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
- 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
- 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, 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 is $O(kn/2^k)$ [lower BPSW06, upper Gro94].

- All existing lower bounds in number-on-the-forehead model use 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 [BPS06].

Application to proof systems

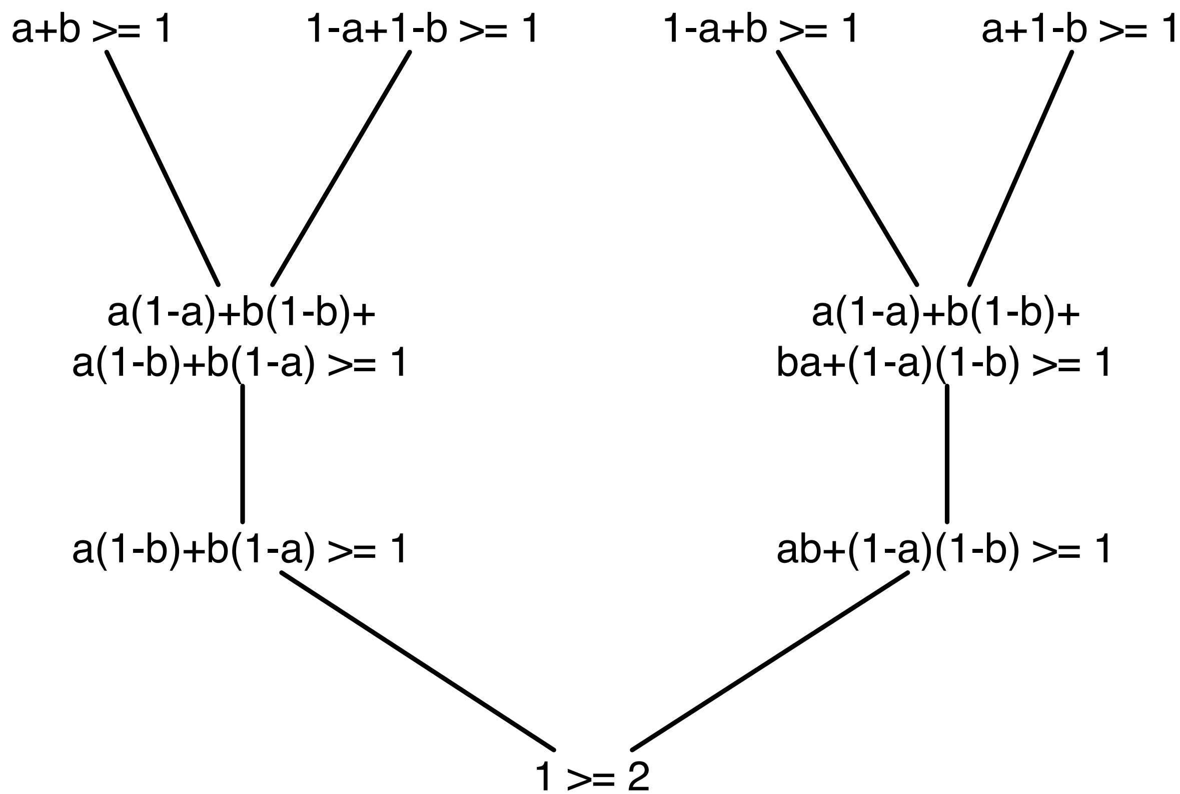
- As linear and semidefinite programming are some of the most sophisticated algorithms we have developed, natural to see how they fare on NP-complete problems.

- 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 [BPS06].

- Semantically entailed proof systems: terms are degree d polynomial inequalities. Derivation rule is Boolean soundness.

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 ($d = 2$), degree d positivstellensatz.

- Exponential bounds were already known for cutting planes and Lovász-Schrijver systems, but relied heavily on the particular geometry 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 $M[x, y] = f(x, y)$.
- A successful protocol partitions M into monchromatic rectangles. This leads to the famous log rank bound.
- More explicitly, the protocol gives us a way to decompose M as

$$M = \sum_i \epsilon_i C_i$$

where $\epsilon_i \in \{-1, 1\}$ and C_i characteristic function of a combinatorial rectangle.

