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

Troy Lee
Rutgers University

Adi Shraibman
Weizmann Institute of Science

A brief history of disjointness

- Alice holds set S_1, Bob S_2. Do they share a common element?

A brief history of disjointness

- Alice holds set S_1, Bob S_2. Do they share a common element?
- Deterministic communication complexity n bits

A brief history of disjointness

- Alice holds set S_1, Bob S_2. Do they share a common element?
- Deterministic communication complexity n bits
- co-Nondeterministic complexity is $O(\log n)$.

A brief history of disjointness

- Alice holds set S_1, Bob S_2. Do they share a common element?
- Deterministic communication complexity n bits
- co-Nondeterministic complexity is $O(\log n)$.
- Randomized complexity $\Theta(n)$ [KS87, Raz92]

A brief history of disjointness

- Alice holds set S_1, Bob S_2. Do they share a common element?
- Deterministic communication complexity n bits
- co-Nondeterministic complexity is $O(\log n)$.
- Randomized complexity $\Theta(n)$ [KS87, Raz92]
- Quantum complexity $\Theta(\sqrt{n})$ [lower Raz03, upper AA03]

Number-on-the-forehead model

- k-players, input $x_1, \ldots, x_k$. Player i knows everything but x_i.
- Large overlap in information makes showing lower bounds difficult. Only available method is discrepancy method.
- Lower bounds have application to powerful models such as depth three circuits and complexity of proof systems.
- Best lower bounds are of the form $n/2^k$. Bound of $n/2^{2k}$ for generalized inner product function [BNS89].

Disjointness in the number-on-the-forehead model

- Best lower bound $\Omega(\frac{\log n}{k-1})$, and best upper bound $O(kn/2^k)$ [lower BPSW06, upper Gro94].

Disjointness in the number-on-the-forehead model

- Best lower bound $\Omega(\frac{\log n}{k-1})$, and best upper bound $O(kn/2^k)$ [lower BPSW06, upper Gro94].

- Kushilevitz and Nisan: "The only technique from two-party complexity that generalizes to multiparty complexity is the discrepancy method."

Disjointness in the number-on-the-forehead model

- Best lower bound $\Omega(\frac{\log n}{k-1})$, and best upper bound $O(kn/2^k)$ [lower BPSW06, upper Gro94].
- Kushilevitz and Nisan: "The only technique from two-party complexity that generalizes to multiparty complexity is the discrepancy method." For disjointness, discrepancy can only show bounds of $O(\log n)$.

Disjointness in the number-on-the-forehead model

- Best lower bound $\Omega(\frac{\log n}{k-1})$, and best upper bound $O(kn/2^k)$ [lower BPSW06, upper Gro94].

- Kushilevitz and Nisan: "The only technique from two-party complexity that generalizes to multiparty complexity is the discrepancy method." For disjointness, discrepancy can only show bounds of $O(\log n)$.

- Researchers have studied restricted models—bound of $n^{1/3}$ for three players where first player speaks and dies [BPSW06]. Bound of $n^{1/k}/k^k$ in one-way model [VW07].

Our results

- We show 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.

Our results

- We show 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.

- Chattopadhyay and Ada independently obtained similar results

Application to proof systems

- As linear and semidefinite programming are some of the most sophisticated algorithms we have developed, natural to see how they fare on NP-complete problems.

- One way to formalize this is through proof complexity: for example cutting planes, Lovász-Schrijver proof systems.

Application to proof systems

- As linear and semidefinite programming are some of the most sophisticated algorithms we have developed, natural to see how they fare on NP-complete problems.

