Disjointness is hard in the multi-party number-on-the-forehead model

Troy Lee
Department of Computer Science
Rutgers University *

Adi Shraibman
Department of Mathematics
Weizmann Institute of Science †

Abstract

We show that disjointness requires randomized communication $\Omega\left(\frac{n^{1/2k}}{(k-1)2^{k-1}2^{2^k-1}}\right)$ in the general k-party number-on-the-forehead model of complexity. The previous best lower bound was $\Omega\left(\frac{\log n}{k-1}\right)$. By results of Beame, Pitassi, and Segerlind, this implies $2^{n^{\Omega(1)}}$ lower bounds on the size of tree-like Lovász-Schrijver proof systems needed to refute certain unsatisfiable CNFs, and super-polynomial lower bounds on the size of any tree-like proof system whose terms are degree-d polynomial inequalities for $d = \log\log n - O(\log\log\log n)$.

To prove our bound, we develop a new technique for showing lower bounds in the number-on-the-forehead model which is based on the norm induced by cylinder intersections. This bound naturally extends the linear program bound for rank useful in the two-party case to the case of more than two parties, where the fundamental concept of monochromatic rectangles is replaced by monochromatic cylinder intersections. Previously, the only general method known for showing lower bounds in the unrestricted number-on-the-forehead model was the discrepancy method, which can only show bounds of size $O(\log n)$ for disjointness.

To analyze the bound given by our new technique for the disjointness function, we extend an elegant framework developed by Sherstov in the two-party case which relates polynomial degree to communication complexity. Using this framework we are able to obtain bounds for any tensor of the form $F(x_1, \ldots, x_k) = f(x_1 \wedge \ldots \wedge x_k)$ where f is a function which only depends on the number of ones in the input.

1 Introduction

Since its introduction by Yao [Yao79] nearly thirty years ago, communication complexity has become a key concept in complexity theory and theoretical computer science in general. Part of its appeal is that it can be applied to many different computational models, for example to formula

*Supported by a National Science Foundation Mathematical Sciences Postdoctoral Fellowship. Email: troyjlee@gmail.com

†Email: adi.shraibman@weizmann.ac.il

size and circuit depth, branching programs, VLSI design, and time-space trade-offs for Turing machines (see [KN97] for more details).

Perhaps the area of communication complexity which remains the most mysterious today is the k-party "number-on-the-forehead" model, originally introduced by Chandra, Furst and Lipton [CFL83]. In this model, k parties wish to compute a function $f : (\{0,1\}^n)^k \to \{-1,1\}$ where player i has the input $x_i \in \{0,1\}^n$ "on his forehead." That is to say, player i has knowledge of the entire input *except* for the string x_i. The communication is written "on the blackboard" so that all players have knowledge of each message. The large overlap in the player's knowledge is part of what makes showing lower bounds in this model so difficult. This difficulty, however, is rewarded by the richness of consequences of such lower bounds: for example, showing a linear lower bound on an explicit function for $k = n^\epsilon$ many players would give an explicit function which requires superpolynomial size ACC circuits.

While showing such bounds still seems distant, we do know of explicit functions which require reasonably large communication in this model. Babai, Nisan, and Szegedy [BNS89] show that the inner product function generalized to k-parties requires randomized communication $\Omega(n/4^k)$, and for other explicit functions slightly larger bounds of size $\Omega(n/2^k)$ are known.

For other basic functions, however, there are huge gaps in our knowledge. One example is the disjointness function, where the goal of the players is to determine if there is an index j such that every string x_i has a one in position j. The best protocol known for disjointness has communication $O(kn/2^k)$ [Gro94]—this upper bound in fact holds for any function whose value only depends on the size of the intersection of the strings x_i. On the other hand, the best lower bound in the general number-on-the-forehead model is $\Omega(\log n/(k-1))$ [BPSW06, Tes02]. A major obstacle toward proving better lower bounds on disjointness is that the discrepancy method, a very common and general technique in communication complexity, can only show bounds of size $O(\log n)$. Because of this limitation, disjointness is always one of the most recalcitrant problems in any model—for example, in the two-party randomized and quantum models, determining the complexity of disjointness was a long-standing open problem which required the development of novel techniques to resolve [KS87, Raz92, Raz03].

In the multiparty case, this difficulty is compounded by the fact that discrepancy is essentially the only method available to show lower bounds in the general number-on-the-forehead model. Indeed, Kushilevitz and Nisan [KN97] say, "The only technique from two-party communication complexity that generalizes to the multiparty case is the discrepancy method."

We show that disjointness requires randomized communication $\Omega\left(\frac{n^{1/2k}}{(k-1)2^{k-1}2^{2^{k-1}}}\right)$ in the general k-party number-on-the-forehead model by developing a new technique which extends the discrepancy method in a natural way. This bound is in the same ballpark as the previously best known results for very restricted classes of protocols. Viola and Wigderson [VW07b] show a bound of $\Omega(n^{1/(k-1)}/k^{O(k)})$ on the communication needed by one-way protocols to solve k-party disjointness. For the $k = 3$ case, Beame, Pitassi, Segerlind, and Wigderson [BPSW06] show a bound of $\Omega(n^{1/3})$ for protocols where the first party speaks once, and then the two other parties interact arbitrarily.

To analyze the bound given by our technique for disjointness, we use the elegant framework developed by Sherstov [She07b] which relates the polynomial degree of a function f to the com-

plexity of a matrix—or tensor, in our case—which is formed from f in a structured way. This allows us to show lower bounds not only for disjointness, but for any function whose value depends only on the size of the intersection of the players' inputs.

After our work was completed and a version of our manuscript circulated, Chattopadhyay and Ada independently obtained similar lower bounds [Cha07b].

1.1 Consequences for Lovász-Schrijver proof systems and beyond

There is an additional motivation to studying the complexity of disjointness in the number-on-the-forehead model. Beame, Pitassi, and Segerlind [BPS06] show that bounds on disjointness imply strong lower bounds on the size of refutations of certain unsatisfiable formulas, for a very general class of proof systems. We now introduce and motivate the study of these proof systems.

As linear and semidefinite programming are some of the most sophisticated polynomial time algorithms which have been developed, it is natural to ask how they fare when pitted against NP-complete problems. For many NP-complete problems, there is a very natural approach to solving them via linear or semidefinite programming: namely, we first formulate the problem as optimizing a convex function over the Boolean cube, i.e. with variables subject to the quadratic constraints $x_i^2 = x_i$. We then relax these quadratic constraints to linear or semidefinite constraints to obtain a program which can be solved in polynomial time. For example, a linear relaxation of $x_i^2 = x_i$ may simply be the constraint $0 \leq x_i \leq 1$. Such a relaxation already gives a linear program with approximation ratio of 2 for the problem of vertex cover. Semidefinite constraints are in general more complicated, but there are several "automatic" ways of generating valid semidefinite inequalities—that is, semidefinite inequalities satisfied by all Boolean solutions of the original problem. Perhaps the best known of these is the Lovász-Schrijver "lift and project" method [LS91]. The seminal 0.878-approximation algorithm for MAXCUT of Goemans and Williamson [GW95] can be obtained by relaxing the natural Boolean programming problem with semidefinite constraints obtained by one application of the Lovász-Schrijver method.

