Simplified Lower Bounds on the Multiparty Communication Complexity of Disjointness

Anup Rao*  Amir Yehudayoff†

August 30, 2014

Abstract

We show that the deterministic number-on-forehead communication complexity of set disjointness for $k$ parties on a universe of size $n$ is $\Omega(n/k^4)$. This gives the first lower bound that is linear in $n$, nearly matching Grolmusz’s upper bound of $O(\log^2(n) + k^2n/2^k)$. We also simplify Sherstov’s proof showing an $\Omega(\sqrt{n}/(k2^k))$ lower bound for the randomized communication complexity of set disjointness.

1 Introduction

Given a family of $k$ sets $\mathcal{F} = (X_1, \ldots, X_k)$ over the universe $[n]$, the disjointness function is defined as

$$\text{Disjoint}(\mathcal{F}) = \begin{cases} 1 & \text{if } \bigcap_{i=1}^k X_i = \emptyset, \\ 0 & \text{otherwise}. \end{cases}$$

We study the communication complexity of computing disjointness in the number-on-forehead model [CFL83]. We consider $k$ parties that attempt to compute $\text{Disjoint}(\mathcal{F})$ by exchanging messages about $X_1, \ldots, X_k$, until one of the parties announces the value of $\text{Disjoint}(\mathcal{F})$. The $i$'th party can see all of the inputs except for $X_i$, and can send messages that depend on the inputs she sees and all previous messages. All messages are visible to all parties. The communication complexity is the minimum number of bits that needs to be transmitted to compute $\text{Disjoint}(\mathcal{F})$. In a randomized communication protocol, the parties use shared randomness to pick a deterministic communication protocol, and then run the chosen deterministic protocol. The protocol computes $\text{Disjoint}(\mathcal{F})$ correctly if it outputs $\text{Disjoint}(\mathcal{F})$ with probability at least $2/3$, for every family $\mathcal{F}$. For formal definitions of multiparty communication complexity and its significance, we refer the interested reader to [KN97].

Grolmusz [Gro94] gave a beautiful deterministic protocol showing that $\text{Disjoint}(\mathcal{F})$ can be computed deterministically with communication $O(\log^2(n) + k^2n/2^k)$. Chattopadhyay [Cha08] used

---

*Computer Science and Engineering, University of Washington, anuprao@cs.washington.edu. Supported by an Alfred P. Sloan Fellowship, the National Science Foundation under agreement CCF-1016565, an NSF Career award, and by the Binational Science Foundation under agreement 2010089.

†Department of Mathematics, Technion-IIT, Haifa, Israel, amir.yehudayoff@gmail.com. Horev Fellow – supported by the Taub Foundation. Supported by the Israel Science Foundation and by the Binational Science Foundation under agreement 2010089.
similar ideas to give a protocol with communication $O(k \log(n) + n/2^k)$. This paper is about proving lower bounds on the communication complexity.

1.1 Motivation and related work

Lower bounds on multiparty communication complexity are important because several computational models such as circuits, branching programs, and propositional proofs can be used to obtain efficient communication protocols. Strong enough communication complexity lower bounds for the computation of any explicit function can therefore be used to prove lower bounds on these models [BNS92, CT93, BHK01, Raz00, VW08]. In particular, lower bounds on the communication complexity of disjointness have many applications (see the recent survey [CP10]). Such lower bounds imply lower bounds on proof systems [BPS07], circuit lower bounds [HC91, RW93, NW93, Vio07], lower bounds on communication for problems related to combinatorial auctions [CS04, NS06, Nis02, DN11, HM07, PSS08], and oracle separations for complexity classes [AW09].

Attempts to prove lower bounds for disjointness have led to many interesting ideas. When the number of parties is $k = 2$, Kalyanasundaram and Schnitger [KS92] proved that $\Omega(n)$ communication is required in the randomized setting. Alternate proofs and tight bounds have since been obtained [Raz92, BYJKS04, BM13] using methods involving information theory. These methods have found many other applications that we do not discuss here.

When $k$ is large, Tesson [Tes03] and Beame, Pitassi, Segerlind and Wigderson [BPSW06] proved that the deterministic communication complexity is $\Omega(\log(n)/k)$. Then Sherstov [She09, She11] introduced the pattern matrix method for proving lower bounds in the case $k = 2$. The method was used to separate certain circuit classes by relating their complexity to analytic properties of boolean functions, like their approximate degree. This technique was generalized to $k > 2$ by Chattopadhyay [Cha07], Lee and Shraibman [LS09], and Chattopadhyay and Ada [CA08]. These last two papers proved lower bounds of the type $\Omega\left(n^{1/(k+1)}/2^{\Omega(k)}\right)$ on the randomized communication complexity. Beame and Huynh-Ngoc [BHN09] extended these methods further to prove that the randomized communication complexity is at least $2^{\Omega\left(\sqrt{\log(n)/k}\right)} 2^{-k}$. Finally, Sherstov [She12, She13] proved the best known lower bounds prior to our work, showing that the randomized communication complexity is at least $\Omega(\sqrt{n}/(k2^k))$. In fact, Sherstov proved lower bounds for a broader class of functions, as we discuss below.