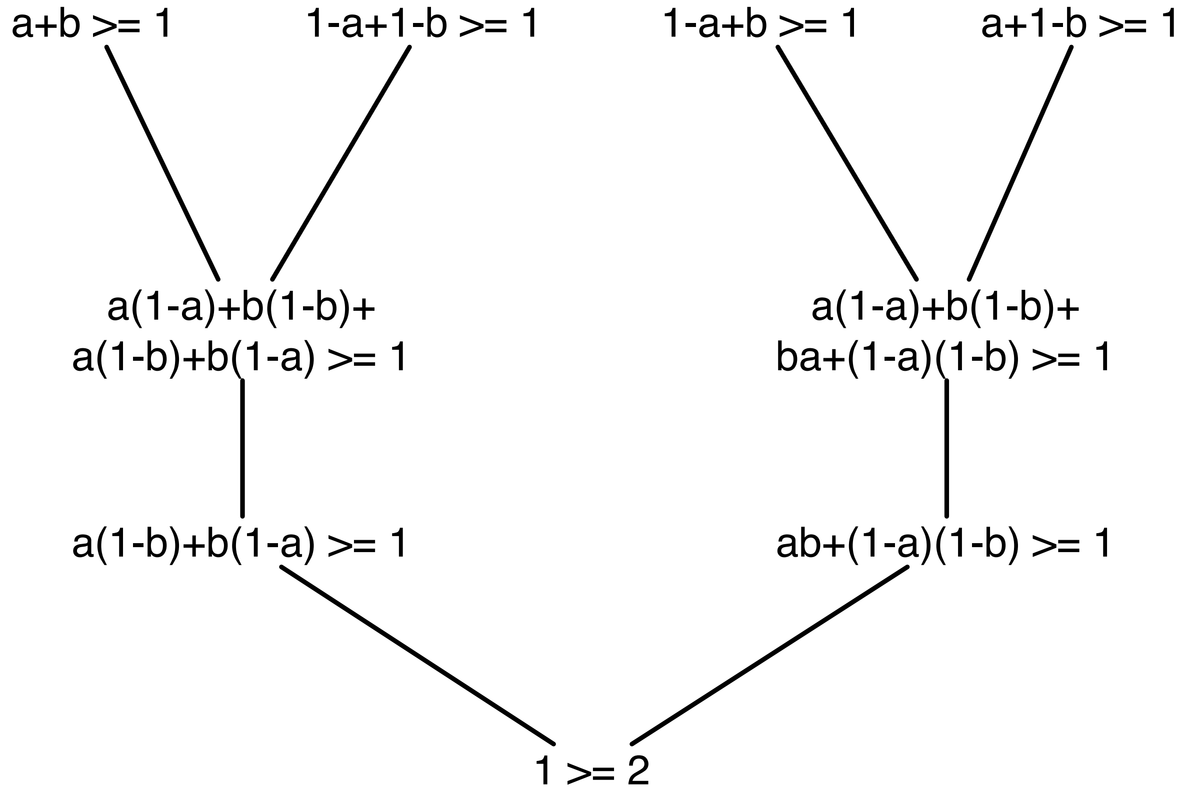
- One way to formalize this is through proof complexity: for example cutting planes, Lovász-Schrijver proof systems.

- Beame, Pitassi, and Segerlind show that lower bounds on disjointness imply lower bounds for a very general class of proof systems, including the above [BPS06].

Semantically entailed proof systems

- Say trying to show a CNF formula ϕ is not satisfiable
- Refutation is a binary tree with nodes labeled by degree d polynomial inequalities and derives $0 \geq 1$.
- Axioms are clauses of ϕ
- Derivation rule is Boolean soundness: if every 0/1 assignment which satisfies f and g also satisfies h, then one may conclude h from f, g.

Example: $(a \vee b) \wedge (\neg a \vee \neg b) \wedge (\neg a \vee b) \wedge (a \vee \neg b)$



Application to proof systems

- Via [BPS06] and our results on disjointness, we obtain super-polynomial lower bounds on the size of *tree-like* degree d semantically entailed proofs needed to refute certain CNFs for any $d = \log\log n - O(\log\log\log n)$.

- Examples: cutting planes, Lovász-Schrijver systems ($d = 2$).

- Exponential bounds known for cutting planes and tree-like Lovász-Schrijver systems, but rely heavily on specific properties of these proof systems. Even for $d = 2$ no such bounds were known in general.

Review of two-party complexity

- Alice and Bob wish to compute a distributed function $f : X \times Y \to \{-1, +1\}$. Consider a $|X|$-by-$|Y|$ matrix where $A[x, y] = f(x, y)$.

- Structural theorem: successful c-bit protocol partitions A into 2^c monchromatic rectangles.

- In particular, the protocol gives us a way to decompose A as

$$A = \sum_i \epsilon_i C_i$$

where $\epsilon_i \in \{-1, 1\}$ and C_i is a 0/1 valued rank-one matrix.

A relaxation

- Define a quantity

$$\mu(A) = \min \left\{ \sum |\alpha_i| : A = \sum_i \alpha_i C_i \right\}$$

  where each C_i is a $0/1$ valued rank-one matrix.

- We have $D(A) \geq \log \mu(A)$.

- The log rank bound is a relaxation in a different direction—each C_i can be an arbitrary rank one matrix, but we count their number rather than their "weight".

Randomized complexity

- For randomized complexity, a protocol gives a decomposition not of A but of a matrix close to A in ℓ_∞ norm.

- To capture this, we consider an approximate version of μ: for $\alpha \geq 1$

$$\mu^\alpha(A) = \min_{A' : J \leq A \circ A' \leq \alpha J} \mu(A')$$

  where J is the all ones matrix.

