proof. Straightforward by induction on T. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Predicates. A *predicate* is a mapping $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. We say that a *value change* occurs at index $t \in \{1, 2, \ldots, n\}$ if $D(t) \neq D(t - 1)$. The *degree* of D, denoted $\deg(D)$, is the total number of value changes of D. For example, the familiar predicate $\text{PARITY}(t) \equiv (t \bmod 2)$ has degree n, whereas a constant predicate has degree 0. It is not hard to show that $\deg(D)$ is the least degree of a real univariate polynomial p such that

$$\text{sign}(p(t)) = (-1)^{D(t)} \qquad \text{for } t = 0, 1, \ldots, n,$$

hence the term *degree*. Finally, given two predicates $D_1, D_2 : \{0, 1, \ldots, n\} \to \{0, 1\}$, their XOR is the predicate $D_1 \oplus D_2 : \{0, 1, \ldots, n\} \to \{0, 1\}$ defined by

$$(D_1 \oplus D_2)(t) \quad \overset{\text{def}}{=} \quad D_1(t) \oplus D_2(t).$$

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

$$[F(i, j)]_{i \in I, \; j \in J},$$

where I and J are some index sets. In specifying matrices, we will use the symbol $*$ for entries whose values are irrelevant, as in the proofs of Lemmas 3.2 and 3.5. Recall that the *spectral norm* of a matrix $A \in \mathbb{R}^{m \times n}$ is given by

$$\|A\| \stackrel{\text{def}}{=} \max_{x \in \mathbb{R}^n, \; \|x\|_2 = 1} \|Ax\|_2,$$

where $\| \cdot \|_2$ is the Euclidean norm on vectors.

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

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

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

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

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

Proposition 2.2. *Let $f : \{0, 1\}^n \to \mathbb{R}$ be given. Then*

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

Symmetric functions. Denote the group of permutations $[n] \to [n]$ by S_n. A function $\phi : \{0, 1\}^n \to \mathbb{R}$ is called *symmetric* if $\phi(x)$ is uniquely determined by $x_1 + \cdots + x_n$. Equivalently, ϕ is symmetric if

$$\phi(x) = \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

for every $x \in \{0, 1\}^n$ and every $\sigma \in S_n$. Observe that for every $\phi : \{0, 1\}^n \to \mathbb{R}$ (symmetric or not), the derived function

$$\phi_{\text{sym}}(x) \stackrel{\text{def}}{=} \frac{1}{n!} \sum_{\sigma \in S_n} \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$$

is symmetric. The symmetric functions on $\{0, 1\}^n$ are intimately related to univariate polynomials, as demonstrated by Minsky and Papert's *symmetrization argument:*

Proposition 2.3 (Minsky & Papert [22]). *Let $\phi : \{0, 1\}^n \to \mathbb{R}$ be symmetric with $\hat{\phi}(S) = 0$ for $|S| > r$. Then there is a polynomial $p \in P_r$ with*

$$\phi(x) = p(x_1 + \cdots + x_n) \qquad \text{for all } x \in \{0, 1\}^n.$$

Minsky and Papert's observation has seen numerous uses in the literature [1,25,26].

2.2 The Unbounded-Error Model of Communication

We continue the review started in the Introduction.

Readers with background in communication complexity will note that the unbounded-error model is exactly the same as the *private-coin randomized model* [19, Chap. 3], with one exception: in the latter case the correct answer is expected with probability at least $2/3$, whereas in the former case the success probability need only *exceed* $1/2$ (say, by an exponentially small amount). This difference has far-reaching implications. For example, the fact that the parties in the unbounded-error model do not have a *shared* source of random bits is crucial: allowing shared randomness would make the complexity of every function a constant, as one can easily verify. By contrast, introducing shared randomness into the randomized model has minimal impact on the complexity of any given function [23].

As one might expect, the weaker success criterion in the unbounded-error model has a drastic impact on the complexity of certain functions. For example, the well-known *disjointness* function on n-bit strings has complexity $O(\log n)$ in the

9

unbounded-error model and $\Omega(n)$ in the randomized model [11, 29]. Furthermore, explicit functions are known [2, 31] with unbounded-error complexity $O(\log n)$ that require $\Omega(\sqrt{n})$ communication in the randomized model to even achieve advantage $2^{-\sqrt{n}/5}$ over random guessing.

More generally, the unbounded-error complexity of a function $f : X \times Y \to \{0, 1\}$ is never much more than its complexity in the other standard models. For example, it is not hard to see that

$$U(f) \leqslant \min\{N^0(f), N^1(f)\} + O(1)$$
$$\leqslant D(f) + O(1),$$

where D, N^0, and N^1 refer to communication complexity in the *deterministic,* 0-*nondeterministic,* and 1-*nondeterministic* models, respectively. Continuing,

$$U(f) \leqslant R_{1/3}(f) + O(1)$$
$$\leqslant O\left(R_{1/3}^{\mathrm{pub}}(f) + \log\log\left[|X| + |Y|\right]\right),$$

where $R_{1/3}$ and $R_{1/3}^{\mathrm{pub}}$ refer to the *private-* and *public-coin randomized* models, respectively. As a matter of fact, one can show that

$$U(f) \leqslant O\left(Q_{1/3}^*(f) + \log\log\left[|X| + |Y|\right]\right),$$

where $Q_{1/3}^*$ refers to the *quantum model with prior entanglement.* An identical inequality is clearly valid for the quantum model *without* prior entanglement. See [3, 19] for rigorous definitions of these various models; our sole intention was to point out that the unbounded-error model is at least as powerful.

Unlike other models of communication complexity, the unbounded-error model has a particularly natural interpretation in matrix-analytic terms. Specifically, let $M = [M_{ij}]$ be a real matrix without zero entries. Define the *sign-rank* of M, denoted $\mathrm{dc}(M)$, by:

$$\mathrm{dc}(M) \overset{\mathrm{def}}{=} \min_A \{\mathrm{rank}\, A : M_{ij}A_{ij} > 0 \quad \text{for all } i, j\}.$$

In words, $\mathrm{dc}(M)$ is the least rank of a real matrix A whose entries each have the same sign as the corresponding entry of M. A term equivalent to sign-rank is *dimension complexity,* hence the notation $\mathrm{dc}(M)$. Paturi and Simon (1986) made the following important observation.

Theorem 2.4 (Paturi and Simon [27, Thm. 2]). *Let X, Y be finite sets and $f : X \times Y \to \{0, 1\}$ a given function. Put $M = [(-1)^{f(x,y)}]_{x \in X, y \in Y}$. Then*

$$U(f) = \log \mathrm{dc}(M) \pm O(1).$$

Paturi and Simon's original observation concerned $X = Y = \{0, 1\}^n$, but their proof readily extends to arbitrary sets. In words, the unbounded-error complexity of a function essentially equals the logarithm of the sign-rank of its communication matrix. This equivalence is very helpful: sometimes it is more convenient to reason in terms of communication protocols, and sometimes the matrix formulation offers more insight.

The power of the unbounded-error model arguably makes it the most challenging model in which to prove communication lower bounds. In a breakthrough result, Forster [5] has recently proved the first nontrivial lower bound in the unbounded-error model for an explicit function. (By contrast, hard functions have long been known [3, 19] for all other communication models.) Forster's proof generalizes to yield the following result, which serves as a crucial starting point for our work.

Theorem 2.5 (Forster et al. [6, Thm. 3]). *Let X, Y be finite sets and $M = [M_{xy}]_{x \in X, y \in Y}$ a real matrix without zero entries. Then*

$$\mathrm{dc}(M) \geqslant \frac{\sqrt{|X|\,|Y|}}{\|M\|} \, \min_{x,y} |M_{xy}|.$$

We close this overview by discussing some closure properties of the unbounded-error model. Given functions $f, g : X \times Y \to \{0, 1\}$, recall that their XOR is the function $f \oplus g : X \times Y \to \{0, 1\}$ defined by

$$(f \oplus g)(x, y) \quad \overset{\mathrm{def}}{=} \quad f(x, y) \oplus g(x, y).$$

We have:

Proposition 2.6 (Folklore). *Let $f, g : X \times Y \to \{0, 1\}$ be arbitrary. Then*

$$U(f \oplus g) \leqslant U(f) + U(g).$$

Proof. Alice and Bob can evaluate f and g individually and output the XOR of the two answers. It is straightforward to verify that this strategy is correct with probability greater than $1/2$. $\qquad\square$

In what follows, we will be interested primarily in the complexity of predicates $D : \{0, 1, \ldots, n\} \to \{0, 1\}$. Specifically, we define $U(D)$ to be the unbounded-error communication complexity of the function $f : \{0, 1\}^n \times \{0, 1\}^n \to \{0, 1\}$ given by

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

11

2.3 Pattern Matrices

An important ingredient of this work is a certain family of real matrices that we call *pattern matrices.* They arose in two earlier works by the author [32,34] and proved useful in obtaining strong lower bounds on communication. Relevant definitions and results from [32] follow.