These results use powerful techniques such as Fourier analysis, Gowers norms, directional derivatives, and bounds on the approximate degree. The last two works of Sherstov are the main inspiration for our work.

1.2 Results

In what follows, $k$ is the number of players in the number-on-forehead model, and $n$ is the size of the universe. Our work follows the ideas in the recent papers of Sherstov [She12, She13]. We prove a linear lower bound on the multiparty communication complexity of disjointness:

**Theorem 1.** The deterministic communication complexity of disjointness is $\Omega\left(\frac{n}{k^2}\right)$.

Given our interpretation of Sherstov’s work in [She12], the proof of Theorem 1 is short. We also simplify the proof of the randomized lower bound from [She13]:

**Theorem 2 ([She13]).** The randomized communication complexity of disjointness is $\Omega\left(\frac{\sqrt{n}}{k2^k}\right)$. 
Sherstov proved lower bounds for functions of the type $f(\text{Disjoint} (\mathcal{F}_1), \ldots, \text{Disjoint} (\mathcal{F}_m))$, where $f$ is a multivariate function, and $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are families on disjoint parts of the universe. In our proof, we focus on the case where $f$ is symmetric. We symmetrize his proof by viewing $f$ as a univariate function $f : \{0, 1, \ldots, m\} \to \{0, 1\}$, rather than as a multivariate function.

The proof begins by bounding the discrepancy of the parity of several independent instances of disjointness. Here we use two different bounds that Sherstov proved [She12, She13], as black boxes. The first bound, stated as Theorem 5 in this paper, is used for the deterministic case, and the second, stated as Theorem 6, is used in the randomized case.

The rest of Sherstov’s proof of Theorem 2 is a method to control the error in an approximation of the function $f$, using the bounds on the discrepancy. In the symmetrized proof, this corresponds to a bound on the error in an approximation of the Kronecker delta function (i.e. the univariate function $f$ that is the indicator of $m$). We bound the error via the following theorem, which shows that every polynomial that is not correlated with any parity has a low-degree approximation:

**Theorem 3.** Let $m$ be a power of 2. For $j \in [m]$, let $J$ denote the smallest power of 2 such that $J \geq j$. Let $Y_1, \ldots, Y_m \in \{0, 1\}$ be distributed uniformly and independently. Suppose $f$ is a real univariate polynomial of degree at most $m$, and $\delta \geq 0$ is such that for every $j \geq d > 0$,

$$ \left| \mathbb{E} \left[ f((Y_1 + \ldots + Y_J)m/J) \cdot (-1)^{Y_1+\ldots+Y_J} \right] \right| \leq 2^{-12Jd} \delta. $$

Then there exists a polynomial $g$ of degree at most $d-1$ such that $|g(x) - f(x)| \leq \delta$ for all $x \in [0, m]$.

To prove Theorem 3, we define a useful basis $b_0(x), \ldots, b_m(x)$ for the space of polynomials, where each $b_i$ is of degree $i$. Given this basis, the polynomial $g$ is just the projection of $f$ to the space spanned by $b_0, \ldots, b_{d-1}$. This basis may be of independent interest. For the analogous part of the proof, Sherstov finds a low-degree approximation of $f$ using a different basis. We prove Theorem 3 in Section 3.

Finally, we state a corollary that may be useful in other applications.

**Corollary 4.** There exists a function $\ell(k) \leq O(k^2 4^k)$ with the following property. Suppose each family $\mathcal{F}_i, i \in [m]$, is supported on a disjoint universe of size $\ell$. Let $f : \{0, 1, \ldots, m\} \to \{0, 1\}$ be an arbitrary function. If the randomized communication complexity of $f(\sum_{i=1}^m \text{Disjoint}(\mathcal{F}_i))$ is $C$, then there exists a polynomial $g$ of degree at most $C - 1$ such that $|f(x) - g(x)| \leq 1/3 + 2^{-3C}$ for all $x \in \{0, 1, \ldots, m\}$.

The proof of the corollary follows easily from the ideas in Section 2.2.

