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 $\Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$ 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 a broad class of tree-like proof systems 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 is limited to bounds of size $O(\log n)$ for disjointness.

To analyze the bound given by our new technique for the disjointness function, we build on an elegant framework developed by Sherstov in the two-party case and Chattopadhyay in the multi-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.

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

†Email: adi.shraibman@weizmann.ac.il

1 Introduction

Since its introduction [1, 24], communication complexity has become a key concept in complexity theory and theoretical computer science in general. Part of its appeal is that it has applications to many different computational models, for example to formula size and circuit depth, proof complexity, branching programs, VLSI design, and time-space trade-offs for Turing machines (see [16] 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 [6]. 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, by results of [25, 13], showing a super-polylogarithmic lower bound on an explicit function for super-polylogarithmic many players would give an explicit function which requires super-polynomial size ACC^0 circuits.

While showing such bounds remains a challenging open problem, we do know of explicit functions which require large communication in this model for $\Omega(\log n)$ many players. Babai, Nisan, and Szegedy [3] 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 some 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 $O(kn/2^k)$ [12]—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

1093-0159/08 $25.00 © 2008 IEEE
DOI 10.1109/CCC.2008.29

lower bound in the general number-on-the-forehead model is $\Omega(\log n/(k-1))$ [5, 23]. A major obstacle toward proving better lower bounds on disjointness is that it has a very simple co-nondeterministic protocol—if the strings do intersect, a $\log n$ size proof consists of the index where all players have a one. One of the most common means of showing lower bounds on randomized complexity, the discrepancy method, also lower bounds (co)-nondeterministic complexity, and thus is limited to logarithmic lower bounds for disjointness. Even in the two-party case, determining the randomized and quantum complexity of disjointness was a long-standing open problem which required the development of novel techniques to resolve [15, 19, 20].

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 [16] say, "The only technique from two-party communication complexity that generalizes to the multiparty case is the discrepancy method."

Besides this technical challenge, additional motivation was given to studying the number-on-the-forehead complexity of disjointness by Beame, Pitassi, and Segerlind, who show that lower bounds on disjointness imply lower bounds on a very general class of proof systems which includes cutting planes and Lovasz-Schrijver proof systems.

We show that disjointness requires randomized communication $\Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right)$ in the general k-party number-on-the-forehead model. To do this, we develop a cylinder intersection norm $\mu(A)$ which measures how efficiently a tensor A can be decomposed as a sum of cylinder intersections. As a two-party protocol decomposes the communication matrix into monochromatic rectangles, in the multiparty number-on-the-forehead case we have the analogous structural theorem that a successful c bit number-on-the-forehead protocol decomposes the communication tensor into 2^c many monochromatic cylinder intersections. The μ norm lower bounds how efficiently A can be so decomposed.

For randomized complexity, the analogous structural theorem says not that A itself can be decomposed, but a matrix *near* A. To handle this, we look at an approximate version of the cylinder intersection norm μ^α where $1 \leq \alpha < \infty$ represents the measure of approximation. The limiting case $\mu^\infty(A)$ is exactly the usual discrepancy method; for bounded α we obtain a technique which is strictly stronger than the discrepancy method.

Disjointness has a $O(\log n)$ co-nondeterministic protocol—the proof simply being the name of an index where all players share a one, thus our results show a separation between deterministic and non-deterministic complexity for up to $k = \log\log n - O(\log\log\log n)$ many players. Building on our work, David and Pitassi [9] are able to separate deterministic and non-deterministic complexity for up to $k = \log n$ players using a non-explicit function f.

Chattopadhyay and Ada [8] independently obtained similar bounds on disjointness with similar techniques. They essentially use the dual version of the μ^α norm as we have defined it here, which they call generalized discrepancy.

1.1 Comparison with previous work

For restricted models of computation, bounds are known which are stronger than ours. Wigderson showed that for one-way three-party number-on-the-forehead protocols, disjointness requires communication $\Omega(n^{1/2})$ (this result appears in [2]). Recently, Viola and Wigderson extended this approach to show a bound of $\Omega(n^{1/(k-1)}/k^{O(k)})$ for one-way k-party protocols. These results actually show bounds on a pointer jumping function, which reduces to disjointness.

Beame, Pitassi, Segerlind, and Wigderson [5] devised a method based on a direct product theorem to show a $\Omega(n^{1/3})$ bound on the complexity of three-party disjointness in a model stronger than one-way where the first player speaks once, and then the two remaining players interact arbitrarily.