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

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

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

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

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

Definition 2.7 (Pattern matrix). For $\phi : \{0, 1\}^t \to \mathbb{R}$, the (n, t, ϕ)-*pattern matrix* is the real matrix A given by

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

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

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

The author has recently conducted [32] a complete and exact spectral analysis of pattern matrices. All we will need is the following expression for their spectral norm.

Theorem 2.8 (Sherstov [32, Thm. 4.3]). *Let* $\phi : \{0, 1\}^t \to \mathbb{R}$ *be given. Let A be the* (n, t, ϕ)-*pattern matrix. Then*

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

3 Reduction to Dense Predicates

For a predicate D, recall that $U(D)$ denotes its unbounded-error communication complexity. Let

$$U(n, k)$$

stand for the minimum $U(D)$ over all predicates $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ with $\deg(D) = k$. In this notation, our ultimate goal will be to show that $U(n, k) = \tilde{\Omega}(k)$. This section takes a step in that direction. First, we reduce the task of analyzing $U(n, k)$ to that of analyzing $U(n, \lceil \alpha n \rceil)$, where $\alpha \geqslant 1/4$. This focuses our efforts on high-degree predicates. We then further reduce the problem to *dense* predicates, i.e., high-degree predicates that change value at more or less even intervals in $\{0, 1, \ldots, n\}$. These reductions are essential because dense predicates behave more predictably and are much easier to analyze than arbitrary predicates. Dense predicates will be the focus of all later sections.

We start with some preparatory work (Section 3.1) and obtain our reductions in the two subsections that follow (Sections 3.2 and 3.3).

3.1 Preliminary Notions

An obvious representation of a predicate $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ is the vector $(D(0), D(1), \cdots, D(n))$. Unfortunately, this representation is poorly suited to analyzing the number of value changes of D. We therefore start by establishing a more convenient representation. For $i = 0, 1, \ldots, n$, define the predicate

$$T_i(t) \quad \overset{\text{def}}{=} \quad \begin{cases} 1 & \text{if } t \geqslant i, \\ 0 & \text{otherwise.} \end{cases}$$

A moment's reflection reveals that every predicate $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ can be uniquely expressed in the form

$$D = \bigoplus_{i \in S} T_i$$

for some set $S \subseteq \{0, 1, \ldots, n\}$. With this in mind, we define the *characteristic vector* of D to be the characteristic vector of S, i.e., the vector $v = (v_0, v_1, \ldots, v_n)$ given by

$$v_i \quad \overset{\text{def}}{=} \quad \begin{cases} 1 & \text{if } i \in S, \\ 0 & \text{otherwise.} \end{cases}$$

The advantage of this representation is that it allows us to conveniently express the number of value changes of D:

$$\deg(D) = |S \cap \{1, \ldots, n\}| = v_1 + \cdots + v_n,$$

13

as one can easily verify.

We will make a few more simple but useful observations. If D_1 and D_2 are predicates with characteristic vectors $v^{(1)}$ and $v^{(2)}$, then $D_1 \oplus D_2$ has characteristic vector $v^{(1)} \oplus v^{(2)}$. Finally, given a predicate $D : \{0, 1, \ldots, n\} \rightarrow \{0, 1\}$, consider a derived predicate $D' : \{0, 1, \ldots, m\} \rightarrow \{0, 1\}$ given by

$$D'(t) \equiv D(t + \Delta),$$

where $m \geqslant 1$ and $\Delta \geqslant 0$ are fixed integers with $m + \Delta \leqslant n$. Then the characteristic vectors v and v' of D and D', respectively, are related as follows:

$$v' = (v_0 \oplus \cdots \oplus v_\Delta, v_{\Delta+1}, \cdots, v_{\Delta+m}) \in \{0, 1\}^{m+1}.$$

From the standpoint of communication complexity, D' can be computed by hard-wiring some inputs to a protocol for D:

$$D'\left(\left|x_1 x_2 \ldots x_m \quad \bigwedge \quad y_1 y_2 \ldots y_m\right|\right)$$
$$= D\left(\left|x_1 x_2 \ldots x_m 1^\Delta 0^{n-m-\Delta} \quad \bigwedge \quad y_1 y_2 \ldots y_m 1^\Delta 0^{n-m-\Delta}\right|\right).$$

Therefore,

$$U(D') \leqslant U(D).$$

3.2 Reduction from Arbitrary to High-Degree Predicates

We start with a technical lemma. Consider a Boolean vector $v = (v_1, v_2, \ldots, v_n)$. We show that there is a subvector $(v_i, v_{i+1}, \ldots, v_j)$ that is reasonably far from both endpoints of v and yet contains many of the "1" bits present in v.

Lemma 3.1. *Let* $v \in \{0, 1\}^n$, $v \neq 0^n$. *Put* $k \stackrel{\text{def}}{=} v_1 + \cdots + v_n$. *Then there are indices* i, j *with* $i \leqslant j$ *such that*

$$v_i + \cdots + v_j \geqslant \frac{1}{14} \frac{k}{1 + \log(n/k)} \tag{3.1}$$

and

$$\min\{i - 1, n - j\} \geqslant j - i. \tag{3.2}$$

Proof. By symmetry, we can assume that $v_1 + v_2 + \cdots + v_m \geqslant \frac{1}{2}k$ for some index $m \leqslant \lceil n/2 \rceil$. Let $\alpha \in (0, \frac{1}{2})$ be a parameter to be fixed later. Let $T \geqslant 0$ be the smallest integer such that

$$v_1 + v_2 + \cdots + v_{\lfloor m/2^T \rfloor} < (1 - \alpha)^T (v_1 + v_2 + \cdots + v_m).$$

14

Clearly, $T \geqslant 1$. Since $v_1 + v_2 + \cdots + v_{\lfloor m/2^T \rfloor} \leqslant m/2^T$, we further obtain

$$1 \leqslant T \leqslant 1 + \frac{1 + \log(n/k)}{\log(2 - 2\alpha)}.$$

Now,

$$v_{\lfloor m/2^T \rfloor + 1} + \cdots + v_{\lfloor m/2^{T-1} \rfloor} = \underbrace{(v_1 + \cdots + v_{\lfloor m/2^{T-1} \rfloor})}_{\geqslant (1-\alpha)^{T-1}(v_1 + v_2 + \cdots + v_m)} - \underbrace{(v_1 + \cdots + v_{\lfloor m/2^T \rfloor})}_{<(1-\alpha)^T(v_1 + v_2 + \cdots + v_m)}$$

$$> \frac{1}{2}\alpha(1 - \alpha)^{T-1}k$$

$$\geqslant \frac{1}{2}\alpha(1 - \alpha(T - 1))k$$

$$\geqslant \frac{1}{2}\alpha\left(1 - \alpha \cdot \frac{1 + \log(n/k)}{\log(2 - 2\alpha)}\right)k. \tag{3.3}$$

Set $\alpha = 0.23/(1 + \log(n/k))$, $i = \lfloor m/2^T \rfloor + 1$, and $j = \lfloor m/2^{T-1} \rfloor$. Then one easily verifies (3.2), while (3.1) is immediate from (3.3). □

We are now ready to prove the desired reduction to high-degree predicates. Throughout this proof, we will freely use the preliminary notions of Section 3.1, often without mention.