## 2 The lower bounds

Without loss of generality, we assume that $n = mL$, where $m$ is a power of 2 and $L$ is a function of $k$ to be determined. Any family $\mathcal{F} = (X_1, \ldots, X_k)$ can be described using the $m$ families $\mathcal{F}_1, \ldots, \mathcal{F}_m$, each over a universe of size $\ell$, defined as

$$ \mathcal{F}_i = (X_1 \cap [(i-1)\ell + 1, i\ell]), \ldots, X_k \cap [(i-1)\ell + 1, i\ell]). $$

Moreover,

$$ \text{Disjoint}(\mathcal{F}) = \begin{cases} 1 & \text{if } \sum_{i=1}^m \text{Disjoint}(\mathcal{F}_i) = m, \\ 0 & \text{otherwise.} \end{cases} $$

3
Let $S_1, \ldots, S_{k-1} \subseteq [\ell]$ be uniformly random sets conditioned on $|S_1 \cap S_2 \cap \ldots \cap S_{k-1}| = 1$. Let $S_k \subseteq [\ell]$ be uniform and independent. Set

$$F_1 = (S_1, \ldots, S_{k-1}, S_k).$$

### Distribution $\mu$ on $F_1 \subseteq 2^{[\ell]}$

In order to prove Theorems 1 and 2, we consider distributions on families $F$, where each $F_i$ is independent and identically distributed. Sherstov shows that there are distributions of this type under which every protocol with small communication complexity must have low correlation with $(-1)^{\sum_{i=1}^m \text{Disjoint}(F_i)}$.

#### 2.1 The lower bound for deterministic protocols

Consider the distribution $\mu$ given in Figure 1 as a way to sample each $F_i$. The following theorem is an easy consequence of Theorem 4.2 in [She12], and the fact that every communication protocol can be expressed as a sum of cylinder intersections:

**Theorem 5** ([She12]). If each family $F_i$ is sampled independently according to $\mu$, and $\pi$ is a $k$ party protocol with communication complexity $C$, then

$$\left| \mathbb{E}[\pi(F) \cdot (-1)^{\sum_{i=1}^m \text{Disjoint}(F_i)}] \right| \leq 2^C \cdot \left( \frac{2^{k-1} - 1}{\sqrt{\ell}} \right)^m.$$

The proof of Theorem 5 involves ideas analogous to [BNS92] and some subtle reasoning about the distribution $\mu$. For completeness, we give a full exposition of the proof in Appendix B. Given Theorem 5, Theorem 1 easily follows:

**Proof of Theorem 1.** Let $\pi$ be a deterministic protocol that computes $\text{Disjoint}(F)$ with communication complexity $C$. When $\text{Disjoint}(F) = 1$, we have $(-1)^{\sum_{i=1}^m \text{Disjoint}(F_i)} = (-1)^m$ and $\pi(F) = 1$. On the other hand, when $\text{Disjoint}(F) = 0$, we have $\pi(F) = 0$. In addition, $\Pr[\text{Disjoint}(F_i) = 1] = 1/2$ for all $i \in [m]$, which implies $\Pr[\text{Disjoint}(F) = 1] = 2^{-m}$. Thus,

$$\left| \mathbb{E}[\pi(F) \cdot (-1)^{\sum_{i=1}^m \text{Disjoint}(F_i)}] \right| = 2^{-m}. \quad (3)$$

Now set $\ell = 16(2^{k-1} - 1)^2$. Theorem 5 and (3) imply that

$$2^C \cdot (2^{k-1}-1)^m \geq 2^{-m} \Rightarrow C \geq m = \Omega\left(\frac{n}{4^k}\right).$$

$$\square$$
2.2 The lower bound for randomized protocols

The proof of Theorem 1 does not give anything meaningful in the randomized setting, since it may be the case that a randomized protocol has no correlation with \((-1)^{\sum_{i=1}^{m}\text{Disjoint}(F_i)}\). To prove lower bounds on the randomized communication, following Sherstov, we use a more complicated distribution on inputs, as well as approximation theory.

For the rest of this section, we work with the distribution \(\gamma\) described in Figure 2. A crucial feature of this distribution is the following symmetric structure: if \(\rho : [\ell] \rightarrow [\ell]\) is a uniformly random permutation independent of \(F_1\), then the families \((F_1, \rho(F_1))\) have the same joint distribution as two independent samples \((F_1, F'_1)\) from \(\gamma\) conditioned on \(\text{Disjoint}(F_1) = \text{Disjoint}(F'_1)\). Here by \(\rho(F_1)\) we mean the family obtained by permuting the underlying universe. In analogy with Theorem 5, Sherstov shows (Corollary 4.19 in [She13]) that no protocol can be significantly correlated with \((-1)^{\sum_{i=1}^{m}\text{Disjoint}(F_i)}\) under the distribution \(\gamma\):