Our techniques have little to do with the above mentioned results, but rather build on another line of work. Our cylinder intersection norm works very well in conjunction with an elegant framework that evolved in a series of works [21, 22, 7] for proving lower bounds on communication complexity.

The two main ingredients in this framework are pattern tensors and duality of normed spaces. For a symmetric function f, a pattern tensor for f is a structured sub-tensor of $f(x_1 \wedge \cdots \wedge x_k)$. The structure of this tensor allows one to relate properties of the underlying function, such as its degree as a polynomial, to complexity measures of the related pattern tensor. For example, in the case of disjointness the underlying function is OR.

Sherstov first looked at pattern matrices in [21] and used them to relate discrepancy and sign degree. Chattopadhyay substantially generalized the pattern matrix framework to tensors and related sign degree to discrepancy of pattern tensors.

Sherstov [22] then showed a relation between approximate degree and approximate trace norm of a pattern matrix. This simplified a proof of Razborov [20] who first introduced the approximate trace norm technique to show tight lower bounds on the quantum communication complexity of symmetric functions.

For k-tensors with $k \geq 3$ there is no equivalent definition of the trace norm, this is where the cylinder intersection norm and its approximated variants, μ^α, come into play. In

two dimensions in fact μ^α can be seen as a generalization of the approximate trace norm technique.

The way we define approximation is slightly different than the above mentioned results. A nice advantage of our approach is that we can unify all of the above results by relating the α-approximate degree to μ^α. For $\alpha = \infty$ this is the relation between sign degree and discrepancy.

1.2 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 [4] 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 [18]. The seminal 0.878-approximation algorithm for MAXCUT of Goemans and Williamson [11] 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 [4], 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 [14].

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 methods. Beame, Pitassi, and Segerlind [4] 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 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 or is a k-tensor. For a k-tensor A of dimensions $(n_1, \ldots, n_k)$ we say $\mathrm{size}(A) = n_1 \cdots n_k$. A tensor for which all entries are in $\{-1, 1\}$ we call a sign tensor. For a 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 $R^k_\epsilon(A)$. We drop the superscript when the number of players is clear from context.

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 $\hat{f}(S) = (1/2^n)\langle f, \chi_S\rangle$. The degree of f is the size of the largest set S for which $\hat{f}(S)$ is non-zero.

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 M is a k-tensor and $\mu_{\pm 1}(M)$ 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 $\mu(M) = \Theta(\gamma_2(M))$ [17].

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

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

□

Remark 7 *It is nice to note that since a non-deterministic protocol induces a covering of the tensor with cylinder intersections, it follows that $\log \mu^{\infty}$ is a lower bound on non-deterministic communication complexity.*

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

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

When $\alpha = \infty$ we have

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

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

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

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 8 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 9 *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 10

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

Proof: By Theorem 8, 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 $\|A\|^2 \leq \|AA^T\|$. We will try to do a similar thing in upper bounding $\mu^*(Q)$. In analogy with AA^T we make the following definition:

Definition 11 (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*

$$A \bullet_1 A[x_2, x_2', \ldots, x_k, x_k'] = \mathbb{E}_{x_1} \left[\prod_{y_2 \in \{x_2, x_2'\}, \ldots, y_k \in \{x_k, x_k'\}} A[x_1, y_2, \ldots, y_k] \right]$$

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

Notice that when A is a m-by-n matrix $A \bullet_1 A$ corresponds to $(1/m)AA^T$. In analogy with the fact that $\|A\|^2 \leq m \|A \bullet_1 A\|$, the next lemma gives a corresponding statement for the μ^* norm and k-tensors. This lemma originated in the work of Babai, Nisan, and Szegedy [3] and all lower bounds on randomized number-on-the-forehead complexity use some version of this lemma.

Lemma 12 *Let A be a k-tensor. Then*

$$\left(\frac{\mu^*(A)}{\mathrm{size}(A)} \right)^{2^{k-1}} \leq \frac{\mu^*(A \bullet_1 A)}{\mathrm{size}(A \bullet_1 A)} \leq \mathbb{E}[|A \bullet_1 A|]$$

Proof: The second inequality follows since $\mu^*(X) \leq \|X\|_1$ for any real matrix X. The first inequality is standard, and follows by applying the Cauchy-Schwarz inequality repeatedly $k-1$ times. □

4.1 Example: Hadamard tensors

We give an example to show how Lemma 12 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 Lemma 12 which gives:

$$\mu^*(H)^2 \leq N^4 \, \mathbb{E}[|H \bullet_1 H|] = N^3$$

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

Ford and Gál [10] extend the notion of matrix orthogonality to tensors, defining what they call Hadamard tensors.