Lemma 3.2 (Reduction from arbitrary to high-degree predicates). *For all integers* n, k *with* $1 \leqslant k \leqslant n$,

$$U(n, k) \geqslant \frac{5}{6} K \min_{\substack{m = K, \ldots, n, \\ 1/4 \leqslant \alpha \leqslant 1}} \left\{ \frac{1}{m} U(m, \lceil \alpha m \rceil) \right\},$$

where

$$K \stackrel{\text{def}}{=} \left\lceil \frac{1}{14} \frac{k}{1 + \log(n/k)} \right\rceil.$$

Proof. Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be any predicate with $\deg(D) = k$. Let $v = (v_0, v_1, \ldots, v_n)$ be the characteristic vector of D. Apply Lemma 3.1 to $(v_1, \ldots, v_n)$ and let i, j be the resulting indices ($i \leqslant j$). Put

$$m \stackrel{\text{def}}{=} j - i + 1.$$

Since $v_i + \cdots + v_j \geqslant K$, we have

$$K \leqslant m \leqslant n. \tag{3.4}$$

15

Define predicates $D^{-(m-1)}, \cdots, D^0, \cdots, D^{m-1}$, each a mapping $\{0, 1, \ldots, m\} \to \{0, 1\}$, by:

$$D^r(t) \equiv D(t + i - 1 + r) \qquad \text{for } r = -(m-1), \ldots, (m-1).$$

Then (3.2) shows that each each of these predicates can be computed by taking a protocol for D and fixing all but the first m variables to appropriate values. Thus,

$$U(D) \geqslant U(D^r) \qquad \text{for } r = -(m-1), \ldots, (m-1). \tag{3.5}$$

The characteristic vector of D^0 is $(*, v_i, \ldots, v_j)$ for some $* \in \{0, 1\}$, which means that $\deg(D^0) = v_i + \cdots + v_j$. If $\deg(D^0) > m/2$, then the theorem is true for D in view of (3.4) and (3.5). Thus, we can assume the contrary:

$$K \leqslant v_i + \cdots + v_j \leqslant \frac{1}{2}m. \tag{3.6}$$

If we write the characteristic vectors of $D^{-(m-1)}, \ldots, D^{m-1}$ one after another as row vectors, we obtain the following matrix A:

$$A = \begin{bmatrix}
* & * & * & * & * & \cdots & * & * & * & v_i \\
* & * & * & * & * & \cdots & * & * & v_i & v_{i+1} \\
* & * & * & * & * & \cdots & * & v_i & v_{i+1} & v_{i+2} \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
* & v_i & v_{i+1} & v_{i+2} & v_{i+3} & \cdots & v_{j-3} & v_{j-2} & v_{j-1} & v_j \\
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots \\
* & v_{j-2} & v_{j-1} & v_j & * & \cdots & * & * & * & * \\
* & v_{j-1} & v_j & * & * & \cdots & * & * & * & * \\
* & v_j & * & * & * & \cdots & * & * & * & *
\end{bmatrix}.$$

Let T be a suitably large integer to be named later, and let $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \ldots, \mathbf{u}^{(T)}$ be independent random vectors, each selected uniformly from among the rows of A. Put

$$\mathbf{u} \overset{\text{def}}{=} \mathbf{u}^{(1)} \oplus \mathbf{u}^{(2)} \oplus \cdots \oplus \mathbf{u}^{(T)}.$$

We will index the columns of A and the components of all these vectors by $0, 1, \ldots, m$ (left to right). Let p_r stand for the fraction of 1s in the rth column of A. Every column of A, except the zeroth, contains $v_i, \ldots, v_j$ and some $m - 1$ additional values. One infers from (3.6) that

$$\frac{K}{2m} \leqslant p_r \leqslant \frac{3}{4} \qquad (r = 1, \ldots, m). \tag{3.7}$$

16

Therefore,

$$
\begin{aligned}
\mathbf{E}\left[(\mathbf{u})_1 + \cdots + (\mathbf{u})_m\right] &= \sum_{r=1}^{m} \mathbf{E}\left[(\mathbf{u}^{(1)})_r \oplus \cdots \oplus (\mathbf{u}^{(T)})_r\right] \\
&= \sum_{r=1}^{m} \left(\frac{1}{2} - \frac{1}{2}(1 - 2p_r)^T\right) \qquad \text{by Proposition 2.1} \\
&\geqslant \frac{1}{2}m\left(1 - \frac{1}{e^{TK/m}}\right) \qquad \text{by (3.6), (3.7).}
\end{aligned}
$$

Fix $T \stackrel{\text{def}}{=} \lceil(\ln 2)m/K\rceil$. Then by the last calculation, there is a vector $u = (u_0, u_1, \ldots, u_m)$ that satisfies $u_1 + \cdots + u_m \geqslant m/4$ and is the XOR of some T rows of A. In other words, there is a predicate $D^{\oplus} : \{0, 1, \ldots, m\} \to \{0, 1\}$ that satisfies $\deg(D^{\oplus}) \geqslant m/4$ and is the XOR of some $T \leqslant \frac{6m}{5K}$ predicates from among $D^{-(m-1)}, \ldots, D^{m-1}$. This completes the proof in view of (3.5) and Proposition 2.6. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

3.3 Reduction from High-Degree to Dense Predicates

The proof of Lemma 3.2 made critical use of random walks on $\{0, 1\}^n$. The work in this section also relies heavily on random walks, except the argument is now more involved. In particular, we will need the following lemma that bounds the mixing time of a random walk.

Lemma 3.3 (Razborov [28, Lem. 1][1]). *Fix a probability distribution μ on $\{0, 1\}^n$. Let $\{v^{(1)}, v^{(2)}, \ldots, v^{(n)}\}$ be a basis for $\{0, 1\}^n$ as a vector space over $GF(2)$. Put*

$$
p \stackrel{\text{def}}{=} \min\left\{\mu(0^n), \mu(v^{(1)}), \mu(v^{(2)}), \ldots, \mu(v^{(n)})\right\}.
$$

Let $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(T)}$ be independent random vectors, each distributed according to μ. Then for every $v \in \{0, 1\}^n$,

$$
\left|\mathbf{Pr}\left[\mathbf{u}^{(1)} \oplus \cdots \oplus \mathbf{u}^{(T)} = v\right] - \frac{1}{2^n}\right| \leqslant \frac{1}{e^{2Tp}}.
$$

We are ready to formally define dense predicates and give the promised reduction.

Definition 3.4 (Dense predicate). Let n, b be positive integers and $d \geqslant 0$ a real number. A predicate D is called (n, b, d)-*dense* if D is a predicate $\{0, 1, \ldots, n\} \to \{0, 1\}$ with characteristic vector $(v_0, v_1, \ldots, v_n)$ satisfying

$$
v_{rb+1} + v_{rb+2} + \cdots + v_{(r+1)b} \geqslant d \qquad \text{for all } r = 0, 1, 2, \ldots, \left\lfloor\frac{n}{b}\right\rfloor - 1.
$$

---

[1]Razborov's proof is in Russian. For an English translation, see Jukna [10, Lem. 24.3].

Lemma 3.5 (Reduction from high-degree to dense predicates). *Let* D : $\{0, 1, \ldots, n\} \to \{0, 1\}$ *be a predicate with* $\deg(D) \geqslant \frac{1}{4}n$. *Let* b *be any integer with* $1 \leqslant b \leqslant \frac{1}{350}n$. *Then*

$$U(D) \geqslant \frac{b}{n \log n} \, U(D'),$$

where D' *is a certain* $(m, \lceil \log n \rceil b, \frac{1}{700}b)$-*dense predicate and* $\frac{1}{350}n \leqslant m \leqslant n$.