<table>
<thead>
<tr>
<th>Distribution (\gamma) on (F_1 \subseteq 2^{[\ell]})</th>
</tr>
</thead>
<tbody>
<tr>
<td>Let (t) be such that (\ell = 2^{k-1} + t(2^k - 1)). Let (b \in {0, 1}) be uniformly random. Define (A = A(b) = {(0, v) : v \in {0, 1}^k, v_1 + v_2 + \ldots + v_k = b \mod 2}), and (W = {(r, v) \in [t] \times {0, 1}^k : v \neq 1^k}). Define the (k) sets (S_1, S_2, \ldots, S_k) as (S_j = {(r, v) \in A \cup W : v_j = 1}). Sample a bijection(^a) (\sigma : A \cup W \rightarrow [\ell]) uniformly at random and independently of (b), and set (F_1 = (\sigma(S_1), \sigma(S_2), \sigma(S_3), \ldots, \sigma(S_k))).</td>
</tr>
</tbody>
</table>

\(^a\)Observe \(|A \cup W| = \ell\). |

Figure 2: The distribution \(\gamma\)

**Theorem 6 ([She13]).** If each family \(F_i\) is sampled independently according to \(\gamma\), and \(\pi\) is a protocol with communication complexity \(C\), then

\[
\left| \mathbb{E} \left[ \pi(F) \cdot (-1)^{\sum_{i=1}^{m}\text{Disjoint}(F_i)} \right] \right| \leq 2^C \cdot \left( \frac{c_0 k^2 4^k}{\ell} \right)^{m/4},
\]

where \(c_0 > 0\) is a universal constant.

The proof of Theorem 6 is delicate, mainly due to the symmetric structure of the distribution \(\gamma\) (especially if one wishes to optimize the dependence on \(k\)). This symmetric structure is, on the other hand, very useful, and we shall exploit it next.
Given any protocol $\pi$ computing $\text{Disjoint}(F)$, define $f_{\pi}$ as the unique degree $m$ polynomial so that for all $t \in \{0, 1, \ldots, m\}$,

$$f_{\pi}(t) = \Pr \left[ \pi(F) = 1 \left| \sum_{i=1}^{m} \text{Disjoint}(F_i) = t \right. \right].$$  \hspace{1cm} (4)

---

**Protocol $\tau_{\pi,j}(F_1, \ldots, F_j)$**

1. Let $J$ denote the smallest power of 2 such that $J \geq j$. Note that $m/J$ is an integer, since $m$ is assumed to be a power of 2.

2. Using public randomness, sample $J - j$ families $F_{j+1}, F_{j+2}, \ldots, F_J$ according to $\gamma$, conditioned on the event that $\text{Disjoint}(F_{j+1}) = \text{Disjoint}(F_{j+2}) = \text{Disjoint}(F_J) = 0$.

3. Let $G = (F_1, \ldots, F_1, F_2, \ldots, F_2, \ldots, F_J, \ldots, F_J)$ be the $m$ families obtained by repeating each family $F_i$ exactly $m/J$ times.

4. Let $\rho_1, \rho_2, \ldots, \rho_m : \ell \to \ell, \eta : [m] \to [m]$ be independent uniformly random permutations chosen using public randomness.

5. Output $\pi(\rho_1(G_{\eta(1)}), \rho_2(G_{\eta(2)}), \ldots, \rho_m(G_{\eta(m)}))$.

---

Figure 3: The protocol $\tau_{\pi,j}$

Since the protocol computes $\text{Disjoint}(F)$ with probability at least $2/3$, we have that $|f_{\pi}(t)| \leq 1/3$, for $t = 0, 1, \ldots, m-1$, and $|1 - f_{\pi}(m)| \leq 1/3$. The following well known theorem \cite{EZ64, RC66, NS94} shows that any such function must have degree $\sqrt{m}/3$:

**Theorem 7** (\cite{EZ64, RC66, NS94}). Let $\epsilon \in (0, 1/2)$. If $f : [0, m] \to \mathbb{R}$ is a polynomial such that $|f(t)| \leq \epsilon$ for $t = 0, 1, \ldots, m-1$, and $|1 - f(m)| \leq \epsilon$, then the degree of $f$ is at least $\sqrt{m(1 - 2\epsilon)}/3$.

Theorem 7 is proved via a clever reduction to Markov’s bound on the magnitude of derivatives in bounded polynomials. We include the short proof in Appendix A. We remark that Theorem 7 is tight — one can use Chebyshev polynomials to give a polynomial $f$ of degree $O(\sqrt{m})$ satisfying the constraints.