Definition 13 (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 14 (Ford and Gál [10]) *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 Lemma 12. This gives us

$$\mu^*(H)^{2^{k-1}} \leq N^{k2^{k-1}} \, \mathbb{E}[|H \bullet_1 H|]$$

The "Hadamard" property of H lets us easily upper bound $\mathbb{E}[|H \bullet_1 H|]$. Note that each entry of $H \bullet_1 H$ is of magnitude at most one, and the probability of a non-zero entry is at most

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

by a union bound. Hence, we obtain

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

Putting everything together, we have

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

□

Remark 15 *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 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 $\mu^\alpha(A_f)$, where A_f is a pattern tensor.

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 16 *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 17 *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 18

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

Proof: Follows from duality theory of linear programming. □

Corollary 19 (cf. Sherstov Corollary 3.3.1 [22]) *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 $\leq d$.*
2. *$\|v\|_1 = 1$.*
3. *$\langle v, f\rangle \geq \frac{\alpha-1}{\alpha+1}$.*

Furthermore, 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 define a natural generalization of the pattern matrices of Sherstov [22] to the tensor case. Note that we need a slightly different definition of pattern tensor than used by Chattopahyay [7] to allow the reduction to disjointness.

A pattern k-tensor is described by natural numbers m, M and a function $\phi : \{0,1\}^m \to \mathbb{R}$. Let $x = (x_1, \ldots, x_m)$ where each x_i is a tensor of rank $k-1$ with side length M. Let $S^i \in [M]^m$ for $i = 1, \ldots, k-1$ be ordered sets. We will let $S^i[t] \in [M]$ refer to the t^{th} element of S^i, which can be thought of as a pointer into the i^{th} dimension of x_t. The sets $\bar{S} = (S^1, \ldots, S^k)$ select an m bit string from x as follows:

$$x|_{\bar{S}} = x_1\left[S^1[1], \ldots, S^{k-1}[1]\right], \ldots, x_m\left[S^1[m], \ldots, S^{k-1}[m]\right].$$

We then define the (k, m, M, ϕ) pattern tensor, denoted $A_{k,m,M,\phi}$, as

$$A_{k,m,M,\phi}[x, S^1, \ldots, S^{k-1}] = \phi(x|_{\bar{S}})$$

Now we are ready to state our main theorem.

Theorem 20 *For non-negative integers k, m 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 $1 \leq \alpha < \alpha_0 < \infty$, provided $M \geq 2e(k-1)2^{2^{k-1}} m/\deg_{\alpha_0}(f)$. Furthermore,

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

provided $M \geq 2e(k-1)2^{2^{k-1}} m/\deg_\infty(f)$

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 q be the vector from Corollary 19 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, q) pattern tensor, where $c = \text{size}(Q)/2^m$. 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 $\langle q, f\rangle \geq (\alpha_0 - 1)/(\alpha_0 + 1)$, 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:

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

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

Upper bound on $\mu^*(\mathbf{Q})$ We will bound $\mathbb{E}[|Q \bullet_1 Q|]$ and apply Lemma 11 to obtain an upper bound on $\mu^*(Q)$. Using the Fourier decomposition of q, we obtain a decomposition of Q as

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

where A_T is the pattern tensor with function χ_T.

Now we have

$$\text{size}(Q)^{2^{k-1}}\mathbb{E}[|Q \bullet_1 Q|] =$$
$$\frac{\text{size}(Q)^{2^{k-1}}}{c^{2^{k-1}}}\mathbb{E}_{\bar{S}_0,\bar{S}_1}\left[\left|\mathbb{E}_x\left[\prod_{\ell=0}^{2^{k-1}-1}\sum_{T\subseteq[m]} \hat{q}(T)\chi_T(x|_{\bar{S}_\ell})\right]\right|\right]$$
$$\leq \mathbb{E}_{\bar{S}_0,\bar{S}_1}\left[\sum_{\substack{T_0,\ldots,T_{2^{k-1}}\\ |T_i|>d}}\left|\mathbb{E}_x\left[\prod_{\ell\in\{0,1\}^{k-1}}\chi_{T_\ell}(x|_{\bar{S}_\ell})\right]\right|\right]$$

Here we have used the fact that $\hat{q}(T_\ell) = 1/2^m\langle q, \chi_{T_\ell}\rangle \leq 1/2^m$, and that $\hat{q}(T_\ell) = 0$ whenever $|T_\ell| \leq d = \deg_{\alpha_0}(f)$ by Corollary 19.