As these techniques have given impressive results in approximation algorithms, it is natural to ask if they can also be used to efficiently obtain exact solutions. Namely, how many inequalities need to be added in general until all fractional optima are eliminated and only true Boolean solutions remain?

One way to address this question is to consider proof systems with derivation rules based on linear programming or the Lovász-Schrijver method. Our particular application will look at the size of proofs needed to refute unsatisfiable formulas. Given a CNF ϕ, we can naturally represent the satisfiability of ϕ as the satisfiability of a system of linear inequalities, one for each clause. For example, the clause $x_1 \vee x_4 \vee \neg x_5$ would be represented as $x_1 + x_4 + (1 - x_5) \geq 1$. Suppose that ϕ is unsatisfiable. Then consider a proof system in which the "axioms" are the inequalities obtained from the clauses of ϕ, and the goal is to derive the contradiction $0 \geq 1$. By the results of [BPS06], our results on disjointness imply that there are unsatisfiable formulas such that any refutation obtained by generating new inequalities by the Lovász-Schrijver method in a "tree-like" way requires size $2^{n^{\Omega(1)}}$. For a standard formulation of the Lovász-Schrijver method known as LS_+, bounds of size $2^{\Omega(n)}$ for tree-like proofs have already been shown by very different methods [IK06].

The advantage of the number-on-the-forehead communication complexity approach, however, is that it can also be applied to much more powerful proof systems which are currently untouchable by other approaches. Beame, Pitassi, and Segerlind [BPS06] show that lower bounds on k-party communication complexity of disjointness give lower bounds on the size of tree-like proofs of certain unsatisfiable CNFs $\phi(x)$, where the derivation rule is as follows: from inequalities f, g of degree $k-1$ in x, we are allowed to conclude a degree $k-1$ inequality h if every Boolean assignment to x which satisfies f and g also satisfies h. Lovász-Schrijver proof systems are a special case of such degree-2 systems. Our bounds on disjointness imply the existence of unsatisfiable formulas whose refutation requires super-polynomial size tree-like degree-k proofs, for any $k = \log\log n - O(\log\log\log n)$. The aforementioned lower bounds on LS_+ proof systems strongly rely on specific geometrical properties of the Lovász-Schrijver operator—showing superpolynomial bounds on the size of tree-like proofs in the more general degree-k model was previously open even in the case $k = 2$.

2 Preliminaries and notation

We let $[n] = \{1, \ldots, n\}$. For multi-party communication complexity it is convenient to work with tensors, the generalization of matrices to higher dimensions. If an element of a tensor A is specified by k indices, we say that A has rank k. A tensor for which all entries are in $\{-1, 1\}$ we call a sign tensor. For a distributed function $f : X_1 \times \ldots \times X_k \to \{-1, 1\}$, we define the communication tensor corresponding to f to be a rank k tensor A_f where $A_f[x_1, \ldots, x_k] = f(x_1, \ldots, x_k)$. We identify f with its communication tensor. For a set $Z \subseteq X_1 \times \ldots \times X_k$ we let $\chi(Z)$ be its characteristic tensor where $\chi(Z)[x_1, \ldots, x_k] = 1$ if $(x_1, \ldots, x_k) \in Z$ and is 0 otherwise.

For a sign tensor A, we denote by $D^k(A)$ the deterministic communication complexity of A in the k-party number-on-the-forehead model. The corresponding randomized communication complexity with error bound $\epsilon \geq 0$ is denoted $R^k_\epsilon(A)$. When we drop the superscript it is to be assumed that the number of parties is equal to the rank of A.

We use the shorthand $A \geq c$ to indicate that all of the entries of A are at least c. The Hadamard or entrywise product of two tensors A and B is denoted by $A \circ B$. Their inner product is denoted $\langle A, B\rangle = \sum_{x_1,\ldots,x_k} A[x_1, \ldots, x_k]B[x_1, \ldots, x_k]$. The ℓ_1 and ℓ_∞ norms of a tensor A are $\|A\|_1 = \sum_{x_1,\ldots,x_k} |A[x_1, \ldots, x_k]|$ and $\|A\|_\infty = \max_{x_1,\ldots,x_k} |A[x_1, \ldots, x_k]|$, respectively.

We also need some basic elements of Fourier analysis. For $S \subseteq [n]$ we define $\chi_S : \{0, 1\}^n \to \{-1, 1\}$ as $\chi_S(x) = (-1)^{\sum_{i\in S} x_i}$. As the χ_S form an orthogonal basis, for any function $f : \{0, 1\}^n \to \mathbb{R}$ we have a unique representation

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

where $\hat{f}(S) = (1/2^n)\langle f, \chi_S\rangle$.

3 The Method

In this section we present a method for proving lower bounds on randomized communication complexity in the number-on-the-forehead model that generalizes and significantly strengthens the discrepancy method.

3.1 Cylinder intersection norm

In two-party communication complexity, a key role is played by combinatorial rectangles—subsets of the form $Z_1 \times Z_2$ where Z_1 is a subset of inputs to Alice and Z_2 is a subset of inputs to Bob. The analogous concept in the number-on-the-forehead model of multi-party communication complexity is that of a cylinder intersection.

Definition 1 (Cylinder intersection) *A subset $Z_i \subseteq X_1 \times \ldots \times X_k$ is called a cylinder in the i^{th} dimension if membership in Z_i does not depend on the i^{th} coordinate. That is, for every $(z_1, \ldots, z_i, \ldots, z_k) \in Z_i$ and $z_i' \in X_i$ it also holds that $(z_1, \ldots, z_i', \ldots, z_k) \in Z_i$. A set Z is called a cylinder intersection if it can be expressed as $Z = \cap_{i=1}^{k} Z_i$ where each Z_i is a cylinder in the i^{th} dimension.*

The reason why cylinder intersections are so important is that a successful protocol partitions the communication tensor into cylinder intersections, each of which is monochromatic with respect to the function f. This leads us to our next definition:

Cylinder intersection norm We denote by μ the norm induced by the absolute convex hull of the characteristic functions of all cylinder intersections. That is, for a k-tensor M

$$\mu(M) = \min \left\{ \sum_i |\alpha_i| : M = \sum_i \alpha_i \chi(Z_i) \right\}$$

where each Z_i is a cylinder intersection, and $\chi(Z_i)$ is a k-tensor where $\chi(Z_i)[x_1, \ldots, x_k] = 1$ if $(x_1, \ldots, x_k) \in Z_i$ and 0 otherwise.

Remark 2 *In our definition of μ above we chose to take $\chi(Z_i)$ as $\{0, 1\}$ tensors. One can alternatively take them to be ± 1 valued tensors—a form which is sometimes easier to bound—without changing much. One can show*

$$\mu(M) \geq \mu_{\pm 1}(M) \geq 2^{-k} \mu(M).$$

where $\mu_{\pm 1}(M)$ is defined as above with $\chi(Z_i)$ taking values from $\{-1, 1\}$.