- One can show that $R_\epsilon(A) \geq \log \mu^\alpha(A) - \log(\alpha)$ for $\alpha = 1/(1-2\epsilon)$.

Dual formulation

- Now we have a lower bound technique, but it seems hard to use as is a minimization problem.

Dual formulation

- Now we have a lower bound technique, but it seems hard to use as is a minimization problem.
- We look at the dual formulation to get a maximization problem which is more convenient for showing lower bounds.

Dual formulation

- Now we have a lower bound technique, but it seems hard to use as is a minimization problem.
- We look at the dual formulation to get a maximization problem which is more convenient for showing lower bounds.
- By definition, the dual norm is

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

- So we see $\mu^*(Q) = \max_C \; |\langle Q, C\rangle|$ where C is $0/1$ valued rank one matrix.

Dual formulation

- By theory of duality we then get

$$\mu(A) = \max_{Q} \frac{\langle A, Q \rangle}{\mu^*(Q)}$$

- This form is more convenient for showing lower bounds— it suffices to exhibit a matrix Q that has non-negligible correlation with A and such that $\mu^*(Q)$ is small.

Dual formulation, approximate versions

The approximate versions of μ also have attractive dual formulations:

$$\mu^\alpha(A) = \max_Q \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)\|Q\|_1}{2\mu^*(Q)}$$

Dual formulation, approximate versions

The approximate versions of μ also have attractive dual formulations:

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

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

Comparison with discrepancy

Discrepancy with respect to probability distribution P is defined as

$$\begin{aligned}\mathrm{disc}_P(A) &= \max_C \langle A \circ P, C \rangle \\ &= \mu^*(A \circ P).\end{aligned}$$

Comparison with discrepancy

Discrepancy with respect to probability distribution P is defined as

$$\begin{aligned}\mathrm{disc}_P(A) &= \max_C \langle A \circ P, C \rangle \\ &= \mu^*(A \circ P).\end{aligned}$$

Thus

$$\frac{1}{\mathrm{disc}(A)} = \max_{\substack{P \geq 0 \\ \|P\|_1 = 1}} \frac{1}{\mu^*(A \circ P)}$$

Comparison with discrepancy

Discrepancy with respect to probability distribution P is defined as

$$\begin{aligned}\mathrm{disc}_P(A) &= \max_C \langle A \circ P, C \rangle \\ &= \mu^*(A \circ P).\end{aligned}$$

Thus

$$\frac{1}{\mathrm{disc}(A)} = \max_{\substack{P \geq 0 \\ \|P\|_1 = 1}} \frac{1}{\mu^*(A \circ P)} = \max_{P \geq 0} \frac{\langle A, A \circ P \rangle}{\mu^*(A \circ P)}$$

Comparison with discrepancy

Discrepancy with respect to probability distribution P is defined as

$$\begin{aligned}\mathrm{disc}_P(A) &= \max_C \langle A \circ P, C\rangle \\ &= \mu^*(A \circ P).\end{aligned}$$

Thus

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

Number-on-the-forehead model

- Instead of a communication matrix, we now have a communication tensor $A[x_1, \ldots, x_k] = f(x_1, \ldots, x_k)$.

- Instead of combinatorial rectangles we now have cylinder intersections.

- Message of player i does not depend on x_i. Behavior can be described as a function ϕ for which

$$\phi(x_1, \ldots, x_i, \ldots, x_k) = \phi(x_1, \ldots, x_i', \ldots, x_k).$$

- We call such a function a cylinder function.

Number-on-the-forehead model

- A cylinder intersection is the intersection of sets which are cylinders. Characteristic function can be written as

$$\phi^1(x_1, \ldots, x_k) \cdots \phi^k(x_1, \ldots, x_k)$$

  where each ϕ^i is a $0/1$ valued cylinder function in the i^{th} dimension.

- Structural theorem: a successful c-bit k-player NOF protocol decomposes the communication tensor into 2^c monochromatic k-fold cylinder intersections.

Our lower bound technique

- Analogous to the two-player case, for a k-tensor A we define

$$\mu(A) = \min \left\{ \sum_i |\alpha_i| : A = \sum_i \alpha_i C_i \right\}$$

  where each C_i is characteristic function of a k-fold cylinder intersection.

- $D_k(A) \geq \log \mu(A)$

Our lower bound technique

- Analogous to the two-player case, for a k-tensor A we define

$$\mu(A) = \min \left\{ \sum_i |\alpha_i| : A = \sum_i \alpha_i C_i \right\}$$

where each C_i is characteristic function of a k-fold cylinder intersection.