Now fix sets $\bar{S}_0, \bar{S}_1$. We will count for how many sets $\{T_\ell\}$ the expectation over x is zero. Consider first a simpler question, to evaluate

$$\mathbb{E}_{x\in\{0,1\}^m}\left[\prod_i \chi_{T_i}(x)\right].$$

A moments reflection shows that this will be nonzero if and only if $\sum T_i = 0 \mod 2$, where we think of T_i as being the characteristic vector of the corresponding set. The next lemma gives the generalization of this where the argument to χ_{T_i} is not x, but $x|_{S_\ell}$ a selection of m-bits from a longer string x. Analogously to the simple case, the expectation now will be nonzero if and only if $\sum S_i|_{T_i} = 0 \mod 2$. By this notation, we think of $S_\ell = (X_\ell^1, \ldots, X_\ell^m)$ as a vector of m many $0/1$ valued tensors where each X_ℓ^i has exactly one nonzero entry—namely the i^{th} position selected by S_ℓ. By $S_\ell|_{T_\ell}$ we zero out those tensors X_ℓ^i where $i \notin T_\ell$.

Lemma 21 *Fix $\bar{S}_0, \bar{S}_1$ and $\{T_\ell\}_{\ell \in \{0,1\}^{k-1}}$.*

$$\mathbb{E}_x \left[\prod_{\ell \in \{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S_\ell}) \right] = \delta \left(\sum_\ell S_\ell|_{T_\ell}, 0 \right),$$

where δ is the kronecker delta function. Notice that this implies in particular that the expectation is zero unless $\sum_\ell T_\ell = 0$.

We continue in the main line of the proof and delay the proof of this lemma to the end. Call a k-cube degenerate if it contains less than 2^k many points. Let g be the number of degenerate cubes selected by $\bar{S}_0, \bar{S}_1$. That is, the number of $t \in [m]$ such that $S_0^i[t] = S_1^i[t]$ for some $i \in [k-1]$. Clearly, by Lemma 21, the expectation will be zero if any of the sets T_ℓ contain an element t for which $\bar{S}_0, \bar{S}_1$ select a nondegenerate cube. Thus the number of sets $T_0, \ldots, T_{2^{k-1}-1}$ which lead to a nonzero expectation is at most

$$\left(\sum_{r=d}^{g} \binom{g}{r} \right)^{2^{k-1}} \leq 2^{g 2^{k-1}}.$$

Remark 22 *We note without proof that our analysis above is nearly tight, and cannot be improved much without using more information about the Fourier coefficients of q.*

Now we bound the probability that $\bar{S}_0, \bar{S}_1$ have g many degenerate cubes. The probability that $S_0^i[t] = S_1^i[t]$ is $1/M$. Thus by a union bound, the probability that a single cube is degenerate is at most $(k-1)/M$. Finally, as each index is chosen independently, the probability of g many degenerate cubes is at most

$$\binom{m}{g} \left(\frac{k-1}{M} \right)^g$$

Putting everything together we have

$$\begin{aligned}
\mathbb{E}[|Q \bullet_1 Q|] &\leq \frac{1}{\text{size}(Q)^{2^{k-1}}} \sum_{g=d}^{m} \binom{m}{g} \left(\frac{k-1}{M} \right)^g 2^{g 2^{k-1}} \\
&\leq \frac{1}{\text{size}(Q)^{2^{k-1}}} \sum_{g=d}^{m} \left(\frac{e(k-1)2^{2^{k-1}} m}{dM} \right)^g \\
&\leq \frac{2^{-d+1}}{\text{size}(Q)^{2^{k-1}}}
\end{aligned}$$

provided that $M \geq 2e(k-1)2^{2^{k-1}} m/d$.