We further remark that in the two dimensional case, μ is very closely related to a semidefinite programming quantity γ_2 introduced to communication complexity by Linial and Shraibman. Indeed, for matrices M we have $\mu(M) = \Theta(\gamma_2(M))$ [LS07].

A successful communication protocol for a sign k-tensor M partitions M into monochromatic cylinder intersections, $Z_1, Z_2, \ldots, Z_{2^{D^k(M)}}$. Hence $M = \sum_i \alpha_i \chi(Z_i)$ where the coefficients α_i are either 1 or -1. Therefore

Theorem 3 *For every sign k-tensor M, $D^k(M) \geq \log(\mu(M))$.*

A randomized protocol is simply a probability distribution over deterministic protocols. This gives us the following fact:

Fact 4 *A sign k-tensor M satisfies $R^k_\epsilon(M) \leq c$ if and only if there are sign k-tensors A_i for $i = 1, \ldots, \ell$ satisfying $D^k(A_i) \leq c$ and a probability distribution $(p_1, \ldots, p_\ell)$ such that*

$$\|M - \sum_i p_i A_i\|_\infty \leq 2\epsilon.$$

To lower bound randomized communication complexity we consider an approximate variant of the cylinder intersection norm.

Definition 5 (Approximate cylinder intersection norm) *Let M be a sign k-tensor, and $\alpha \geq 1$. We define the α-approximate cylinder intersection norm as*

$$\mu^\alpha(M) = \min_{M'} \{\mu(M') : 1 \leq M \circ M' \leq \alpha\}$$

In words, we take the minimum of the cylinder intersection norm over all tensors M' which are signed as M and have entries with magnitude between 1 and α. Considering the limiting case as $\alpha \to \infty$ motivates us to define

$$\mu^\infty(M) = \min_{M'} \{\mu(M') : 1 \leq M \circ M'\}$$

One should note that $\mu^\alpha(M) \leq \mu^\beta(M)$ for $1 \leq \beta \leq \alpha$.

The following theorem is an immediate consequence of the definition of approximate cylinder norm and Fact 4.

Theorem 6 *Let M be a sign k-tensor, and $0 \leq \epsilon < 1/2$. Then*

$$R^k_\epsilon(M) \geq \log(\mu^\alpha(M)) - \log(\alpha_\epsilon)$$

where $\alpha_\epsilon = 1/(1-2\epsilon)$ and $\alpha \geq \alpha_\epsilon$.

Proof: Let p_i and A_i for $1 \leq i \leq \ell$ be as in Fact 4. We take

$$B = \frac{1}{1-2\epsilon} \sum_{i=1}^{\ell} p_i A_i.$$

Notice that $1 \leq B \circ M \leq \alpha_\epsilon$, and hence by Definition 5

$$\mu^{\alpha_\epsilon}(M) \leq \mu(B).$$

Employing the fact that μ is a norm and Theorem 3, we get

$$\mu(B) \leq \frac{1}{1-2\epsilon} \sum_i p_i \mu(A_i) \leq \frac{1}{1-2\epsilon} \sum_i p_i 2^{D^k(A_i)} \leq \frac{2^{R^k_\epsilon(M)}}{1-2\epsilon}.$$

□

3.2 Employing duality

We now have a quantity, $\mu^\alpha(M)$, which can be used to prove lower bounds for randomized communication complexity in the number-on-the-forehead model. As this quantity is defined in terms of a minimization, however, it seems in itself a difficult quantity to bound from below.

In this section, we employ the duality theory of linear programming to find an equivalent formulation of $\mu^\alpha(M)$ in terms of a maximization problem. This makes the task of proving lower bounds for $\mu^\alpha(M)$ much easier, as the $\forall$ quantifier we had to deal with before is now replaced by an $\exists$ quantifier.

As it turns out, in order to prove lower bounds on $\mu^\alpha(M)$ we will need to understand the dual norm of μ, denoted μ^*. The standard definition of a dual norm is

$$\mu^*(Q) = \max_{M:\mu(M)\leq 1} \langle M, Q \rangle,$$

for every tensor Q. Since the unit ball of μ is the absolute convex hull of the characteristic vectors of cylinder intersections, we can alternatively write

$$\mu^*(Q) = \max_Z |\langle Q, \chi(Z) \rangle|$$

where the maximum is taken over all cylinder intersections Z.

We will use the following form for our lower bounds:

Theorem 7 *Let M be a sign tensor and $1 \leq \alpha$.*

$$\mu^\alpha(M) = \max_Q \frac{(1+\alpha)\langle M, Q \rangle + (1-\alpha)\|Q\|_1}{2}$$
$$\textit{s.t. } \mu^*(Q) \leq 1$$

When $\alpha = \infty$ we have

$$\mu^\infty(M) = \max_{Q: M \circ Q \geq 0} \langle M, Q \rangle$$
$$\textit{s.t. } \mu^*(Q) \leq 1$$

Proof: We treat the case $1 \leq \alpha < \infty$ first. We can write $\mu^\alpha(M)$ as a linear program as follows. For each cylinder intersection Z_i let $X_i = \chi(Z_i)$. Then

$$\begin{aligned} \mu^\alpha(M) &= \min_{p,q} \sum_i p_i + q_i \\ \text{s.t.} \quad & 1 \leq \left(\sum_i (p_i - q_i) X_i \right) \circ M \leq \alpha \\ & p_i, q_i \geq 0 \end{aligned}$$

Taking the dual of this program in the straightforward way, we obtain

$$\begin{aligned} \mu^\alpha(M) &= \max_Q \frac{(1+\alpha)\langle M, Q\rangle + (1-\alpha)\|Q\|_1}{2} \\ \text{s.t.} \quad & |\langle X_i, Q\rangle| \leq 1, \text{ for all } X_i \end{aligned}$$

For $\alpha = \infty$ we get the same program as above without the constraint $(\sum_i (p_i - q_i) X_i) \circ M \leq \alpha$. Dualizing this program gives the desired result. □

Observing the bounds in Theorem 7 we see that to lower bound $\mu^\alpha(M)$ it suffices to find a tensor Q with $\mu^*(Q) \leq 1$ that has a large inner product with M. In Section 4 we discuss a technique for showing bounds on μ^*.

3.3 The discrepancy method

Virtually all lower bounds in the general number-on-the-forehead model have used the discrepancy method, which we now recall.

Definition 8 *Let M be a sign k-tensor, and let P be a probability distribution on its entries. The discrepancy of M with respect to P, written $\mathrm{disc}_P(M)$ is*

$$\mathrm{disc}_P(M) = \max_Z \langle M \circ P, \chi(Z)\rangle$$

where the maximum is taken over cylinder intersections Z. We further define the general discrepancy as

$$\mathrm{disc}(M) = \min_P \mathrm{disc}_P(M)$$

where the minimum is taken over all probability distributions P.

The discrepancy method turns out to be equivalent to $\mu^\infty(M)$.