Example: Parity of two bits

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Example: Parity of two bits

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} +$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

A relaxation

- Define a quantity

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

  where each C_i is a combinatorial rectangle.

- Then we have $D(M) \geq \log \mu(M)$.

- 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"

Number-on-the-forehead model

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

- 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 function which is a cylinder in the i^{th} dimension.

- As a two-player protocol decomposes communication matrix into monochromatic rectangles, number-on-the-forehead decomposes communication tensor into monochromatic cylinder intersections.

Our lower bound technique

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

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

  where each C_i is characteristic function of a cylinder intersection.

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

Our lower bound technique

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

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

  where each C_i is characteristic function of a cylinder intersection.

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

- Now we have a lower bound technique, but how do we use it?

Dual norm

- In order to show lower bounds on μ it is helpful to look at its dual norm
- By definition, $\mu^*(Q) = \max_{B:\mu(B)\leq 1} |\langle Q, B\rangle|$
- So we see
$$\mu^*(Q) = \max_C |\langle Q, C\rangle|$$
where C is the characteristic function of a cylinder intersection.

Dual formulation

- By theory of duality we then get

$$\mu(M) = \max_Q \frac{\langle M, Q \rangle}{\mu^*(Q)}$$

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

Randomized Models

- The method can be easily modified for randomized models. Instead of M, the important thing is then tensors which are *close* to M.
- Define $\mu^{\alpha}(M) = \min'_{M} \{\mu(M') : 1 \leq M \circ M' \leq \alpha\}$.
- Motivates the definition $\mu^{\infty}(M) = \min_{M'} \{\mu(M') : 1 \leq M \circ M'\}$
- $R_{\epsilon}(M) \geq \log \mu^{\alpha_{\epsilon}}(M) - \log \alpha_{\epsilon}$, where $\alpha_{\epsilon} = 1/(1 - 2\epsilon)$.

Dual formulation, approximate versions

The approximate versions of μ also have attractive dual formulations:

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

Dual formulation, approximate versions

The approximate versions of μ also have attractive dual formulations:

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

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

Comparison with discrepancy

For a sign tensor M discrepancy is defined as

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

Comparison with discrepancy

For a sign tensor M discrepancy is defined as

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

Thus

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

Comparison with discrepancy

For a sign tensor M discrepancy is defined as

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

Thus

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

Comparison with discrepancy

For a sign tensor M discrepancy is defined as

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

Thus

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

Overview of proof

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

Overview of proof

- Tasks: choose Q. Show $\langle M, Q \rangle$ is non-negligible. Upper bound $\mu^*(Q)$.
- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b].

Overview of proof

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

- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b].

- If M is "derived" from a function f in a structured way, we can relate $\mu^\alpha(M)$ to the approximate degree of f.

Overview of proof

- Tasks: choose Q. Show $\langle M, Q\rangle$ is non-negligible. Upper bound $\mu^*(Q)$.
- We will follow the elegant "pattern matrix" framework of Sherstov [She07a,She07b].
- If M is "derived" from a function f in a structured way, we can relate $\mu^\alpha(M)$ to the approximate degree of f.
- Namely, we can use a "witness" q to the high degree of f to construct Q with the right properties.

Pattern Tensors

Chattopadhyay extends Sherstov's pattern matrices to the multiparty case [Cha07]. We adapt this definition to accommodate disjointness.