Proof. Let $(v_0, v_1, \ldots, v_n)$ be the characteristic vector of D. Apply Lemma 3.1 to $(v_1, \ldots, v_n)$ and let i, ℓ be the resulting indices $(i \leqslant \ell)$. It will be convenient to work with a somewhat smaller subvector $v \stackrel{\text{def}}{=} (v_i, \ldots, v_j)$, where we define $j \in \{i, \ldots, \ell\}$ to be the largest integer so that $b \mid (j - i + 1)$. Since $b \leqslant \frac{1}{350}n$ and $v_i + \cdots + v_\ell \geqslant \frac{1}{168}n$, this gives:

$$v_i + \cdots + v_j \geqslant \frac{1}{350}n. \tag{3.8}$$

Defining $m \stackrel{\text{def}}{=} j - i + 1$, we infer that $\frac{1}{350}n \leqslant m \leqslant n$, as desired. We view $v = (v_i, \ldots, v_j)$ as composed of consecutive blocks, each b bits long:

$$v = \left(\underbrace{\boxed{v_i, \ldots, v_{i+b-1}}}_{\text{block 1}}, \underbrace{\boxed{v_{i+b}, \ldots, v_{i+2b-1}}}_{\text{block 2}}, \cdots \cdots , \underbrace{\boxed{v_{j-b+1}, \ldots, v_j}}_{\text{block } m/b} \right). \tag{3.9}$$

For $r = 1, 2, \ldots, b$, define the *rth layer of* v, denoted $z^{(r)}$, to be the vector obtained by taking the rth component from each of the above blocks:

$$z^{(r)} \stackrel{\text{def}}{=} (v_{i-1+r}, v_{i-1+b+r}, \ldots, v_{j-b+r}) \in \{0, 1\}^{m/b}.$$
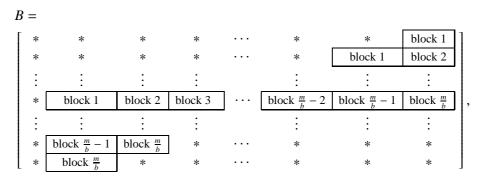
We say of a layer z that it is *perfect* if it does not have $\lceil \log n \rceil$ consecutive components equal to 0. If more than $\frac{1}{700}b$ of the layers are perfect, take D' to be the predicate with characteristic vector $(v_0 \oplus \cdots \oplus v_{i-1}, v_i, \ldots, v_j)$. Clearly, D' is $(m, \lceil \log n \rceil b, \frac{1}{700}b)$-dense. Furthermore, $U(D') \leqslant U(D)$, by the same argument as in Lemma 3.2. As a result, the theorem holds in this case.

Thus, we may assume that at least $(1 - \frac{1}{700})b$ of the layers are not perfect. In view of (3.8), at most $(1 - \frac{1}{350})b$ layers can be zero vectors. Therefore, $\frac{1}{700}b$ or more layers are nonzero *and* not perfect. These are the only layers we will consider in the remainder of the proof.

Define predicates $D^{-(m-b)}, D^{-(m-2b)}, \ldots, D^{-b}, D^0, D^b, \ldots, D^{m-2b}, D^{m-b}$, each a mapping $\{0, 1, \ldots, m\} \to \{0, 1\}$, by $D^r(t) \equiv D(t + i - 1 + r)$. These are a subset of the predicates from the proof of Lemma 3.2, and again

$$U(D) \geqslant U(D^r) \qquad \text{for each } r. \tag{3.10}$$

Writing the characteristic vectors of these predicates one after another as row vectors yields the following matrix B:

$$B =$$

$$\begin{bmatrix} * & * & * & * & \cdots & * & * & \boxed{\text{block 1}} \\ * & * & * & * & \cdots & * & \boxed{\text{block 1}} & \text{block 2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ * & \boxed{\text{block 1}} & \text{block 2} & \text{block 3} & \cdots & \boxed{\text{block } \frac{m}{b} - 2} & \text{block } \frac{m}{b} - 1 & \text{block } \frac{m}{b} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ * & \boxed{\text{block } \frac{m}{b} - 1} & \text{block } \frac{m}{b} & * & \cdots & * & * & * \\ * & \boxed{\text{block } \frac{m}{b}} & * & * & \cdots & * & * & * \end{bmatrix},$$

where the blocks refer to the partition in (3.9). Let T be a suitably large integer to be named later, and let $\mathbf{u}^{(1)}, \mathbf{u}^{(2)}, \dots, \mathbf{u}^{(T)}$ be independent random vectors, each selected uniformly from among the rows of B. Put

$$\mathbf{u} \stackrel{\text{def}}{=} \mathbf{u}^{(1)} \oplus \mathbf{u}^{(2)} \oplus \cdots \oplus \mathbf{u}^{(T)}.$$

We will index the columns of B and the components of $\mathbf{u}$ by $0, 1, \dots, m$ (left to right). Key to analyzing the distribution of $\mathbf{u}$ is the following claim.

Claim 3.5.1. *Let $T \geqslant (m/b) \ln n$. Let $\Delta \in \{1, 2, \dots, b\}$ be such that the layer $z^{(\Delta)}$ is nonzero and not perfect. Let $s \in \{0, b, 2b, 3b, \dots\}$ be such that $s + \lceil \log n \rceil b \leqslant m$. Then*

$$\mathbf{Pr}\left[(\mathbf{u})_{s+\Delta} = (\mathbf{u})_{s+b+\Delta} = \cdots = (\mathbf{u})_{s+(\lceil \log n \rceil - 1)b + \Delta} = 0 \right] \leqslant \frac{2}{n}.$$

Proof. Let B' be the matrix whose columns are the following columns of B: $s + \Delta, s + b + \Delta, \dots, s + (\lceil \log n \rceil - 1)b + \Delta$, in that order. Since $z^{(\Delta)}$ is nonzero and not perfect, it has $\lceil \log n \rceil + 1$ consecutive components with values either $0, 0, \dots, 0, 1$ or $1, 0, 0, \dots, 0$. Consequently, B' must contain one of the following submatrices, each of size $(\lceil \log n \rceil + 1) \times \lceil \log n \rceil$:

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \\ 0 & 0 & 0 & \cdots & 0 & 1 & * \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 1 & * & \dots & * & * & * \\ 1 & * & * & \dots & * & * & * \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} * & * & * & \cdots & * & * & 1 \\ * & * & * & \cdots & * & 1 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ * & 1 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}.$$

The claim now follows immediately from Lemma 3.3, since $2^{-\lceil \log n \rceil} + e^{-2T \cdot \frac{b}{2m}} \leqslant 2/n$. $\qquad\qquad\qquad\square$

We now return to the proof of the lemma. Fix $T = \lceil (m/b) \ln n \rceil$. Let $s = 0$ and apply Claim 3.5.1 with every $\Delta \in \{1, 2, \ldots, b\}$ for which the layer $z^{(\Delta)}$ is nonzero and not perfect. Since there are at least $\frac{1}{700} b$ such choices for Δ, we conclude by the union bound that

$$\mathbf{Pr}\left[(\mathbf{u})_1 + (\mathbf{u})_2 + \cdots + (\mathbf{u})_{\lceil \log n \rceil b} < \frac{1}{700} b\right] \leqslant b \cdot \frac{2}{n}.$$

The same calculation applies to the next set of $\lceil \log n \rceil b$ components of $\mathbf{u}$ (i.e., $s = \lceil \log n \rceil b$), and so on. Applying a union bound across all these $m/(\lceil \log n \rceil b)$ calculations, we find that with probability

$$1 - \frac{m}{\lceil \log n \rceil b}\left(b \cdot \frac{2}{n}\right) > 0,$$

the predicate whose characteristic vector is $\mathbf{u}$ is $(m, \lceil \log n \rceil b, \frac{1}{700} b)$-dense. Fix any such predicate D'. Since D' is the XOR of $T \leqslant (n \log n)/b$ predicates from among $D^{-(m-b)}, \ldots, D^{m-b}$, the lemma follows by (3.10) and Proposition 2.6. $\qquad\square$

4 A Lower Bound for Approximation by Polynomials

Crucial to our study of dense predicates are certain approximation problems to which they give rise. Roughly speaking, the hardness of such an approximation problem for low-degree polynomials translates into the communication hardness of the associated predicate. This section carries out the first part of the program, namely, showing that the approximation task at hand is hard for low-degree polynomials. We examine this question in its basic mathematical form, with no extraneous considerations to obscure our view. How communication fits in this picture will become clear in the next two sections.

For a finite set $X \subset \mathbb{R}$, a function $f : X \to \mathbb{R}$, and an integer $r \geqslant 0$, define

$$\epsilon^*(f, X, r) \stackrel{\text{def}}{=} \min_{p \in P_r} \max_{x \in X} |p(x) - f(x)|.$$

In words, $\epsilon^*(f, X, r)$ is the least error (in the uniform sense) to which a degree-r polynomial can approximate f on X. The following well-known fact from approximation theory is useful in estimating this error.