We shall prove that if the communication of $\pi$ is much less than $\sqrt{n}/(k2^k)$, then Theorems 6 and 3 imply that $f_{\pi}$ can be approximated by a polynomial whose degree is much less than $\sqrt{m}$, contradicting Theorem 7. For each $j \in [m]$, define the protocol $\tau_{\pi,j}$ as in Figure 3. The protocol $\tau_{\pi,j}$ takes $j$ families of sets $F_1, \ldots, F_j$ distributed independently according to $\gamma$ as input. Using shared public randomness, and no communication, the input is used to generate $m$ families $H_1, \ldots, H_m$ independently distributed according to $\gamma$, conditioned on $\sum_i \text{Disjoint}(H_i) = (m/J) \sum_i \text{Disjoint}(F_i)$. This is where the symmetric structure of $\gamma$ is used: the distribution of $H_1, \ldots, H_m$ is as stated due to this structure. Finally, $\tau_{\pi,j}$ simulates $\pi$ on $H_1, \ldots, H_m$. We have:

**Lemma 8.** For every fixed input $F_1, \ldots, F_j$, with $j \leq m$,

$$\Pr[\tau_{\pi,j}(F_1, \ldots, F_j) = 1] = f_{\pi} \left( \sum_{i=1}^{j} \text{Disjoint}(F_i) \right) \frac{m}{J}. $$
Lemma 8 and Theorem 6 together imply that the correlation of \( f_\pi \) with parity is small:

**Lemma 9.** Let \( J \) be the smallest power of 2 so that \( J \geq j \). If the communication complexity of \( \pi \) is \( C \), and \( Y_1, \ldots, Y_j \in \{0,1\} \) are uniformly random and independent, then

\[
\mathbb{E} \left[ f_\pi((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1 + \ldots + Y_j} \right] \leq 2^C \cdot \left( \frac{c_0 k^2 4^k}{\ell} \right)^{j/4},
\]

where \( c_0 > 0 \) is a universal constant.

**Proof.** If \( F_1 \) is distributed according to \( \gamma \), then \( \text{Disjoint}(F_1) \) is a uniformly random bit. Thus,

\[
\mathbb{E} \left[ f_\pi((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1 + \ldots + Y_j} \right] = \mathbb{E} \left[ f_\pi \left( \sum_{i=1}^j \text{Disjoint}(F_i) \right)m/J \cdot (-1)\sum_{i=1}^j \text{Disjoint}(F_i) \right]
\]

Using Lemma 8,

\[
= \mathbb{E} \left[ \tau_{\pi,j}(F_1, \ldots, F_j) \cdot (-1)\sum_{i=1}^j \text{Disjoint}(F_i) \right] \leq 2^C \cdot \left( \frac{c_0 k^2 4^k}{\ell} \right)^{j/4},
\]

where the last inequality is by Theorem 6, since the communication complexity of \( \tau_{\pi,j} \) is equal to that of \( \pi \).

Given Lemma 9, the proof is completed as follows:

**Proof of Theorem 2.** We set \( \ell = 2^{16} c_0 k^2 4^k \), so that the right hand side of Lemma 9 is \( 2^C - 16 j \). Fix any randomized protocol \( \pi \) that computes \( \text{Disjoint}(\mathcal{F}) \) on the distribution induced by \( \gamma \) with \( C \) bits of communication. Let \( f_\pi \) be as defined in (4).

By Lemma 9, \( f_\pi \) satisfies the hypothesis of Theorem 3 with \( d = C \). Thus we conclude that there is a degree \( C - 1 \) polynomial \( g \) that agrees with \( f_\pi \) up to an error of \( 2^{-3C} \). Theorem 7 implies that

\[
C \geq \sqrt{m(1 - 2(1/3 + 2^{-3C}))}/3 = \Omega(\sqrt{n}/(k2^k)).
\]

**3 Approximating functions that are not correlated with parity**

Here we prove Theorem 3, which shows that if a polynomial has low correlation with parity, then it can be approximated by a low degree polynomial. In what follows, let \( Y_1, \ldots, Y_m \in \{0,1\} \) be independent and uniformly random bits, and let \( I, J \) be the smallest powers of 2 such that \( I \geq i \) and \( J \geq j \).

To prove the theorem, we define a useful basis for the space of polynomials. Let \( b_0(x) = 1 \). For \( i > 0 \), let

\[
b_i(x) = 2^i \binom{x I/m}{i} a_i(x) = 2^i x(x - m/I)(x - 2m/I) \ldots (x - (i - 1)m/I).
\]

Since \( b_i \) is of degree \( i \), the polynomials \( b_0, \ldots, b_m \) form a basis for the space of polynomials of degree at most \( m \). To prove Theorem 3, we express \( f \) in this basis and then argue that all coefficients corresponding to high degree terms are negligible. The polynomials in our basis can be bounded by the following lemma:
Lemma 10. For every \( i \in \{0, 1, \ldots, m\} \), \( \max_{x \in [0, m]} |b_i(x)| \leq 8^i \).