□

Proof:[of Lemma 21] Let us now turn to the proof of the lemma. Suppose that $\sum_\ell S_\ell|_{T_\ell} \neq 0$. Then there is some $t \in \cup T_\ell$ such that $\sum_\ell X_\ell^t \neq 0$. Let W be a nonzero entry of this sum, and we assume wlog that $X_0^t[W] = 1$. Now consider

$$\begin{aligned}
&\mathbb{E}_x \left[\prod_{\ell \in \{0,1\}^{k-1}} \chi_{T_\ell}(x|_{S_\ell}) \right] = \\
&\chi_{T_0}(x|_{S_0}) \mathbb{E}_{x:x[W]=1} \left[\prod_{\ell \in \{0,1\}^{k-1}-0} \chi_{T_\ell}(x|_{S_\ell}) \right] + \\
&\chi_{T_0}(x|_{S_0}) \mathbb{E}_{x:x[W]=0} \left[\prod_{\ell \in \{0,1\}^{k-1}-0} \chi_{T_\ell}(x|_{S_\ell}) \right] \\
&= 0.
\end{aligned}$$

If, on the other hand, $\sum_\ell S_\ell|_{T_\ell} = 0$ then for each x the product term will simply be one as each character will be taken to an even power. Thus the expectation will be one. □

6 Applications

6.1 Symmetric functions

In this section, we apply Theorem 20 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 $g_n : \{0,1,\ldots,n\} \to \{-1,1\}$.

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 $\text{DISJ}_{k,n} = -F_{k,n,\text{OR}}$.

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

Theorem 23 *Let $f_n : \{0,1\}^n \to \{-1,1\}$ be any symmetric function. Fix $0 \le \epsilon < 1/2$, and let $\alpha_0 > 1/(1-2\epsilon)$. Let $c_k = 2e(k-1)2^{2^{k-1}}/\deg_{\alpha_0}(f)$, then*

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

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

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

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 $F_{k,n',f}$. Let $y^1, \ldots, y^k$ be inputs to $F_{k,n',f}$.

The reduction is as follows: The inputs $(x, S^1, \ldots, S^{k-1})$ to $A_{k,m,M,f}$ are mapped to inputs $(x, y^1, \ldots, y^{k-1})$ of $F_{k,n',f}$ as follows. The input x is mapped to itself. For each $j \in \{1, \ldots, k-1\}$, we interpret the n' length strings $y^j = (y^j_1, \ldots, y^j_m)$ as consisting of m many rank $k-1$ tensors. We set $y^j_t[I_1, \ldots, I_{k-1}] = 1$ if $I_j = S^j[t]$ and 0 otherwise.

To see that this is indeed a reduction, observe that

$$\begin{aligned} F_{k,n',f}(x, y^1, \ldots, y^{k-1}) &= f_{n'}(x \wedge (y^1 \wedge y^2 \ldots \wedge y^{k-1})) \\ &= g_{n'}(|x \wedge (y^1 \wedge y^2 \ldots \wedge y^{k-1})|) \\ &= g_m(|x|_{\bar{S}}|) \\ &= f_m(x|_{\bar{S}}) \\ &= A_{k,m,M,f}(x, S^1, \ldots, S^{k-1}). \end{aligned}$$

The third equality follows from the fact that the vector $y^1_t \wedge y^2_t \ldots \wedge y^{k-1}_t$ is equal to 1 in coordinate $(I_1, \ldots, I_{k-1})$ 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

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

where the last inequality follows from Theorem 20.

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

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

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 24 *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 25 (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 23, we get the following simple lower bound.

Corollary 26 *Let $f_n(x) = g_n(|x|)$ be a symmetric function and set $c_k = 2e(k-1)2^{2^{k-1}}/\deg_3(f)$. 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,

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

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

[1] H. Abelson. Lower bounds on information transfer in distributed computations. In *Proceedings of the 19th IEEE Symposium on Foundations of Computer Science*, pages 151–158. IEEE, 1978.

[2] L. Babai, T. Hayes, and P. Kimmel. The cost of the missing bit: communication complexity with help. *Combinatorica*, 21:455–488, 2001.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[9] M. David and T. Pitassi. Separating nof communication classes RP and NP. Technical Report TR-08-014, ECCC, 2008.

[10] 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.

[11] 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.

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

[13] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

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

[15] 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.

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

[17] 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.

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

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

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

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

[22] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th ACM Symposium on the Theory of Computing*. ACM, 2008.

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

[24] 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.

[25] A. Yao. On ACC and threshold circuits. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 619–627. IEEE, 1990.