Theorem 9

$$\mu^\infty(M) = \frac{1}{\mathrm{disc}(M)}.$$

Proof: By Theorem 7, for every sign tensor M

$$\mu^\infty(M) = \max_{Q \circ M \geq 0} \{\langle M, Q\rangle : \mu^*(Q) \leq 1\}$$

We can rewrite this as

$$\mu^\infty(M) = \max_{Q \circ M \geq 0} \frac{\langle M, Q\rangle}{\mu^*(Q)} = \max_{P : P \geq 0} \frac{\langle M, M \circ P\rangle}{\mu^*(M \circ P)}$$

As both numerator and denominator are homogeneous, we have

$$\begin{aligned}
\mu^\infty(M) &= \max_{\substack{P : P \geq 0 \\ \|P\|_1 = 1}} \frac{\langle M, M \circ P\rangle}{\mu^*(M \circ P)} \\
&= \max_{\substack{P : P \geq 0 \\ \|P\|_1 = 1}} \frac{1}{\mu^*(M \circ P)} \\
&= \frac{1}{\mathrm{disc}(M)}.
\end{aligned}$$

□

4 Techniques to bound $\mu^*(Q)$

In the last section, we saw that to bound the randomized number-on-the-forehead communication complexity of a tensor M, it suffices to find a tensor Q such that $\langle M, Q\rangle$ is large and $\mu^*(Q)$ is small. The first quantity is simply a sum and is in general not too hard to compute. Upper bounding $\mu^*(Q)$ is more subtle. In this section, we review some techniques for doing this.

In upper bounding the magnitude of the largest eigenvalue of A, a common thing is to consider the matrix AA^T, and use the fact that $\|A\|^2 \leq \lambda_1(AA^T)$. We will try to do a similar thing in upper bounding $\mu^*(Q)$. In analogy with AA^T we make the following definition:

Definition 10 (Contraction product) *Let A be a k-tensor with entries indexed by elements from $X_1 \times \ldots \times X_k$. We define the contraction product of A along X_1, denoted $A \bullet_1 A$, to be a $2(k-1)$-tensor with entries indexed by elements from $X_2 \times X_2 \times \ldots \times X_k \times X_k$. The $x_2, x_2', \ldots, x_k, x_k'$ entry is defined to be*

$$A \bullet_1 A[x_2, x_2', \ldots, x_k, x_k'] = \sum_{x_1} \prod_{y_2 \in \{x_2, x_2'\}, \ldots, y_k \in \{x_k, x_k'\}} A[x_1, y_2, \ldots, y_k]$$

The contraction product may be defined along other dimensions mutatis mutandis.

Notice that when A is a matrix $A \bullet_2 A$ corresponds to AA^T. When A is a m_1-by-m_2 matrix, the fact that $\|A\|^2 \leq \lambda_1(A \bullet_2 A)$ implies that $\mu^*(A)^2 \leq m_1 \mu^*(A \bullet_1 A)$ or $\mu^*(A)^2 \leq m_2 \mu^*(A \bullet_2 A)$. The next claim gives the general result for k-tensors. This approach is fairly standard and one can find similar statements in, for example, [Cha07, VW07a].

Claim 11 *Let A be a k-tensor with dimensions $(m_1, \ldots, m_k)$. Then*

$$\mu^*(A)^{2^{k-1}} \leq m_1^{2^{k-1}-1} m_2^{2^{k-1}-2} \cdots m_k^{2^{k-1}-2} \mu^*(A \bullet_1 A)$$

Proof: Let $\phi_1, \ldots, \phi_k$ be 0/1 valued functions which maximize $\mu^*(A)$, and where ϕ_i does not depend on the i^{th} variable x_i. Then we have

$$\begin{aligned}
\mu^*(A) &= \sum_{x_1,\ldots,x_k} A[x_1, \ldots, x_k] \phi_1(x_2, \ldots, x_k) \cdots \phi_k(x_1, \ldots, x_{k-1}) \\
&= \sum_{x_1,\ldots,x_{k-1}} \phi_k(x_1, \ldots, x_{k-1}) \sum_{x_k} A[x_1, \ldots, x_k] \phi_1(x_2, \ldots, x_k) \cdots \phi_{k-1}(x_1, \ldots, x_{k-2}, x_k)
\end{aligned}$$

Applying the Cauchy-Schwarz inequality we find

$$\begin{aligned}
\mu^*(A)^2 &\leq m_1 m_2 \cdots m_{k-1} \sum_{x_1,\ldots,x_{k-1}} \left(\sum_{x_k} A[x_1, \ldots, x_k] \phi_1(x_2, \ldots, x_k) \cdots \phi_{k-1}(x_1, \ldots, x_{k-2}, x_k) \right)^2 \\
&\leq m_1 m_2 \cdots m_{k-1} \sum_{\substack{x_1,\ldots,x_{k-1} \\ x_k, x'_k}} \prod_{y_k \in \{x_k, x'_k\}} A[x_1, \ldots, y_k] \phi_1(x_2, \ldots, y_k) \cdots \phi_{k-1}(x_1, \ldots, x_{k-2}, y_k)
\end{aligned}$$

The result follows by repeating the above process in turn for each variable $x_{k-1}, \ldots, x_2$. □

4.1 Example: Hadamard tensors

We give an example to show how Claim 11 can be used in conjunction with our μ method. Let H be a N-by-N Hadamard matrix. We show that $\mu^\infty(H) \geq \sqrt{N}$. Indeed, simply let the witness matrix Q be H itself. Incidentally, this corresponds to taking the uniform probability distribution in the discrepancy method. With this choice we clearly have $H \circ Q \geq 0$, and so

$$\mu^\infty(H) \geq \frac{\langle H, H \rangle}{\mu^*(H)} = \frac{N^2}{\mu^*(H)}$$

Now we bound $\mu^*(H)$ using Claim 11 which gives:

$$\mu^*(H)^2 \leq N \mu^*(H \bullet_1 H) = N^3$$

As $H \bullet_1 H$ has nonzero entries only on the diagonal, and these entries are of magnitude N.

Ford and Gál [FG05] extend the notion of matrix orthogonality to tensors, defining what they call Hadamard tensors.

Definition 12 (Hadamard tensor) *Let H be a sign k-tensor of dimensions $(N, \ldots, N)$. We say that H is a Hadamard tensor if*

$$(H \bullet_1 H)[x_2, x_2', \ldots, x_k, x_k'] = 0$$

whenever $x_i \neq x_i'$ for all $i = 2, \ldots, k$.