Proof. We show that the maximum of \( b_i \) is attained when \( x = m \), and so

\[
\max_{x \in [0, m]} |b_i(x)| = |b_i(m)| = 2^i \left( \frac{mI/m}{i} \right) \leq 2^i \cdot 2^i I \leq 8^i.
\]

Note that the magnitude of \( b_i \) is symmetric around the point \((i-1)m/(2I)\),
\[
|b_i(x + (i-1)m/(2I))| = |b_i(-x + (i-1)m/(2I))|.
\]

So the maximum is attained with \( x \in [(i-1)m/(2I), m] \). For any such \( x \) that is not a root of \( b_i \),
\[
\left| \frac{b_i(x + m/I)}{b_i(x)} \right| = \left| \frac{x + m/I}{x - (i-1)m/I} \right| \geq \left| \frac{x + m/I}{x} \right| > 1,
\]
proving that the maximum is attained with \( x \in [m - m/I, m] \). For such \( x \), every term \((x - jm/I)\)
with \( j \in \{0, 1, \ldots, i-1\} \) in \( b_i(x) \) is non-negative, and so the maximum is attained when \( x = m \).

The basis polynomials behave nicely under the random experiments from \[[1]\]:

Lemma 11. For all \( i \in \{0, 1, 2, \ldots, m\} \) and \( j \in [m] \),

\[
|\mathbb{E} \left[ b_i((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j} \right]| \begin{cases} 
= 0 & \text{if } i < j, \\
= 1 & \text{if } i = j, \\
= 0 & \text{if } j < i \leq J, \\
\leq 8^i & \text{if } J < i.
\end{cases}
\]

Proof. When \( i < j \), the polynomial \( b_i(y_1 + \ldots + y_j) \) has degree \( i \) in the variables \( y_1, \ldots, y_j \). Since
every monomial must exclude one of the \( j \) variables, the contribution of each of the monomials
to the expectation is 0. When \( i = j \), \( b_i((y_1 + \ldots + y_j)m/I) = 2^i(y_1 + \ldots + y_i) \) is non-zero only when
\( y_1 = y_2 = \ldots = y_i = 1 \). Thus the expectation is \( 2^{-1} \cdot 2^i \binom{i}{i} = 1 \) in this case. When \( j < i \leq J \),
we have \( I = J \). Since for \( r \in [i-1] \), \( b_i(\text{rm}/I) = 2^i \binom{i}{i} = 0 \), the expectation is 0. When \( i > J \), by
Lemma \[10\], the expectation is at most \( 8^i \).

Theorem \[3\] now follows by straightforward induction:

Proof of Theorem \[3\]. Write \( f(x) = \sum_{j=0}^{m} a_j b_j(x) \), and let \( g(x) \) be the degree \( d - 1 \) polynomial
\( g(x) = \sum_{j=0}^{d-1} a_j b_j(x) \). To prove the theorem, we show that \( |g(x) - f(x)| \leq \delta \) for all \( x \in [0, m] \). Lemma \[11\] and \[1\] imply that for \( j = d, \ldots, m \),
\[
|a_j| - \sum_{i=J+1}^{m} 8^i |a_i| \leq |\mathbb{E} \left[ f((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j} \right]| \leq 8^{-4J} \delta
\]
\[
\Rightarrow |a_j| \leq 8^{-4J} \delta + \sum_{i=J+1}^{m} 8^i |a_i|.
\] (5)
We now prove by induction that for \( j = m, m - 1, \ldots, d \),

\[
\sum_{t=j}^{J} |a_t| \leq 8^{-3J} \delta.
\] (6)

When \( m/2 < j \leq m \), (5) implies

\[
\sum_{t=j}^{m} |a_t| \leq (m/2)8^{-4m} \delta \leq 8^{-3m} \delta.
\]

In the general case (5) implies

\[
(2/J) \sum_{t=j}^{J} |a_t| \leq 8^{-4J} \delta + \sum_{t=J+1}^{m} 8^t |a_t| \leq 8^{-4J} \delta + \sum_{r=\log(J)+1}^{\log(m)} 8^{2^r} \sum_{t=1+2^{r-1}}^{2^r} |a_t|
\]

Applying the induction hypothesis, we get

\[
\leq 8^{-4J} \delta + \sum_{r=\log(J)+1}^{\log(m)} 8^{2^r} \sum_{t=1+2^{r-1}}^{2^r} 8^{-3 \cdot 2^r} \delta \leq 8^{-4J} \delta + \sum_{q=0}^{\infty} 8^{-q} \leq 8^{-3J} (2/J) \delta,
\]

which proves the general case of (6).

Finally, for every \( x \in [0, m] \), Lemma 10 and (6) imply

\[
|g(x) - f(x)| \leq \sum_{j=d}^{m} |a_j| |b_j(x)| \leq \sum_{r=[\log d]}^{\log m} 8^{-3 \cdot 2^r} \delta \cdot 8^{2^r} \leq 8^{-2d} \delta \sum_{q=0}^{\infty} 8^{-2q} \leq \delta.
\]

\[
\Box
\]

Acknowledgements

We thank Paul Beame, Pavel Hrubeš, and Alexander Sherstov for useful discussions.