- $D_k(A) \geq \log \mu(A)$

- As before we define the approximate version to lower bound randomized complexity:

$$\mu^\alpha(A) = \min_{A' : J \leq A \circ A' \leq \alpha J} \mu(A')$$

Dual formulation

- Now we see that
$$\mu^*(Q) = \max_C |\langle Q, C\rangle|$$
where C is the characteristic function of a cylinder intersection.

- Connection to discrepancy: $\mathrm{disc}_P(A) = \mu^*(A \circ P)$.

$$\mu^\alpha(A) = \max_Q \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)\|Q\|_1}{2\mu^*(Q)}$$

Dual formulation

- Now we see that
$$\mu^*(Q) = \max_C |\langle Q, C\rangle|$$
where C is the characteristic function of a cylinder intersection.

- Connection to discrepancy: $\mathrm{disc}_P(A) = \mu^*(A \circ P)$.

$$\mu^\alpha(A) = \max_Q \frac{(1+\alpha)\langle A, Q\rangle + (1-\alpha)\|Q\|_1}{2\mu^*(Q)}$$

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

Overview of proof

- Task: choose Q. Show $\langle A, Q\rangle$ is non-negligible. Upper bound $\mu^*(Q)$.

Overview of proof

- Task: choose Q. Show $\langle A, Q \rangle$ is non-negligible. Upper bound $\mu^*(Q)$.
- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b], and its extension to the tensor case by Chattopadhyay [Cha07].

Overview of proof

- Task: choose Q. Show $\langle A, Q \rangle$ is non-negligible. Upper bound $\mu^*(Q)$.
- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b], and its extension to the tensor case by Chattopadhyay [Cha07].
- If A is formed from a function f of a single variable in a structured way, we can relate $\mu^\alpha(A)$ to the α-approximate degree of f.

Overview of proof

- Task: choose Q. Show $\langle A, Q \rangle$ is non-negligible. Upper bound $\mu^*(Q)$.

- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b], and its extension to the tensor case by Chattopadhyay [Cha07].

- If A is formed from a function f of a single variable in a structured way, we can relate $\mu^\alpha(A)$ to the α-approximate degree of f.

- Namely, a "witness" q to the high approximate degree of f can be used to construct Q with the right properties.

Pattern Matrix

- Alice holds m-many strings $x = (x_1, \ldots, x_m)$ each of length M.
- Bob holds $S = (a_1, \ldots, a_m)$, each $a_i \in M$, to select bits of x.
- For a function $f : \{0,1\}^m \to \{-1,+1\}$, pattern matrix is defined as

$$A_f[x, S] = f(x_1[a_1], \ldots, x_m[a_m]).$$

- If $f = \mathrm{OR}$ then this is special case of disjointness on mM bits.

Pattern Tensors

- For simplicity, $k = 3$. Now Alice has m many M-by-M *matrices* $x = (x_1, \ldots, x_m)$.

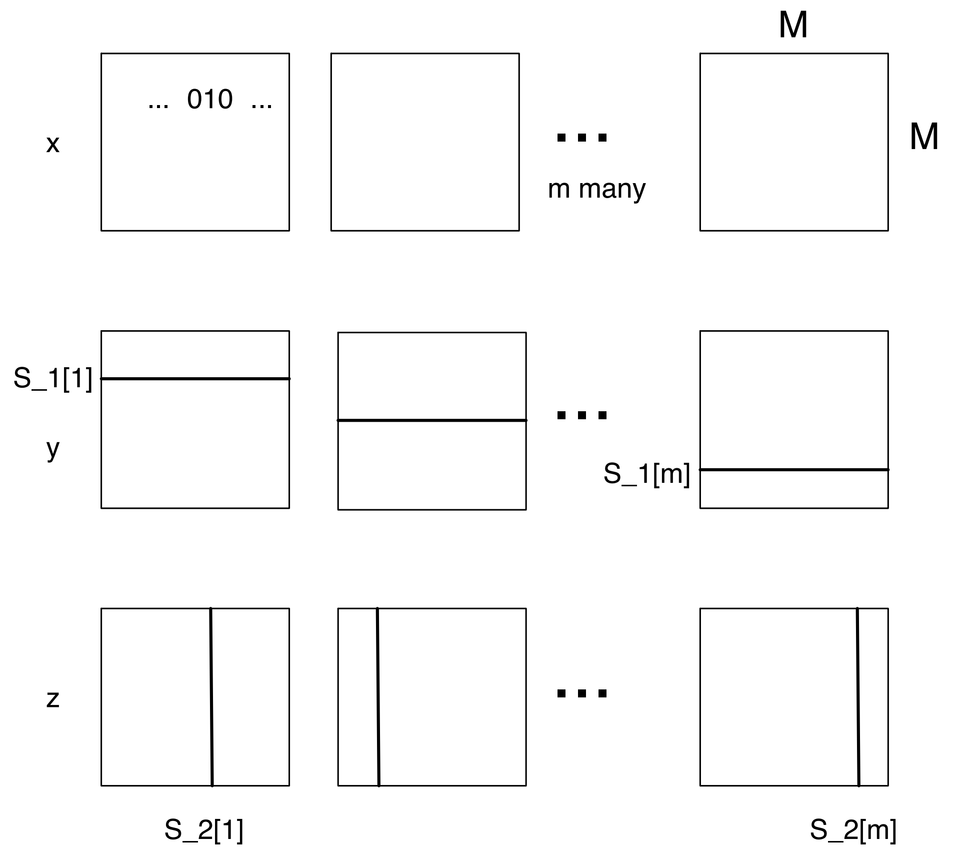
- Bob, Charlie hold $S_1, S_2 \in [M]^m$ to select rows resp. columns of x.