The simple proof above for Hadamard matrices can be easily extended to Hadamard tensors:

Theorem 13 (Ford and Gál [FG05]) *Let H be a rank k Hadamard tensor. Then*

$$\mu^{\infty}(H) \geq \left(\frac{N}{k-1}\right)^{1/2^{k-1}}$$

Proof: We again take the witness Q to be H itself. This clearly satisfies $H \circ Q \geq 0$, and so

$$\mu^{\infty}(H) \geq \frac{\langle H, H \rangle}{\mu^*(H)} = \frac{N^k}{\mu^*(H)}$$

It now remains to upper bound $\mu^*(H)$ which we do by Claim 11. This gives us

$$\mu^*(H)^{2^{k-1}} \leq N^{k2^{k-1}-2k+1} \mu^*(H \bullet_1 H)$$

The "Hadamard" property of H lets us easily upper bound $H \bullet_1 H$. We have

$$\Pr[\vee_{i=2}^{k}(x_i = x_i')] \leq \frac{k-1}{N}$$

by a union bound. Thus the number of non-zero entries in $H \bullet_1 H$ is at most $N^{2(k-1)}\frac{k-1}{N}$. Each non-zero entry has magnitude N. Hence, we obtain

$$\mu^*(H)^{2^{k-1}} \leq (k-1)N^{k2^{k-1}-2k+1}N^{2k-2} = (k-1)N^{k2^{k-1}-1}.$$

Putting everything together, we have

$$\mu^{\infty}(H) \geq \left(\frac{N}{k-1}\right)^{1/2^{k-1}}$$

□

Remark 14 *By doing a more careful inductive analysis, Ford and Gál obtain this result without the $k-1$ term in the denominator. They also construct explicit examples of Hadamard tensors.*

5 Lower bounds on μ^α for pattern tensors

In a series of works, Sherstov [She07a, She07b] develops a beautiful framework for proving lower bounds on the discrepancy and the communication complexity of sign matrices A of a particular form, which he calls pattern matrices. A pattern matrix A_f is derived from a Boolean function f in a structured way, according to a particular "pattern". For such matrices, Sherstov is able to relate discrepancy to the sign degree of f. He terms this the degree-discrepancy lemma [She07a]. In a follow-up work [She07b], he relates the bounded-error approximate degree of f to an approximate version of the trace norm developed by Razborov [Raz03] to give a new, more transparent, proof for the lower bounds on the quantum communication complexity of symmetric predicates.

Chattopadhyay [Cha07] extends Sherstov's degree-discrepancy lemma to relate the sign degree of f to discrepancy, or equivalently $\mu^\infty(A_f)$, of an appropriately defined pattern *tensor* A_f. In this section, we take the natural step to generalize this result to relate the approximate degree of f to $\mu^\alpha(A_f)$ for any α. In fact, the advantage of our approach is that μ^α provides a uniform framework in which one can view all of these results, and seamlessly treats both the case of discrepancy ($\alpha = \infty$) and bounded-error (bounded α).

In Section 5.1 we describe a key lemma which relates the approximate polynomial degree of f to the existence of a hard input "distribution" for f. This will only truly correspond to a distribution in the case of discrepancy—otherwise it can take on negative values. Then in Section 5.2 we use this distribution, together with the machinery developed in Section 4 to show our main result relating the α-approximate degree of f to $\mu^\alpha(A_f)$.

5.1 Dual polynomials

We define approximate degree in a slightly non-standard way so that we may simultaneously treat the bounded α and $\alpha = \infty$ cases.

Definition 15 *Let $f : \{0,1\}^n \to \{-1,1\}$. For $\alpha \geq 1$ we say that a function g gives an α-approximation to f if $1 \leq g(x)f(x) \leq \alpha$ for all $x \in \{0,1\}^n$. Similarly we say that g gives an ∞-approximation to f if $1 \leq g(x)f(x)$ for all $x \in \{0,1\}^n$. We let the α-approximate degree of f, denoted $\deg_\alpha(f)$, be the smallest degree of a function g which gives an α-approximation to f.*

Remark 16 *In a more standard scenario, one is considering a 0/1 valued function f and defines the approximate degree as $\deg'_\epsilon(f) = \min\{\deg(g) : \|f - g\|_\infty \leq \epsilon\}$. Letting $f_\pm$ be the sign representation of f, one can see that for $0 \leq \epsilon < 1/2$ our definition is equivalent to the standard one in the following sense: $\deg'_\epsilon(f) = \deg_{\alpha_\epsilon}(f_\pm)$ where $\alpha_\epsilon = \frac{1+2\epsilon}{1-2\epsilon}$.*

For a fixed degree d, let $\alpha_d(f)$ be the smallest value of α for which there is a degree d polynomial which gives an α-approximation to f. Notice that $\alpha_d(f)$ can be written as a linear program. Namely, let $B(n,d) = \sum_{i=0}^d \binom{n}{i}$, and Φ be a 2^n-by-$B(n,d)$ incidence matrix, with rows labelled by strings $x \in \{0,1\}^n$ and columns labelled by monomials of degree at most d. We set $\Phi(x,m) = (-1)^{m(x)}$, where $m(x)$ is the evaluation of the monomial m on input x. Then

$$\alpha_d(f) = \min_y \{\|\Phi y\|_\infty : 1 \leq \Phi y \circ f\}$$

If this program is infeasible with value α—that is, if there is no degree d polynomial which gives an α-approximation to f—then the feasibility of the dual of this program will give us a "witness" to this fact. It is this witness that we will use to construct a tensor Q which witnesses that μ^α is large.

Lemma 17

$$\alpha_d(f) = \max_v \left\{ \frac{1 + \langle v, f \rangle}{1 - \langle v, f \rangle} : \|v\|_1 = 1, v^T \Phi = 0 \right\}$$

Proof: Follows from duality theory of linear programming. □

Corollary 18 (cf. Sherstov Corollary 3.3.1 [She07b]) *Let $f : \{0,1\}^n \to \{-1,1\}$ and let $d = \deg_\alpha(f)$. Then there exists a function $v : \{0,1\}^n \to \mathbb{R}$ such that*

1. *$\langle v, g \rangle = 0$ for any function g of degree $< d$.*
2. *$\|v\|_1 = 1$.*
3. *$\langle v, f \rangle \geq \frac{\alpha - 1}{\alpha + 1}$.*

In particular, when $\alpha = \infty$, there is a function $v : \{0,1\}^n \to \mathbb{R}$ satisfying items (1), (2), and such that $v(x)f(x) \geq 0$ for all $x \in \{0,1\}^n$.

5.2 Pattern Tensors

We now define a pattern tensor of rank k. Take m and M such that m divides M. Divide $[M]$ into m many contiguous blocks, each of size M/m. Let $S^1, S^2, \ldots, S^{k-1} \subseteq [M/m]^m$, be vectors of m elements from $[M/m]$. We will use the notation $S^i[t]$ to refer to the t^{th} element of S^i. Finally, let $x \in \{0,1\}^{m(M/m)^{k-1}}$, where we think of x as a k-tensor of dimensions $(m, M/m, \ldots, M/m)$.

For a function $\phi : \{0,1\}^m \to \mathbb{R}$ define the (k, m, M, ϕ) pattern tensor, denoted $A_{k,m,M,\phi}$ by

$$A_{k,m,M,\phi}(x, S^1, S^2, \ldots, S^{k-1}) = \phi(x_{1,i_1^1,i_1^2,\ldots,i_1^{k-1}} \ldots x_{m,i_m^1,i_m^2,\ldots,i_m^{k-1}})$$

where $S^j = \{i_1^j, i_2^j, \ldots, i_m^j\}$ for $j = 1, \ldots, k-1$. We will use the shorthand $x|_{S^1,\ldots,S^{k-1}}$ to refer to the m-bit string $x_{1,i_1^1,i_1^2,\ldots,i_1^{k-1}} \ldots x_{m,i_m^1,i_m^2,\ldots,i_m^{k-1}}$.

Now we are ready to state our main theorem.