Fact 4.1 (see, e.g., [30, Thm. 1.15]). *Let $X = \{x_1, x_2, \ldots, x_{r+2}\}$ be a set of $r + 2$ distinct reals. Let $f : X \to \mathbb{R}$ be given. Put $\omega(x) \stackrel{\text{def}}{=} (x - x_1)(x - x_2) \cdots (x - x_{r+2})$. Then*

$$\epsilon^*(f, X, r) = \frac{\left| \sum_{i=1}^{r+2} [f(x_i)/\omega'(x_i)] \right|}{\sum_{i=1}^{r+2} [1/|\omega'(x_i)|]}.$$

To develop some intuition for the work in this section, consider the following approximation problem. Let $f : \{0, 1, \ldots, n\} \to \{0, 1\}$ be defined by

$$f(x) = \begin{cases} 1 & \text{if } x = \lfloor n/2 \rfloor, \\ 0 & \text{otherwise.} \end{cases}$$

It is well-known that any polynomial that approximates f within $1/3$ has degree $\Omega(n)$. For example, this follows from work by Paturi [26]. The approximation problem of interest to us is similar, except that our points need not be as evenly spaced as $0, 1, \ldots, n$ but rather may form clusters. As a result, Paturi's results and methods do not apply, and we approach this question differently, using the first-principles formula of Fact 4.1. Specifically, our main result in this section is as follows.

Lemma 4.2 (Inapproximability by low-degree polynomials). *Let positive integers L, d and a real number $B \geqslant d$ be given. Let $\{x_{ij} : i = 1, \ldots, L; j = 1, \ldots, d\}$ be a set of Ld distinct reals, where $x_{ij} \in [(i-1)B, iB]$ and*

$$|x_{ij} - x_{i'j'}| \geqslant 1 \qquad \text{for } (i, j) \neq (i', j'). \tag{4.1}$$

Let $x_0 \in [\frac{1}{4}LB, \frac{3}{4}LB]$. Then any polynomial p with

$$p(x_0) = 1, \qquad |p(x_{ij})| < \frac{1}{2} \left(\frac{1}{LB} \right)^{4d+1} \qquad \text{for all } i, j$$

has degree at least $\left(\frac{1}{2}L - 1 \right) d$.

Proof. Define $f(x)$ by

$$f(x) = \begin{cases} 1 & \text{if } x = x_0, \\ 0 & \text{if } x = x_{ij} \text{ for some } i, j. \end{cases}$$

By symmetry, we can assume that $x_0 \in [\frac{1}{4}LB, \frac{1}{2}LB]$. Fix an integer $\ell \leqslant \lceil \frac{1}{2}L \rceil$ so that $x_0 \in [(\ell - 1)B, \ell B]$. Put

$$X \stackrel{\text{def}}{=} \{x_0\} \cup \{x_{ij} : i = 1, \ldots, 2\ell - 1; \quad j = 1, \ldots, d\}.$$

21

With $\omega(x) \overset{\text{def}}{=} \prod_{y \in X}(x - y)$, Fact 4.1 implies that

$$\epsilon^*(f, X, |X| - 2) \geqslant \frac{1}{|X|} \frac{\min_{x \in X} |\omega'(x)|}{|\omega'(x_0)|}. \qquad (4.2)$$

We proceed to estimate the denominator and numerator of (4.2). If $x_0 = x_{ij}$ for some i, j, the lemma is vacuous. Thus, we can assume that the quantity

$$\delta \overset{\text{def}}{=} \min_{\substack{i=1,\ldots,2\ell-1, \\ j=1,\ldots,d}} |x_0 - x_{ij}|$$

satisfies $\delta > 0$. We have:

$$|\omega'(x_0)| = \prod_{j=1}^{d} \prod_{i=1}^{2\ell-1} |x_0 - x_{ij}| \leqslant \delta \prod_{j=1}^{d} \prod_{i=1}^{2\ell-1} B \underbrace{\left\lceil \frac{|x_0 - x_{ij}|}{B} \right\rceil}_{\leqslant |i-\ell|+1}$$

$$\leqslant \delta \cdot \left(\ell! \, \ell! \, B^{2\ell-1}\right)^d. \qquad (4.3)$$

On the other hand, every $x_{i'j'} \in X$ satisfies:

$$|\omega'(x_{i'j'})| = \prod_{x \in X \setminus \{x_{i'j'}\}} |x - x_{i'j'}| \overset{(4.1)}{\geqslant} \delta \prod_{j=1}^{d} \prod_{\substack{i=1,\ldots,2\ell-1 \\ i \notin \{i'-1,i',i'+1\}}} |x_{ij} - x_{i'j'}|$$

$$\geqslant \delta \prod_{j=1}^{d} \prod_{\substack{i=1,\ldots,2\ell-1 \\ i \notin \{i'-1,i',i'+1\}}} B \underbrace{\left\lfloor \frac{|x_{ij} - x_{i'j'}|}{B} \right\rfloor}_{\geqslant |i-i'|-1}$$

$$\geqslant \delta \cdot \left(\frac{\ell! \, \ell! \, B^{2\ell-4}}{\ell^4}\right)^d. \qquad (4.4)$$

Now (4.2) yields, in view of (4.3) and (4.4):

$$\epsilon^*(f, X, |X| - 2) \geqslant \frac{1}{2}\left(\frac{1}{LB}\right)^{4d+1},$$

which concludes the proof since $|X| \geqslant \left(\frac{1}{2}L - 1\right)d + 1$. $\qquad \square$

5 Smooth Orthogonalizing Distributions

We now transition to the final ingredient of our proof, *smooth orthogonalizing distributions* for a given predicate D. This informal term refers to a distribution

on $\{0, 1\}^n$ that does not put too little weight on any point (the *smooth* part) and under which $(-1)^{D(x_1+\cdots+x_n)}$ is approximately orthogonal to all low-degree parity functions χ_S (the *orthogonalizing* part). Our task is to establish the existence of such distributions for every dense predicate. Crucial to this undertaking will be the inapproximability result that we proved in Section 4.

For a polynomial p, a predicate $D : \{0, 1, \ldots, n\} \to \{0, 1\}$, and a number $N > 0$, define the *advantage* of p in computing D by

$$\mathrm{adv}(p, N, D) \quad \overset{\text{def}}{=} \quad N \min_{t=0,\ldots,n} \left\{ (-1)^{D(t)} p(t) \right\} \; + \; \sum_{t=0}^{n} \frac{\binom{n}{t}}{2^n} (-1)^{D(t)} p(t).$$

This quantity is conceptually close to the correlation of p and D with respect the binomial distribution. There is a substantial difference, however: if p and D differ in sign at some point, this causes a penalty term to be subtracted. We will be interested in values $N \gg 1$, when even a single error of p results in a large penalty. Define

$$\mathrm{adv}_r(N, D) \quad \overset{\text{def}}{=} \quad \max_p \{ \mathrm{adv}(p, N, D) \},$$

where the maximization is over $p \in P_r$ with $|p(t)| \leqslant 1$ for $t = 0, 1, \ldots, n$. As we now show, this quantity is closely related to smooth orthogonalizing distributions for D.

Theorem 5.1 (Smooth distributions vs. approximation by polynomials). *Given a predicate $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ and an integer $r \geqslant 0$. Then for every $N > 1$, there is a distribution μ on $\{0, 1\}^n$ such that $\mu(x) \geqslant \frac{1}{2^n N}$ for each x and*

$$\left| \mathbf{E}_x \left[(-1)^{D(x_1+\cdots+x_n)} \mu(x) \, \chi_S(x) \right] \right| \leqslant \frac{1}{2^n N} \, \mathrm{adv}_r(N-1, D) \qquad \textit{for } |S| \leqslant r.$$

Proof. Put $f(x) \overset{\text{def}}{=} (-1)^{D(x_1+\cdots+x_n)}$ and consider the following linear program:

$$
\boxed{
\begin{array}{lll}
\text{variables:} & \mu(x) \text{ for all } x; \quad \epsilon & \\[4pt]
\text{minimize:} & \epsilon & \\[4pt]
\text{subject to:} & \left| \displaystyle\sum_{x \in \{0,1\}^n} \mu(x) f(x) \chi_S(x) \right| \leqslant \epsilon & \text{for } |S| \leqslant r, \\[16pt]
& \displaystyle\sum_{x \in \{0,1\}^n} \mu(x) = 1, & \\[16pt]
& \mu(x) \geqslant \dfrac{1}{2^n N} & \text{for each } x.
\end{array}
}
\tag{LP1}
$$