A Approximation theory

The proof relies on a fundamental theorem of Markov, relating the degree of a bounded polynomial to the maximum value of its derivative.

**Theorem 12** (Markov’s Theorem [Che66]). Let \( g : [-1, 1] \rightarrow [-1, 1] \) be computed by a polynomial of degree \( d \). Then \( |g'(y)| \leq d^2 \) for every \( y \in [-1, 1] \).

Markov’s theorem allows us to prove the statement about approximation that we need:

**Proof of Theorem 7**. Let \( d \) be the degree of \( f \) and let \( D = \max_{x\in[0,m]} |f'(x)| \). We can bound \( f \) using \( D \) as follows. The value \( |f(j)| \) is at most \( 1 + \epsilon \) for \( j \in \{0, 1, \ldots, m\} \), and so \( |f(x)| \leq 1 + \epsilon + D/2 \) for \( x \in [0, m] \). On the other hand, \( D \geq \frac{f(m) - f(m-1)}{1} \geq 1 - 2\epsilon \).
Now consider the degree $d$ polynomial $g : [-1, 1] \to [-1, 1]$ given by $g(y) = \frac{f(my/2+m/2)}{1+\epsilon+D/2}$. Since $g'(y) = \frac{(m/2)f'(my/2+m/2)}{1+\epsilon+D/2}$, there is a $y \in [-1, 1]$ such that $|g'(y)| = \frac{Dm/2}{1+\epsilon+D/2}$. By Theorem 12

$$d^2 \geq \frac{Dm/2}{1+\epsilon+D/2} \geq \frac{m(1/2 - \epsilon)}{1+\epsilon+1/2 - \epsilon} = \frac{2m(1/2 - \epsilon)}{3} \Rightarrow d \geq \sqrt{m(1-2\epsilon)/3}.$$

\[ \square \]

**B Bounding the discrepancy for the deterministic case**

Here we give the proof of Theorem 5 [She12]. Let $\mathcal{F} = (T_1, \ldots, T_k)$. We shall need to analyze the discrepancy on a more general class of distributions. Let each family $\mathcal{F}_i$ be sampled independently according to the distribution $\mu$, on a universe of size $\ell_i$. Let $g(\mathcal{F})$ be a cylinder intersection, that is, $g(\mathcal{F}) = \prod_{i=1}^{k} g_i(\mathcal{F})$ where each $g_i$ is 0/1 valued and does not depend on $T_i$. We shall prove that

$$\left| \mathbb{E} \left[ g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right| \leq \prod_{i=1}^{m} \frac{2^{k-1} - 1}{\sqrt{\ell_i}},$$

which implies Theorem 5 since every communication protocol with communication $C$ can be expressed as a sum of $2^C$ cylinder intersections. We prove (7) by induction on $k$.

When $k = 2$, convexity implies that

$$\left| \mathbb{E} \left[ g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right|^2 \leq \mathbb{E}_{T_2} \left[ \mathbb{E}_{T_1} \left[ g_1(\mathcal{F}) \mathbb{E}_{T_i} \left[ g_2(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right] \right]^2

\leq \mathbb{E}_{T_2} \left[ \mathbb{E}_{T_i} \left[ g_2(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right]^2 \leq \mathbb{E}_{T_2, T_1, T'_1} \left[ \mathbb{E}_{T_1, T'_1} \left[ (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i) + \text{Disjoint}(\mathcal{F}'_i)} \right] \right],$$

where here $\mathcal{F} = (T_1, T_2)$ and $\mathcal{F}' = (T'_1, T_2)$. Now for every fixing of $T_1, T'_1$, the inner expectation is 1 when $T_1 = T'_1$, and otherwise it is 0. Thus,

$$\left| \mathbb{E} \left[ g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right|^2 \leq \Pr[T_1 = T'_1] = \prod_{i=1}^{k} \frac{1}{\ell_i},$$

proving the base case.

When $k > 2$, we again use convexity to bound

$$\left| \mathbb{E} \left[ g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right|^2 \leq \mathbb{E}_{T_2, \ldots, T_k} \left[ g_1(\mathcal{F}) \mathbb{E}_{T_i} \left[ \prod_{j=2}^{k} g_j(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i)} \right] \right]^2

\leq \mathbb{E}_{T_1, T'_1} \left[ \mathbb{E}_{T_2, \ldots, T_k} \left[ \prod_{j=2}^{k} g_j(\mathcal{F}) \cdot g_j(\mathcal{F}') \cdot (-1)^{\sum_{i=1}^{m} \text{Disjoint}(\mathcal{F}_i) + \text{Disjoint}(\mathcal{F}'_i)} \right] \right],$$