Theorem 19 *For non-negative integers k, m and $M \geq e(k-1)2^{2^{k-1}}m^2$, and a Boolean function f on m variables*

$$\log \mu^\alpha(A_{k,m,M,f}) \geq \deg_{\alpha_0}(f)/2^{k-1} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1},$$

for every $1 \leq \alpha < \alpha_0 < \infty$. Furthermore,

$$\log \mu^\infty(A_{k,m,M,f}) \geq \deg_\infty(f)/2^{k-1}.$$

Proof: For simplicity we will drop the subscripts and just write A for $A_{k,m,M,f}$. Recall that

$$\mu^{\alpha}(A) = \max_{Q:\|Q\|_1=1} \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)}{2\mu^*(Q)}$$
$$\mu^{\infty}(A) = \max_{Q:Q\circ A\geq 0} \frac{\langle A, Q\rangle}{\mu^*(Q)}.$$

Let v be the vector from Corollary 18 which witnesses that the α_0-approximate degree of f is at least d. We let Q be $1/c$ times the (k, m, M, v) pattern tensor, where $c = 2^{m(M/m)^{k-1}-m}\left(\frac{M}{m}\right)^{(k-1)m}$. With this choice of normalization we have $\|Q\|_1 = 1$.

Lower bound on $\langle \mathbf{A}, \mathbf{Q}\rangle$ First consider the case $1 \leq \alpha < \infty$. Then we have $\langle v, f\rangle \geq (\alpha_0 - 1)/(\alpha_0 + 1)$, and so, by our choice of normalization, $\langle A, Q\rangle \geq (\alpha_0 - 1)/(\alpha_0 + 1)$. This allows us to bound $(1/2)$ the term in the numerator of $\mu^{\alpha}(A)$ as follows:

$$\frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)}{2} \geq \frac{(\alpha_0 - 1)(1+\alpha) + (\alpha_0 + 1)(1-\alpha)}{2(\alpha_0+1)}$$
$$= \frac{\alpha_0 - \alpha}{\alpha_0 + 1}.$$

In the case $\alpha = \infty$, observe that Q inherits the property $Q \circ A \geq 0$ as $v \circ f \geq 0$. The fact that $v \circ f \geq 0$ together with $\|v\|_1 = 1$ gives $\langle v, f\rangle = 1$, which in turn implies $\langle A, Q\rangle = 1$.

Upper bound on $\mu^*(\mathbf{Q})$ We use the Fourier representation of v, namely

$$v[x] = \sum_{T\subseteq[m]} \hat{v}(T)\chi_T(x),$$

which naturally induces a representation of Q as:

$$Q = \frac{1}{c}\sum_{T\subseteq[m]} \hat{v}(T)A_T,$$

where A_T is the (k, m, M, χ_T) pattern tensor.

By Claim 11, to bound $\mu^*(Q)$ it suffices to bound $\mu^*(Q \bullet_x Q) \leq \|Q \bullet_x Q\|_1$. Now we have

$$\|Q \bullet_x Q\|_1 \leq \frac{1}{c^{2^{k-1}}} \sum_{\substack{T_\ell\subseteq[m]\\ \ell\in\{0,1\}^{k-1}}} \sum_{S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}} \left|\sum_x \prod_{\ell\in\{0,1\}^{k-1}} \hat{v}(T_\ell)\chi_{T_\ell}(x|_{S_{\ell_1}^1,\ldots,S_{\ell_{k-1}}^{k-1}})\right|$$
$$\leq \frac{1}{2^{m2^{k-1}}c^{2^{k-1}}} \sum_{\substack{T_\ell\\ |T_\ell|\geq d}} \sum_{S_0^1,S_1^1,\ldots,S_0^{k-1},S_1^{k-1}} \left|\sum_x \prod_{\ell\in\{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S_{\ell_1}^1,\ldots,S_{\ell_{k-1}}^{k-1}})\right|$$

as $\hat{v}(T_\ell) = 1/2^m\langle v, \chi_{T_\ell}\rangle \leq 1/2^m$, and $\hat{v}(T_\ell) = 0$ whenever $|T_\ell| < d$.

Fortunately, for many values of $S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}$ the quantity within absolute values is simply 0. The next claim upper bounds the probability that it is non-zero.

Claim 20 *Fix sets $T_\ell \subseteq [m]$ for $\ell \in \{0,1\}^{k-1}$.*

$$\Pr_{S_0^1, S_1^1, \ldots, S_0^{k-1}, S_1^{k-1}} \left[\sum_x \prod_{\ell \in \{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}}}) \neq 0 \right] \leq \left(\frac{(k-1)m}{M} \right)^{|\cup_\ell T_\ell|}$$

We continue in the main line of proof and delay the proof of this claim to the end. Applying Claim 20 together with the observation that, when the inner sum is non-zero, it has magnitude $2^{m(M/m)^{k-1}}$, we find

$$\|Q \bullet_x Q\|_1 \leq \frac{2^{m(M/m)^{k-1}} \left(\frac{M}{m}\right)^{2(k-1)m}}{2^{m2^{k-1}} c^{2^{k-1}}} \sum_{\substack{T_\ell \\ |T_\ell| \geq d}} \left(\frac{(k-1)m}{M} \right)^{|\cup_\ell T_\ell|}$$

Applying Claim 11, and the fact that $\mu^*(Q \bullet_x Q) \leq \|Q \bullet_x Q\|_1$, we get

$$\mu^*(Q)^{2^{k-1}} \leq \sum_{\substack{T_\ell \\ |T_\ell| \geq d}} \left(\frac{(k-1)m}{M} \right)^{|\cup_\ell T_\ell|}$$

We now quantify this sum over the cardinality of $\cup_\ell T_\ell$. As each T_ℓ is of cardinality at least d, $|\cup_\ell T_\ell| \geq d$ as well. As there are fewer than $2^{r2^{k-1}}\binom{m}{r}$ many collections $\{T_\ell\}_{\ell \in \{0,1\}^{k-1}}$ with $|\cup_\ell T_\ell| = r$, we obtain:

$$\mu^*(Q)^{2^{k-1}} \leq \sum_{r=d}^{m} 2^{r2^{k-1}} \binom{m}{r} \left(\frac{(k-1)m}{M} \right)^r.$$

Finally, using $\binom{m}{r} \leq (em/r)^r$ we find

$$\begin{aligned} \mu^*(Q)^{2^{k-1}} &\leq \sum_{r=d}^{m} \left(\frac{(k-1)2^{2^{k-1}} em^2}{rM} \right)^r \\ &\leq 2^{-d+1}, \end{aligned}$$

for $d > 1$ and $M \geq e(k-1)2^{2^{k-1}} m^2$.

□

We now turn to the proof of Claim 20.

Proof:[of Claim 20] We first develop a simple necessary condition for the sum in question to be non-zero.

Claim 21 *For $\ell \in \{0,1\}^{k-1}$, fix sets $T_\ell \subseteq [m]$, and sets $S^i_{\ell_i} \subseteq [M/m]^m$ where $i = 1, \ldots, k-1$. Then*

$$\sum_x \prod_{\ell \in \{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}}}) \neq 0$$

only if for all $t \in \cup_\ell T_\ell$ there exists $j \in \{1, \ldots, k-1\}$ such that $S^j_0[t] = S^j_1[t]$.