- For simplicity, $k = 3$. Let $M \geq m$ and $f : \{0,1\}^m \to \{-1,1\}$
- First player holds x: vector of m many M-by-M matrices
- Second, third players hold $S_1, S_2 \subset [M]^m$ which will index bits of x
- Define $F(x, S_1, S_2) = f(x_1[S_1[1], S_2[1]], \ldots, x_m[S_1[m], S_2[m]])$

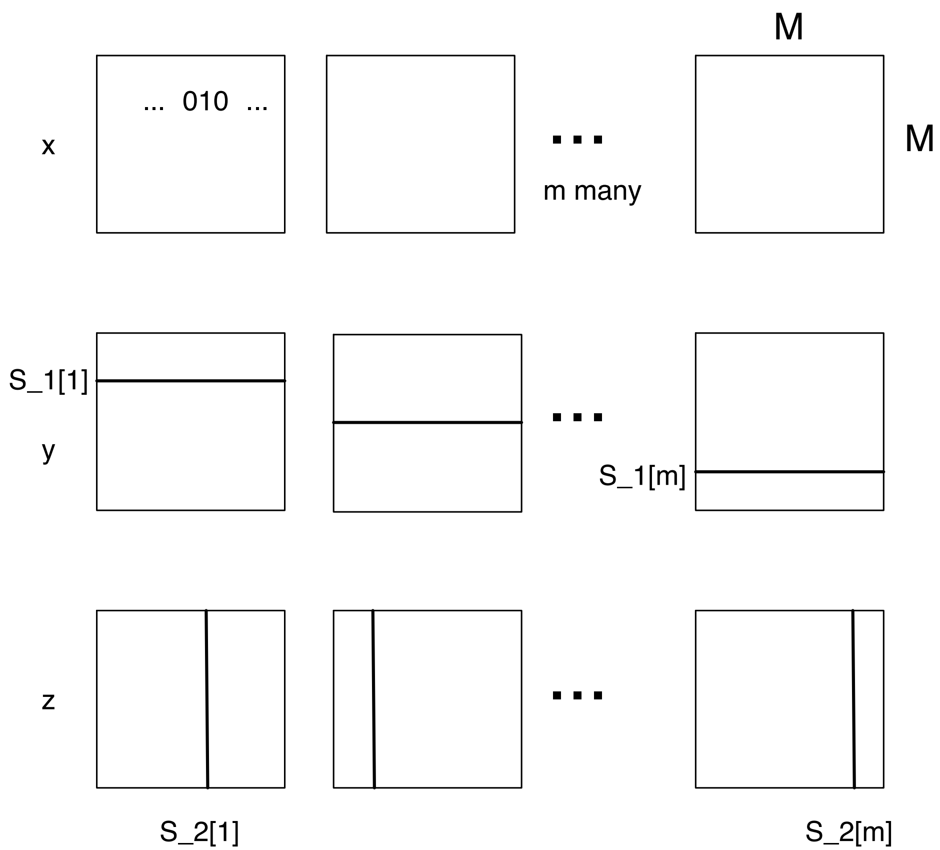
Pattern Tensors

- Every m bit string appears an equal number of times as argument to f.
- When f is the OR function, we can embed F into the disjointness function.

Pattern Tensors

- Every m bit string appears an equal number of times as argument to f.
- When f is the OR function, we can embed F into the disjointness function.
- Think of inputs x, y, z to disjointness as being vectors of m many M-by-M matrices
- x stays the same. Define $y_i[r, c] = 1$ iff $S_1[i] = r$. Similarly, $z_i[r, c] = 1$ iff $S_2[i] = c$.

Picture of the embedding



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 degree-d α-approximation, the dual of this program will be feasible, and its solution q will give us a witness to the hardness of f.

- We will use this witness vector q to construct our tensor Q to witness that μ^{α} is large.

Dual polynomial

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

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

Dual polynomial

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

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

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

Main theorem

Let $\alpha < \alpha_0$.

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

Main theorem

Let $\alpha < \alpha_0$.

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

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

Main theorem

Let $\alpha < \alpha_0$.

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

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

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

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

Where we lose

- $n^{1/2k}$ comes from the reduction. Curse of dimensionality.
- Factor of 2^{2^k} comes in upper bounding $\mu^*(Q)$

More recently. . .

- We can develop an analogous norm γ for the quantum case.
- It turns out that all techniques to upper bound μ^* also work for γ^*
- We can port essentially all known results to quantum case. In particular, we can show bounds of size $n/2^k$ for explicit functions, $n/2^{2k}$ for generalized inner product.