- For a function $f : \{0,1\}^m \to \{-1,+1\}$ define

$$A_f[x, S_1, S_2] = f(x_1[S_1[1], S_2[1]], \ldots, x_m[S_1[m], S_2[m]].$$

- Nice property: every m-bit string appears as input to f equal number of times.

Embedding into disjointness of size mM^2



Building Q from degree witness

- We define approximate degree in a "sign" way
- $\deg_\alpha(f) = \min_g \{\deg(g) : 1 \leq g(x)f(x) \leq \alpha\}$
- In this way, we can uniformly handle both the bounded-error case and the sign or voting polynomial degree which corresponds to $\deg_\infty(f)$.

Dual polynomial

- For a fixed degree d, finding the "best fit" degree d polynomial g can be written as a linear program.
- If f has no α-approximation with degree d, the dual of this program will be feasible, and its solution q will give us a witness to the hardness of f.
- We let $Q = A_q$ be the pattern tensor formed from q, and this will witness that $\mu^{\alpha}(A_f)$ is large.

Dual polynomial

More precisely, if $\deg_\alpha(f) \geq d$ then there exists a polynomial q such that

1. $\|q\|_1 = 1$
2. $\langle f, q \rangle \geq \frac{\alpha-1}{\alpha+1}$
3. q is orthogonal to all polynomials of degree $< d$.

Dual polynomial

More precisely, if $\deg_\alpha(f) \geq d$ then there exists a polynomial q such that

1. $\|q\|_1 = 1$

2. $\langle f, q \rangle \geq \frac{\alpha-1}{\alpha+1}$

3. q is orthogonal to all polynomials of degree $< d$.

We let Q be the pattern tensor formed from q. Item 2 lower bounds $\langle A_f, Q \rangle$. Item 3 is used to upper bound $\mu^*(Q)$.

Main theorem

Let $\alpha < \alpha_0$.

$$\log \mu^{\alpha}(A_f) \geq \frac{\deg_{\alpha_0}(f)}{2^{k-1}} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1}$$

Main theorem

Let $\alpha < \alpha_0$.

$$\log \mu^{\alpha}(A_f) \geq \frac{\deg_{\alpha_0}(f)}{2^{k-1}} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1}$$

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

Main theorem

Let $\alpha < \alpha_0$.

$$\log \mu^{\alpha}(A_f) \geq \frac{\deg_{\alpha_0}(f)}{2^{k-1}} + \log \frac{\alpha_0 - \alpha}{\alpha_0 + 1}$$

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

We can embed the pattern tensor of OR into disjointness to obtain

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

Room for improvement

- We expect the right answer to be $\Omega(n/2^k)$.
- Numerator: $n^{1/2k}$ comes from the reduction. Curse of dimensionality.
- Denominator: factor of 2^{2^k} comes in upper bounding $\mu^*(Q)$.

Conclusion

- Norm approach to communication complexity holds quite generally.
- Our inspiration to the μ norm: γ_2 norm shown to lower bound quantum communication complexity by Linial and Shraibman.
- We can extend the γ_2 norm to tensors to show lower bounds on quantum NOF complexity.
- Turns out all techniques to bound μ also work for γ_2. In particular, we can show bounds of size $n/2^k$ for explicit functions, $n/2^{2k}$ for generalized inner product, and $n^{1/2k}/2^{2^k}$ for disjointness.