It suffices to show that the optimum of this program is at most $\frac{1}{N}\operatorname{adv}_r(N-1,D)$. For this, we pass to the dual:

$$
\begin{aligned}
&\text{variables:} && \alpha_S \ (\text{for } |S| \leqslant r); \quad \xi_x \ (\text{for all } x); \quad \Delta \\[4pt]
&\text{maximize:} && \frac{1}{N}\left((N-1)\Delta + \frac{1}{2^n}\sum_{x \in \{0,1\}^n}(\Delta + \xi_x)\right) \\[4pt]
&\text{subject to:} && f(x)\sum_{|S|\leqslant r}\alpha_S\,\chi_S(x) \geqslant \Delta + \xi_x && \text{for all } x, \\[4pt]
& && \sum_{|S|\leqslant r}|\alpha_S| \leqslant 1, \\[4pt]
& && \alpha_S \in \mathbb{R} && \text{for } |S| \leqslant r, \\[4pt]
& && \xi_x \geqslant 0 && \text{for all } x, \\[4pt]
& && \Delta \in \mathbb{R}.
\end{aligned}
\tag{LP2}
$$

The dual programs (LP1) and (LP2) are both feasible and thus have the same finite optimum. Therefore, our task reduces to proving that the optimum of (LP2) is at most $\frac{1}{N}\operatorname{adv}_r(N-1,D)$. Fix an optimal solution to (LP2). Then

$$
f(x)\sum_{|S|\leqslant r}\alpha_S\,\chi_S(x) = \Delta + \xi_x \qquad \text{for all } x, \tag{5.1}
$$

since in case of a strict inequality ($>$) we could increase the corresponding variable ξ_x by a small amount to obtain a feasible solution with greater value. Furthermore, we claim that

$$
\Delta = \min_{x \in \{0,1\}^n}\left\{f(x)\sum_{|S|\leqslant r}\alpha_S\,\chi_S(x)\right\}. \tag{5.2}
$$

Indeed, let m stand for the right-hand side of (5.2). Then $\Delta \leqslant m$ because each ξ_x is nonnegative. It remains to show that $\Delta \geqslant m$. If we had $\Delta < m$, then (5.1) would imply that $\xi_x \geqslant m - \Delta$ for all x. As a result, we could obtain a new feasible solution $\xi'_x = \xi_x + (\Delta - m)$ and $\Delta' = m$. This new solution satisfies $\Delta' + \xi'_x = \Delta + \xi_x$ for all x. Moreover, $\Delta' > \Delta$, which results in a greater objective value and yields the desired contradiction. In summary, $\Delta = m$.

In view of (5.1) and (5.2), the optimum of (LP2) is

$$
\frac{1}{N}\max_{\phi}\left\{(N-1)\min_{x}\{f(x)\phi(x)\} \;+\; \frac{1}{2^n}\sum_{x}f(x)\phi(x)\right\}, \tag{5.3}
$$

where the maximization is over functions ϕ of the form

$$\phi(x) = \sum_{|S|\leqslant r} \alpha_S \chi_S(x), \qquad \text{where } \sum_{|S|\leqslant r} |\alpha_S| \leqslant 1. \qquad (5.4)$$

Fix ϕ that optimizes (5.3). By (5.4),

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

Put

$$\phi_{\text{sym}}(x) \stackrel{\text{def}}{=} \frac{1}{n!} \sum_{\sigma\in S_n} \phi(x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$

Since f is symmetric, ϕ and ϕ_{sym} have the same objective value in (5.3). By the symmetrization argument (Proposition 2.3), there is a univariate polynomial $p \in P_r$ with

$$\phi_{\text{sym}}(x) = p(x_1 + \cdots + x_n) \qquad \text{for all } x \in \{0,1\}^n.$$

For $t = 0, 1, \ldots, n$,

$$|p(t)| = |p(\underbrace{1 + \cdots + 1}_{t \text{ times}} + 0 + \cdots + 0)| \leqslant \max_{x\in\{0,1\}^n} \{|\phi_{\text{sym}}(x)|\} \leqslant \max_{x\in\{0,1\}^n} \{|\phi(x)|\} \leqslant 1.$$

Replacing $\phi(x)$ by $p(x_1 + \cdots + x_n)$ in (5.3), we see that the optimum of (LP2) is at most

$$\frac{1}{N} \max_p \left\{ (N-1) \min_{t=0,\ldots,n} \left\{ (-1)^{D(t)} p(t) \right\} + \frac{1}{2^n} \sum_{t=0}^{n} \binom{n}{t} (-1)^{D(t)} p(t) \right\},$$

where the maximization is over $p \in P_r$ with $|p(t)| \leqslant 1$ for $t = 0, 1, \ldots, n$. This latter quantity is $\frac{1}{N} \text{adv}_r(N-1, D)$, by definition. $\qquad\square$

Theorem 5.1 states that a smooth orthogonalizing distribution for D exists whenever low-degree polynomials have negligible advantage in computing D. Accordingly, we proceed to examine the advantage achievable by low-degree polynomials.

Lemma 5.2 (Each dense predicate induces a hard approximation problem).
Let D be an $(n, B, 2d+1)$-dense predicate, where n, B, d are positive integers. Assume that $\text{adv}_r(D, N) \geqslant n2^{-n/6}$, where $r < \deg(D)$ and $N > 0$ are given. Then there are $\lfloor \frac{n}{B} \rfloor d$ distinct reals $\{x_{ij} : i = 1, \ldots, \lfloor \frac{n}{B} \rfloor; j = 1, \ldots, d\}$ and a polynomial

$p \in P_r$ *such that:*

$$x_{ij} \in [(i-1)B, iB] \qquad \text{for all } i, j,$$
$$|x_{ij} - x_{i'j'}| \geqslant 1 \qquad \text{for all } (i, j) \neq (i', j'),$$
$$|p(x_{ij})| \leqslant \sqrt{n}/N \qquad \text{for all } i, j,$$
$$p(x_0) = 1 \qquad \text{for some } x_0 \in [\tfrac{1}{4}n, \tfrac{3}{4}n].$$

Proof. Fix $q \in P_r$ with $|q(t)| \leqslant 1$ for $t = 0, 1, \dots, n$ and $\mathrm{adv}(q, D, N) = \mathrm{adv}_r(D, N)$. Fix $k \in \{0, 1, \dots, n\}$ with

$$\binom{n}{k}(-1)^{D(k)}q(k) = \max_{t=0,\dots,n}\left\{\binom{n}{t}(-1)^{D(t)}q(t)\right\}.$$

Since $\deg(q) < \deg(D)$, the quantity $\binom{n}{t}(-1)^{D(t)}q(t)$ is positive for at most n values of $t = 0, 1, \dots, n$. Therefore,

$$\mathrm{adv}(q, D, N) \leqslant n \cdot \frac{\binom{n}{k}}{2^n}(-1)^{D(k)}q(k) \leqslant n \cdot \frac{\binom{n}{k}}{2^n}.$$

Recalling that $\mathrm{adv}(q, D, N) \geqslant n2^{-n/6}$, we infer that $\tfrac{1}{4}n \leqslant k \leqslant \tfrac{3}{4}n$. Put

$$p(t) \stackrel{\mathrm{def}}{=} \frac{1}{|q(k)|}\, q(t).$$

Taking $x_0 \stackrel{\mathrm{def}}{=} k$, we have $\tfrac{1}{4}n \leqslant x_0 \leqslant \tfrac{3}{4}n$ and $p(x_0) = 1$, as desired. It remains to find the points x_{ij}. For this, we need the following claim.