Proof: We show the contrapositive. Assume without loss of generality that there is a $t \in T_{0\ldots0}$ such that for all $j : S^j_0[t] \neq S^j_1[t]$. Then we claim that

$$\sum_x \prod_{\ell \in \{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}}}) = 0.$$

Indeed, if we denote by W the index $(t, S^1_0[t], \ldots, S^{k-1}_0[t])$ of x, we find the above sum is equal to

$$\begin{aligned} \textstyle\sum_{x:x[W]=1} \quad & \chi_{T_{0\ldots0}}(x|_{S^1_0, \ldots, S^{k-1}_0}) \prod_{\ell \in \{0,1\}^{k-1} - (0\ldots0)} \chi_{T_\ell}(x|_{S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}}}) + \\ \textstyle\sum_{x:x[W]=0} \quad & \chi_{T_{0\ldots0}}(x|_{S^1_0, \ldots, S^{k-1}_0}) \prod_{\ell \in \{0,1\}^{k-1} - (0\ldots0)} \chi_{T_\ell}(x|_{S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}}}) \\ = \quad & 0. \end{aligned}$$

If x' is x with the bit in position W flipped, then $\chi_{T_{0\ldots0}}(x|_{S^1_0, \ldots, S^{k-1}_0}) = -\chi_{T_{0\ldots0}}(x'|_{S^1_0, \ldots, S^{k-1}_0})$. The fact that $S^j_0[t] \neq S^j_1[t]$ for all j guarantees that W does not appear in the restriction of x to $(S^1_{\ell_1}, \ldots, S^{k-1}_{\ell_{k-1}})$ for $\ell \neq (0 \ldots 0)$, implying the cancellation of the above two terms. □

Now that we know the sum can be nonzero only if for all $t \in \cup T_\ell$, there exists j such that $S^i_0[t] = S^i_1[t]$, all that remains is to calculate the probability this happens. For a fixed $t \in \cup T_\ell$, we have

$$\Pr_{S^i_{\ell_i}[t] \subseteq [M/m]} \left[\exists i : S^i_0[t] = S^i_1[t]\right] \leq \frac{(k-1)m}{M}$$

by a union bound. It then follows that

$$\Pr_{S^i_{\ell_i} \subseteq [M/m]^m} \left[\forall t \in \cup T_\ell \; \exists i : S^i_0[t] = S^i_1[t]\right] \leq \left(\frac{(k-1)m}{M}\right)^{|\cup T_\ell|},$$

as each $S^i_{\ell_i}[t]$ is chosen independently. □

6 Applications

6.1 Symmetric functions

In this section, we apply Theorem 19 to prove lower bounds on the k-party number-on-the-forehead randomized communication complexity of all symmetric functions. A function $f_n : \{0,1\}^n \to \{-1, 1\}$ is called symmetric if $f_n(x) = g_n(|x|)$ for some function $g_n : [n] \to \{-1, 1\}$.

For a function $f_n : \{0,1\}^n \to \{-1,1\}$ we denote by $F_{k,n,f}$ the function $F_{k,n,f} : (\{0,1\}^n)^k \to \{-1,1\}$ defined by $F_{k,n,f}(x_1,\ldots,x_k) = f(x_1 \wedge x_2 \ldots \wedge x_k)$. In particular, we have $\mathrm{DISJ}_{k,n} = F_{k,n,\mathrm{OR}}$.

Our main result on pattern tensors allows us to say the following about functions $F_{k,n,f}$.

Theorem 22 *Fix $0 \leq \epsilon < 1/2$, and let $\alpha_0 > 1/(1-2\epsilon)$. Set $c_k = e(k-1)2^{2^{k-1}}$. For any symmetric function $f_n : \{0,1\}^n \to \{-1,1\}$*

$$R^k_\epsilon(F_{k,n,f}) \geq \deg_{\alpha_0}(f_m)/2^{k-1} - O(1),$$

for $m = \lfloor (n/c_k^{k-1})^{1/k} \rfloor$.

Proof: Take $m = \lfloor (n/c_k^{k-1})^{1/k} \rfloor$ and $M = \lfloor c_k m^2 \rfloor$, and $n' = m(M/m)^{k-1}$. It is easy to check that $n \geq n'$.

We show that the (k,m,M,f) pattern tensor, $A_{k,m,M,f}$, is a sub-tensor of $F_{k,n',f}$, i.e. that there is a reduction from the problem of computing $A_{k,m,M,f}$ to the problem of computing $F_{k,n',f}$. For purposes of presentation, we consider x and each y_i in the vector of inputs $(y_1,\ldots,y_k)$ to $F_{k,n',f}$ as tensors of dimension $(m, M/m, \ldots, M/m)$.

The reduction is as follows: The **inputs** $(x, S^1, \ldots, S^{k-1})$ to $A_{k,m,M,f}$ are **mapped** to inputs $(x, y_1, \ldots, y_{k-1})$, respectively. The input x is mapped to itself. For each $j \in \{1,\ldots,k-1\}$, we let $y_j[t, I_1, \ldots, I_{k-1}] = 1$ if $I_j = S^j[t]$ and 0 otherwise.

To see that this is indeed a reduction, observe that

$$\begin{aligned}
F_{k,n',f}(x, y_1, \ldots, y_{k-1}) &= f_{n'}(x \wedge (y_1 \wedge y_2 \ldots \wedge y_{k-1})) \\
&= g_{n'}(|x \wedge (y_1 \wedge y_2 \ldots \wedge y_{k-1})|) \\
&= g_m(|x|_{S^1,\ldots,S^{k-1}}|) \\
&= f_m(x|_{S^1,\ldots,S^{k-1}}) \\
&= A_{k,m,M,f}(x, S^1, \ldots, S^{k-1}).
\end{aligned}$$

The third equality follows from the fact that the vector $y_1 \wedge y_2 \ldots \wedge y_{k-1}$ is equal to 1 in coordinate $(t, I_1, \ldots, I_{k-1})$ if and only if $(I_1 = S^1[t]) \wedge (I_2 = S^2[t]) \ldots \wedge (I_{k-1} = S^{k-1}[t])$. Hence, the coordinates that are taken in x when restricting x to $S^1, \ldots, S^{k-1}$ are exactly the coordinates in which the vector $y_1 \wedge y_2 \ldots \wedge y_{k-1}$ is equal to 1. The rest of the steps follow directly from the definitions.

Therefore, taking $\alpha_0 > \alpha > 1/(1-2\epsilon)$ we have

$$\log \mu^\alpha(F_{k,n',f}) \geq \log \mu^\alpha(A_{k,m,M,f}) \geq \deg_{\alpha_0}(f_m)/2^{k-1} - O(1),$$

where the last inequality follows from Theorem 19.

Finally there is a natural reduction from $F_{k,n',f}$ to $F_{k,n,f}$ for $n \geq n'$, which simply restricts some of the coordinates in the input to zero. Thus

$$\log \mu^\alpha(F_{k,n,f}) \geq \log \mu^\alpha(F_{k,n',f}).$$

The application to randomized communication complexity follows from Theorem 6. □

We can instantiate this theorem using a result of Paturi which gives asymptotically optimal bounds on the approximate degree of all symmetric functions. We need the following definition.