(8)
for all \( i \) we have

\[
\text{Since (even conditioned on the value of } Z \text{) the size of } Q_i \text{ is distributed identically to the size of } Q_i', \text{ we have}
\]

\[
\text{where here we adopt the convention that } Z_i/|Q_i| \text{ is 0 when } Z_i = 0, |Q_i| = 0. \text{ We shall prove that for all } i,
\]

\[
\Pr[Z_i = 0] \leq \frac{2^{k-1} - 1}{\ell_i},
\]

and

\[
\mathbb{E}_{Z_i, Q_i} \left[ \frac{Z_i}{|Q_i|} \right] \leq \frac{2(2^{k-1} - 1)}{\ell_i(2^{k-2} - 1)}.
\]

Inequalities (12) and (13) imply that

\[
\text{Inequalities (12) and (13) imply that}
\]

\[
\text{where here we adopt the convention that } Z_i/|Q_i| \text{ is 0 when } Z_i = 0, |Q_i| = 0. \text{ We shall prove that for all } i,
\]

\[
\Pr[Z_i = 0] \leq \frac{2^{k-1} - 1}{\ell_i},
\]
as required.

It only remains to prove (12) and (13). Fix \( i \) for the rest of the proof. We start with (12). Let \( S_1, \ldots, S_k \) be the intersections of \( T_1, \ldots, T_k \) with the \( i \)th part of the universe, and let \( S'_i \) be the intersection of \( T'_i \) with the \( i \)th part of the universe. Observe that for all \( w \neq \emptyset \),

\[
\Pr \left[ Z_i = 0 \mid \bigcap_{j=2}^{k-1} S_j = w \right] = \frac{1}{|w|}.
\]

The total number of choices for sets \( S_1, \ldots, S_{k-1} \) can be counted as the number of ways to pick the common intersection point, times the number of configurations for the rest of the universe: \( \ell_i \cdot (2^{k-1} - 1)^{\ell_i - 1} \). Of these, the number of configurations with \( \bigcap_{j=2}^{k-1} S_j = w \) can be counted as the number of choices for the common intersection point in \( w \), times the number of configurations for the rest of the universe: \( |w| \cdot (2^{k-1} - 2)^{\ell_i - |w|} \). So the probability that the two intersection points are the same is

\[
\Pr[Z_i = 0] = \frac{1}{\ell_i \cdot (2^{k-1} - 1)^{\ell_i - 1}} \cdot \sum_{w \neq \emptyset} \frac{|w| \cdot (2^{k-1} - 2)^{\ell_i - |w|}}{|w|} \\
\leq \frac{1}{\ell_i \cdot (2^{k-1} - 1)^{\ell_i - 1}} \cdot (2^{k-1} - 2 + 1)^{\ell_i} = \frac{2^{k-1} - 1}{\ell_i},
\]

proving (12).

Next we prove (13). Let \( p = \frac{2^{k-2} - 1}{2(2^{k-1} - 1)} \). Let \( V = S_1 \cap \ldots \cap S_{k-1} \) be the intersection set of size 1. We claim that for every non-empty set \( q \), and singleton \( v \),

\[
\Pr[Z_i = 1, Q_i = q, V = v] = \begin{cases} 
\leq p^{q-1}(1 - p)^{\ell_i - |q|} / \ell_i & \text{when } v \subseteq q, \\
0 & \text{otherwise.}
\end{cases} \tag{14}
\]

When \( V \) is not contained in \( Q_i \), the value of \( Z_i \) is always 0. On the other hand, when \( v \subseteq q \),

\[
\Pr[Z_i = 1, Q_i = q, V = v] \leq \Pr[Q_i = q, V = v] \\
= (1/\ell_i) \cdot \Pr[Q_i = q | V = v] \\
\leq p^{q-1}(1 - p)^{\ell_i - |q|} / \ell_i,
\]

since every element \( e \notin v \) is included in \( Q_i \) with probability \( p \), independent of all other such elements. Indeed such an element is included in \( S_1 \) with probability \( \frac{2^{k-2} - 1}{2(2^{k-1} - 1)} \), and given that it is included in \( S_1 \), it is excluded from \( S'_i \) with probability 1/2.

When \( |Q_i| = 0 \), we have that \( Z_i = 0 \), and so \( Z_i / |Q_i| = 0 \) by our convention. Thus (14) gives

\[
\mathbb{E}_{Z_i, Q_i} \left[ \frac{Z_i}{|Q_i|} \right] \leq \sum_{v: |v| = 1} \sum_{q: v \subseteq q} p^{q-1}(1 - p)^{\ell_i - |q|} / |q| \\
= \frac{1}{p^{\ell_i}} \sum_{q \neq \emptyset} p^{q-1}(1 - p)^{\ell_i - |q|} \\
\leq \frac{1}{p^{\ell_i}} \cdot (1 - p + p)^{\ell_i} = \frac{2(2^{k-1} - 1)}{\ell_i(2^{k-2} - 1)}
\]

proving (13).
References