Claim 5.2.1. *Let a, b be integers with $a < b$ and $D(a) \neq D(b)$. Then $|p(\xi)| \leqslant \sqrt{n}/N$ for some $\xi \in [a, b]$.*

Proof. If q vanishes at some point in $[a, b]$, we are done. In the contrary case, q is nonzero and has the same sign at every point of $[a, b]$, which means that either $q(a)(-1)^{D(a)} < 0$ or $q(b)(-1)^{D(b)} < 0$. Since $\mathrm{adv}(q, D, N) \geqslant 0$, we have:

$$\min\{|q(a)|, |q(b)|\} \leqslant \frac{n}{N}\max_{t=0,\dots,n}\left\{\frac{\binom{n}{t}}{2^n}(-1)^{D(t)}q(t)\right\} = \frac{n}{N} \cdot \frac{\binom{n}{k}}{2^n} \cdot |q(k)|$$

$$\leqslant \frac{\sqrt{n}}{N}\,|q(k)|,$$

and hence $\min\{|p(a)|, |p(b)|\} \leqslant \sqrt{n}/N$. $\qquad\qquad\square$

Consider any segment $[(i-1)B + 1, iB]$, for an integer i with $1 \leqslant i \leqslant \lfloor \frac{n}{B} \rfloor$. Since D is $(n, B, 2d+1)$-dense, it changes value at least $2d$ times in $[(i-1)B + 1, iB]$. As a result, there are at least d pairs of integers $(a_1, b_1), \ldots, (a_d, b_d)$ with

$$D(a_1) \neq D(b_1), \quad D(a_2) \neq D(b_2), \quad \ldots, \quad D(a_d) \neq D(b_d)$$

and

$$(i-1)B + 1 \leqslant a_1 < b_1 < a_2 < b_2 < \cdots < a_d < b_d \leqslant iB.$$

In view of Claim 5.2.1, this provides the desired d points in $[(i-1)B + 1, iB]$. $\square$

Our work in this and the previous section furnishes all the key ingredients needed to deduce the existence of smooth orthogonalizing distributions for dense predicates. Putting them together yields the main result of this section:

Theorem 5.3 (Smooth orthogonalizing distributions for dense predicates). *Let D be an $(n, B, 2d+1)$-dense predicate, where n, B, d are positive integers with $B \mid n$ and $n \geqslant 3B$. Then there is a distribution μ on $\{0, 1\}^n$ such that:*

$$\mu(x) \geqslant \frac{1}{2^n} \frac{1}{3n^{4d+1.5}} \qquad \qquad \textit{for each } x,$$

$$\left| \mathop{\mathbf{E}}_{x} \left[(-1)^{D(x_1 + \cdots + x_n)} \mu(x) \chi_S(x) \right] \right| \leqslant 2^{-7n/6} \qquad \textit{for } |S| < \frac{nd}{6B}.$$

Proof. Put $N \stackrel{\text{def}}{=} 3n^{4d+1.5}$. In view of Theorem 5.1, it suffices to show that $\mathrm{adv}_r(D, N-1) < n2^{-n/6}$ for all $r < \frac{nd}{6B}$. So assume, for the sake of contradiction, that $\mathrm{adv}_r(D, N-1) \geqslant n2^{-n/6}$ for some $r < \frac{nd}{6B}$. Since $\deg(D) \geqslant \frac{n}{B}(2d+1)$, we have $r < \deg(D)$. Thus, Lemma 5.2 is applicable and yields $\frac{nd}{B}$ distinct reals $\{x_{ij} : i = 1, \ldots, \frac{n}{B}; j = 1, \ldots, d\}$ and a polynomial $p \in P_r$ such that:

$$
\begin{array}{ll}
x_{ij} \in [(i-1)B, iB] & \text{for all } i, j, \\
|x_{ij} - x_{i'j'}| \geqslant 1 & \text{for all } (i, j) \neq (i', j'), \\
|p(x_{ij})| < \frac{1}{2} \left(\frac{1}{n} \right)^{4d+1} & \text{for all } i, j, \\
p(x_0) = 1 & \text{for some } x_0 \in [\frac{1}{4}n, \frac{3}{4}n].
\end{array}
$$

Applying Lemma 4.2 with $L \stackrel{\text{def}}{=} \frac{n}{B}$, we infer that $r \geqslant \left(\frac{1}{2} \frac{n}{B} - 1 \right) d$, which yields $r \geqslant \frac{nd}{6B}$ since $\frac{n}{B} \geqslant 3$. We have reached the desired contradiction to $r < \frac{nd}{6B}$. $\square$

6 Proof of the Main Result

This section consolidates the preceding developments into our main result, an optimal lower bound on the unbounded-error communication complexity of every symmetric function. As outlined earlier, we will first solve this problem for dense predicates and then extend our work to the general case via the reductions of Section 3.

As a first step, we identify a pattern matrix inside the communication matrix of a given predicate D.

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

$$\left[(-1)^{D(|x \wedge y|)} \right]_{x \in \{0,1\}^m, y \in \{0,1\}^m}.$$

The author has proved an almost identical statement in earlier work [32, Lem. 6.1]. For the reader's convenience, we reproduce that proof with the needed adaptations in Appendix A.

We are now ready to solve the problem for all dense predicates.

Theorem 6.2 (Communication complexity of dense predicates). *Let $\alpha > 0$ be a sufficiently small absolute constant. Let D be an $(m, b\lceil \log n \rceil, \frac{1}{700}b)$-dense predicate, where $\frac{1}{350} n \leqslant m \leqslant n$ and $b = \lfloor \alpha n / \log^2 n \rfloor$. Then*

$$U(D) \geqslant \Omega\left(\frac{n}{\log n} \right).$$

Proof. Throughout the proof we will, without mention, use the assumption that n is large enough. This will simplify the setting of parameters, the manipulation of floors and ceilings, and generally make the proof easier to follow.

Fix an integer $v \in [\frac{1}{8}m, \frac{1}{4}m]$ with $b\lceil \log n \rceil \mid v$. Clearly, $v \geqslant 3b\lceil \log n \rceil$. Define $D' : \{0, 1, \ldots, v\} \rightarrow \{0, 1\}$ by $D'(t) \equiv D(t)$. Since D' is $(v, b\lceil \log n \rceil, \frac{1}{700}b)$-dense, Theorem 5.3 provides a distribution μ on $\{0, 1\}^v$ with

$$\mu(z) \geqslant 2^{-v} \, 2^{-\alpha n / 350 \log n} \qquad \text{for each } z \in \{0, 1\}^v, \tag{6.1}$$

$$\left| \mathop{\mathbf{E}}_z \left[(-1)^{D(|z|)} \mu(z) \chi_S(z) \right] \right| \leqslant 2^{-7v/6} \qquad \text{for } |S| < \frac{v}{6 \cdot 1401 \lceil \log n \rceil}. \tag{6.2}$$

Define $\phi : \{0, 1\}^v \to \mathbb{R}$ by $\phi(z) \stackrel{\text{def}}{=} (-1)^{D(|z|)}\mu(z)$. Restating (6.2),

$$|\hat{\phi}(S)| \leqslant 2^{-7v/6} \qquad \text{for } |S| < \frac{v}{6 \cdot 1401\lceil \log n \rceil}. \qquad (6.3)$$

Furthermore, Proposition 2.2 reveals that

$$\max_{S \subseteq [v]} |\hat{\phi}(S)| \leqslant 2^{-v}. \qquad (6.4)$$

Let A be the $(2v, v, 8^{-v}\phi)$-pattern matrix. By (6.3), (6.4), and Theorem 2.8,

$$\|A\| \leqslant 4^{-v} \, 2^{-v/12 \cdot 1401\lceil \log n \rceil}. \qquad (6.5)$$

By (6.1), every entry of A has absolute value at least $16^{-v} \, 2^{-\alpha n/350 \log n}$. Combining this observation with (6.5) and Theorem 2.5,

$$\text{dc}(A) \geqslant 2^{v/12 \cdot 1401\lceil \log n \rceil} \, 2^{-\alpha n/350 \log n}.$$

Recall that $v \geqslant \frac{1}{8} m \geqslant \frac{1}{8 \cdot 350} n$. Hence, for a suitably small constant $\alpha > 0$,

$$\text{dc}(A) \geqslant 2^{\Omega(n/\log n)}.$$

It remains to relate the sign-rank of A to the communication complexity of D. Let F be the $(2v, v, f)$-pattern matrix, where $f(z) \stackrel{\text{def}}{=} (-1)^{D(|z|)}$. Then $\text{dc}(A) = \text{dc}(F)$ because A and F have the same sign pattern. However, Lemma 6.1 states that F is a submatrix of the communication matrix of D, namely,

$$M \stackrel{\text{def}}{=} \left[(-1)^{D(|x \wedge y|)}\right]_{x \in \{0,1\}^m, y \in \{0,1\}^m}.$$

Thus, $\text{dc}(M) \geqslant \text{dc}(F)$. Summarizing,

$$\text{dc}(M) \geqslant \text{dc}(F) = \text{dc}(A) \geqslant 2^{\Omega(n/\log n)}.$$

In view of Theorem 2.4, the proof is complete. $\qquad\qquad\qquad\qquad\qquad \square$

The hard work is now behind us. What remains is to apply the reductions of Section 3, in reverse order.