Definition 23 *Let* $g_n : [n] \to \{-1, 1\}$. *Define*

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

to be the smallest integers such that g_n *is constant in the interval* $[\ell_0(g_n), n - \ell_1(g_n)]$. *For a symmetric function* $f(x) = g_n(|x|)$ *let* $\ell_0(f) = \ell_0(g_n)$ *and similarly* $\ell_1(f) = \ell_1(g_n)$.

Theorem 24 (Paturi) *Let* $f : \{0, 1\}^n \to \{-1, 1\}$ *be a symmetric function. Then*

$$\deg_3(f) = \Theta\left(\sqrt{n(\ell_0(f) + \ell_1(f))}\right).$$

Using this characterization of approximate degree, and Theorem 22, we get the following simple lower bound.

Corollary 25 *Set* $c_k = e(k-1)2^{2^{k-1}}$. *Let* $f_n(x) = g_n(|x|)$ *be a symmetric function. Then*

$$R_{1/4}(F_{n,k,f}) = \Omega\left(\frac{\sqrt{m(\ell_0(f_m) + \ell_1(f_m))}}{2^{k-1}}\right)$$

where $m = \lfloor (n/c_k^{k-1})^{1/k} \rfloor$. *In particular,*

$$R_{1/4}(\mathrm{DISJ}_{k,n}) = \Omega\left(\frac{n^{1/2k}}{(k-1)2^{k-1}2^{2^{k-1}}}\right)$$

6.2 Proof systems

In this section we formally define the proof systems discussed in the introduction, and the lower bounds which follow from our results on disjointness.

A k-threshold formula is a formula of the form $\sum_j \gamma_j m_j \geq t$, where t, γ_j are integers, and each m_j is a monomial over variables $x_1, \ldots, x_n$. The size of a k-threshold formula is the sum of the sizes of γ_j and t, written in binary. For k-threshold formulas f_1, f_2, g, we say that g is *semantically entailed* by f_1 and f_2 if every 0/1 assignment to $x_1, \ldots, x_n$ that satisfies both f_1 and f_2 also satisfies g.

Let ϕ be an unsatisfiable CNF formula with variables $x_1, \ldots, x_n$. For each clause of ϕ we create a linear threshold formula which is satisfied if and only if the clause is. We refer to these clauses as *axioms*. We say that $\mathcal{P}$ is a $\mathrm{Th}(k)$ refutation of ϕ if

- $\mathcal{P}$ is a sequence $L_1, \ldots, L_t$ of k-threshold formulas.
- Each formula L_j is either an axiom or is semantically entailed by formulas $L_i, L_{i'}$ with $i, i' < j$.
- The final formula L_t is $0 \geq 1$.

The size of $\mathcal{P}$ is the sum of the sizes of $L_1, \ldots, L_t$. We say that $\mathcal{P}$ is *tree-like* if the underlying directed acyclic graph representing the implication structure of the proof is a tree.

We are now ready to state the connection of [BPS06] between the number-on-the-forehead complexity of disjointness and the size of $\mathrm{Th}(k)$ proofs.

Theorem 26 (Beame, Pitassi, and Segerlind [BPS06]) *Let $k \geq 2$. For every n there is a CNF formula ϕ over n variables such that the size of any $\mathrm{Th}(k-1)$ refutation of ϕ is at least*

$$\exp\left(\Omega\left(\left(\frac{R^k_{1/4}(\mathrm{DISJ}_{k,n^{1/7}})}{\log n}\right)^{1/3}\right)\right).$$

Substituting the bounds from our Corollary 25 we obtain the following Corollary:

Corollary 27 *Let $k \geq 2$. For every n there is a CNF formula ϕ over n variables which requires $\mathrm{Th}(k-1)$ refutation proofs of size*

$$\exp\left(\Omega\left(\frac{n^{1/52k}}{\left(\log(n)(k-1)2^{k-1}2^{2^{k-1}}\right)^{1/3}}\right)\right).$$

In particular, for any $k = \log\log n - O(\log\log\log n)$ there is a CNF formula which requires superpolynomial size $\mathrm{Th}(k)$ refutations.

Acknowledgments

We thank Robert Špalek for helpful comments on an earlier version of the paper, and Nate Segerlind for answering our many questions about proof systems.

References

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and Logspace-hard pseudorandom sequences. In *Proceedings of the 21st ACM Symposium on the Theory of Computing*, pages 1–11. ACM, 1989.

[BPS06] P. Beame, T. Pitassi, and N. Segerlind. Lower bounds for Lovász-Schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2006.

[BPSW06] P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product lemma for corruption and the NOF complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

[CFL83] A. Chandra, M. Furst, and R. Lipton. Multi-party protocols. In *Proceedings of the 15th ACM Symposium on the Theory of Computing*, pages 94–99. ACM, 1983.

[Cha07] A. Chattopadhyay. Discrepancy and the power of bottom fan-in depth-three circuits. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 449–458. IEEE, 2007.

[Cha07b] A. Chattopadhyay. Personal communication.

[FG05] J. Ford and A. Gál. Hadamard tensors and lower bounds on multiparty communication complexity. In *Proceedings of the 32th International Colloquium On Automata, Languages and Programming*, pages 1163–1175, 2005.

[Gro94] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and computation*, 112(1):51–54, 1994.

[GW95] M. Goemans and D. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42:1115–1145, 1995.

[IK06] D. Itsykson and A. Kojevnikov. Lower bounds of static Lovász-Schrijver calculus proofs for Tseitin tautologies. *Zapiski Nauchnyh Seminarov POMI*, 340:10–32, 2006.

[KN97] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[KS87] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. In *Proceedings of the 2nd Annual Conference on Structure in Complexity Theory*, pages 41–49, 1987.

[LS91] L. Lovász and A. Schrijver. Cones of matrices and set-functions, and 0-1 optimization. *SIAM Journal of Optimization*, 1:1–17, 1991.

[LS07] N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*. ACM, 2007.

[Raz92] A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106:385–390, 1992.

[Raz03] A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145–159, 2003.

[She07a] A. Sherstov. Separating AC^0 from depth-2 majority circuits. In *Proceedings of the 39th ACM Symposium on the Theory of Computing*. ACM, 2007.

[She07b] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. Technical report, ECCC TR07-100, 2007.

[Tes02] P. Tesson. *Communication complexity questions related to finite monoids and semigroups*. PhD thesis, McGill University, 2002.

[VW07a] E. Viola and A. Wigderson. Norms, XOR lemmas, and lower bounds for GF(2) polynomials and multiparty protocols. In *Proceedings of the 22nd IEEE Conference on Computational Complexity*. IEEE, 2007.

[VW07b] E. Viola and A. Wigderson. One-way multi-party communication lower bound for pointer jumping with applications. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*. IEEE, 2007.

[Yao79] A. Yao. On some complexity questions in distributive computing. In *Proceedings of the 11th ACM Symposium on the Theory of Computing*, pages 209–213. ACM, 1979.