Corollary 6.2.1 (Communication complexity of high-degree predicates). *Let* $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ *be a predicate with* $\deg(D) \geqslant \frac{1}{4}n$. *Then*

$$U(D) \geqslant \Omega\left(\frac{n}{\log^4 n}\right).$$

Proof. Immediate from Lemma 3.5 and Theorem 6.2. □

Corollary 6.2.2 (Communication complexity of arbitrary predicates). *Let D :* $\{0, 1, \ldots, n\} \to \{0, 1\}$ *be a nonconstant predicate. Put $k \stackrel{\text{def}}{=} \deg(D)$. Then*

$$U(D) \geqslant \Omega\left(\frac{k}{[1 + \log(n/k)] \log^4 n}\right).$$

Proof. Immediate from Lemma 3.2 and Corollary 6.2.1. □

At last, we arrive at the main result of this paper.

Theorem 1.1 (Restated from p. 4). *Let $D : \{0, 1, \ldots, n\} \to \{0, 1\}$ be given. Then*

$$U(D) = \tilde{\Theta}(\deg(D)),$$

where the $\tilde{\Theta}$ notation suppresses $\log n$ factors.

Proof. The lower bound on $U(D)$ follows by Corollary 6.2.2. To prove the upper bound, let p be a polynomial of degree $\deg(D)$ with

$$\text{sign}(p(t)) = (-1)^{D(t)} \qquad \text{for } t = 0, 1, \ldots, n.$$

Put

$$M \stackrel{\text{def}}{=} \left[(-1)^{D(|x \wedge y|)}\right]_{x,y}, \qquad R \stackrel{\text{def}}{=} \left[p(x_1 y_1 + \cdots + x_n y_n)\right]_{x,y},$$

where the indices run as usual: $x, y \in \{0, 1\}^n$. Then $M_{xy} R_{xy} > 0$ for all x and y. Therefore,

$$\text{dc}(M) \leqslant \text{rank}(R) \leqslant \sum_{i=0}^{\deg(D)} \binom{n}{i} \leqslant 2^{O(\deg(D) \log n)}.$$

In view of Theorem 2.4, this completes the proof. □

References

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

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

[3] R. de Wolf. *Quantum Computing and Communication Complexity*. PhD thesis, University of Amsterdam, 2001.

[4] V. Feldman, P. Gopalan, S. Khot, and A. K. Ponnuswami. New results for learning noisy parities and halfspaces. In *Proceedings of the 47th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 563–574, 2006.

[5] J. Forster. A linear lower bound on the unbounded error probabilistic communication complexity. *J. Comput. Syst. Sci.*, 65(4):612–625, 2002.

[6] J. Forster, M. Krause, S. V. Lokam, R. Mubarakzjanov, N. Schmitt, and H.-U. Simon. Relations between communication complexity, linear arrangements, and computational complexity. In *Proc. of the 21st Conf. on Foundations of Software Technology and Theoretical Computer Science (FST TCS)*, pages 171–182, 2001.

[7] J. Forster and H. U. Simon. On the smallest possible dimension and the largest possible margin of linear arrangements representing given concept classes. *Theor. Comput. Sci.*, 350(1):40–48, 2006.

[8] M. Goldmann, J. Håstad, and A. A. Razborov. Majority gates vs. general weighted threshold gates. *Computational Complexity*, 2:277–300, 1992.

[9] A. Hajnal, W. Maass, P. Pudlák, G. Turán, and M. Szegedy. Threshold circuits of bounded depth. *J. Comput. Syst. Sci.*, 46(2):129–154, 1993.

[10] S. Jukna. *Extremal Combinatorics with Applications in Computer Science*. Springer-Verlag, Berlin, 2001.

[11] B. Kalyanasundaram and G. Schintger. The probabilistic communication complexity of set intersection. *SIAM J. Discret. Math.*, 5(4):545–557, 1992.

[12] B. Kashin and A. A. Razborov. Improved lower bounds on the rigidity of Hadamard matrices. *Matematicheskie zametki*, 63(4):535–540, 1998. In Russian.

[13] M. Kearns and L. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. *J. ACM*, 41(1):67–95, 1994.

[14] M. Kharitonov. Cryptographic hardness of distribution-specific learning. In *Proc. of the 25th Symposium on Theory of Computing*, pages 372–381, 1993.

[15] A. R. Klivans and R. Servedio. Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. In *Proc. of the 33rd Symposium on Theory of Computing (STOC)*, pages 258–265, 2001.

[16] A. R. Klivans and A. A. Sherstov. Cryptographic hardness for learning intersections of halfspaces. In *Proc. of the 47th Symposium on Foundations of Computer Science (FOCS)*, pages 553–562, 2006.

[17] A. R. Klivans and A. A. Sherstov. Improved lower bounds for learning intersections of halfspaces. In *Proc. of the 19th Conf. on Learning Theory (COLT)*, pages 335–349, 2006.

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

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

[20] N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman. Complexity measures of sign matrices. *Combinatorica*, 2006. To appear. Manuscript at `http://www.cs.huji.ac.il/˜nati/PAPERS/complexity_matrices.ps.gz`.

[21] S. V. Lokam. Spectral methods for matrix rigidity with applications to size-depth trade-offs and communication complexity. *J. Comput. Syst. Sci.*, 63(3):449–473, 2001.

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

[23] I. Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39(2):67–71, 1991.

[24] N. Nisan. The communication complexity of threshold gates. In *Proceedings of "Combinatorics, Paul Erdos is Eighty"*, pages 301–315, 1993.

[25] N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

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

[27] R. Paturi and J. Simon. Probabilistic communication complexity. *J. Comput. Syst. Sci.*, 33(1):106–123, 1986.

[28] A. A. Razborov. Bounded-depth formulae over the basis $\{\&, \oplus\}$ and some combinatorial problems. *Complexity Theory and Applied Mathematical Logic*, vol. "Problems of Cybernetics":146–166, 1988. In Russian.

[29] A. A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

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

[31] A. A. Sherstov. Halfspace matrices. In *Proc. of the 22nd Conf. on Computational Complexity (CCC)*, pages 83–95, 2007.

[32] A. A. Sherstov. The pattern matrix method for lower bounds on quantum communication. Technical Report TR-07-46, The Univ. of Texas at Austin, Dept. of Computer Sciences, September 2007.

[33] A. A. Sherstov. Powering requires threshold depth 3. *Inf. Process. Lett.*, 102(2–3):104–107, 2007.

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

[35] L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.

A Pattern Matrices Inside Communication Matrices

The purpose of this appendix is to prove Lemma 6.1, needed in Section 6.

Lemma 6.1 (Restated from p. 28). *Let $D : \{0, 1, \ldots, m\} \to \{0, 1\}$ be a given predicate. Let F be the $(2v, v, f)$-pattern matrix, where $v \leqslant m/4$ and $f(z) \stackrel{\text{def}}{=} (-1)^{D(|z|)}$. Then F is a submatrix of*

$$\left[(-1)^{D(|x \wedge y|)}\right]_{x \in \{0,1\}^m, \, y \in \{0,1\}^m}. \tag{A.1}$$

Proof (adapted from Sherstov [32, Lem. 6.1]). By definition,

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

We will define one-to-one maps

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

such that

$$|\, x|_V \, \oplus \, w \,| \quad = \quad |\, \alpha(x) \wedge \beta(V, w) \,| \qquad \text{for all } x, V, w. \tag{A.2}$$

Obviously, this will mean that F is a submatrix of (A.1).

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

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

Define β by

$$\beta(V, w) \quad \stackrel{\text{def}}{=} \quad \gamma(i_1, w_1) \, \gamma(i_2, w_2) \, \cdots \, \gamma(i_v, w_v) \, 0^{m-4v},$$

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

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

It is now straightforward to verify (A.2). $\qquad \square$