Beating the Direct Sum Theorem in Communication Complexity

Grigory Yaroslavtsev
Pennsylvania State University

Aarhus Universitetet, Theory seminar

Joint work with Marco Molinaro (CMU) and David Woodruff (IBM)
SODA’13
Results

**Stronger Direct Sum Theorem** in communication complexity for equality-type functions

\[
R_{\delta}(f^{k}) \geq \Omega(k) R_{\delta}(f)
\]

\[
D_{\mu,k,\delta}(f^{k}) \geq \Omega(k) D_{\mu,k,\delta}(f)
\]

Yao’s principle

Optimal lower bounds for sketching problems:

- Johnson-Lindenstrauss transform for n vectors
- Pairwise $\ell_1$- and $\ell_2$-distance estimation
- Matrix multiplication
- Join size estimation of multiple databases
Communication Complexity

- **2 deterministic players:** Alice and Bob
- **Joint function** $f$
- **Communicate and compute** $f(x, y)$

$\Pi(x, y)$ denotes transcript or output
Communication Complexity

- 2 deterministic players: Alice and Bob
- Joint function $f$
- Communicate and compute $f$

- **Ex:** $x, y \in \{0,1\}^n$, want to output $x \equiv y$

**Diagram:**

- Alice and Bob communicate through a joint function $f(x, y)$.
- Inputs $x$ and $y$ are exchanged between Alice and Bob.
- The goal is to compute $f(x, y)$.

**Figure:**

- Alice sends $x$ to Bob.
- Bob sends $y$ to Alice.
- They exchange additional messages labeled $\Pi$.
- The output is $f(x, y)$. 
Communication Complexity

• Consider distribution $\mu$ over inputs

• **Goal:** Compute $f(x, y)$ for all but $\delta \mu$-fraction of inputs while minimizing longest communication

• **Distributional complexity**

  $D_{\mu, \delta}(f) = \text{minimum communication over all } \delta\text{-protocols}$
Multiple Instances

\[ f(x_1, y_1), \quad f(x_2, y_2), \quad \ldots, \quad f(x_k, y_k) \]

\[ f^k(x, y) \]
Multiple Instances

- **Goal**: Compute $f^k(x, y)$ for all but $\delta \mu^k$-fraction of inputs while minimizing longest communication

- **Distributional complexity**: $D_{\mu^k, \delta}(f^k)$
Multiple Instances

Main question: How much can we save against solving each independently?

\[ D_{\mu^k, \delta}(f^k) \geq \Omega(k) \bullet_{\delta} f \]

- Sometimes a bit: in the private randomness model

\[ D_{\mu^k, \delta}(f^k) \geq \Omega(\log n) \quad \text{but} \quad D_{\mu^k, \delta}(E Q_n) = O(n) \quad [FKNN95] \]
Multiple Instances

\[ D_{\mu^k, \delta}(f^k) \geq \Omega(k). D_{\mu, \delta} f \]

- **Direct sum theorems**
  
  - \[ D_{\mu^k} f^k \geq \Omega(\sqrt{k}). D_{\mu^k} f \]  
    \[ \text{BBCR 10} \]
  
  - None attains above bound

- **Direct product theorems**
  
  - \[ D_{\mu^k, 1-(1-\frac{1}{3})^k} f^k \geq \Omega(\sqrt{k}). D_{\mu, \frac{1}{3}} f \]  
    \[ \text{BRWY} \]
Information Complexity

• **Information cost:** For protocol \( \Pi \) and \( (X, Y) \sim \mu \), information revealed about input is

\[
\text{IC}_\mu(\Pi) = I(\Pi(X, Y); X, Y) = H(X, Y) - H(X, Y | \Pi)
\]

• **Information complexity:**

\[
\text{IC}_{\mu, \delta}(f) = \min \text{ information cost over all } \delta\text{-protocols}
\]

**Connection:** Communication is at least information

\[
D_{\mu, \delta}(f) \geq \text{IC}_{\mu, \delta}(f)
\]

• For non-product \( \mu \) we will work with **conditional information complexity** \( \text{IC}_{\mu, \delta}(f | \nu) \)
Protocols with Abortion

**Def:** A protocol $\Pi (\beta, \delta)$-computes $f$ if

- (Abortion) $\Pr(\Pi(X, Y) = \text{abort}) \leq \beta$
- (Error) $\Pr(\Pi(X, Y) \neq f(X, Y) \mid \Pi(X, Y) \neq \text{abort}) \leq \delta$

**Obs:** Stronger guarantee than being wrong with prob. $\approx \beta + \delta$

$\text{IC}_{\mu, \beta, \delta}(f) = \min \text{ information cost} \text{ over all protocols that} \ (\beta, \delta)\text{-compute } f$
Theorem: For every communication problem

Solving $k$ copies with error $\delta$ requires solving each copy with constant abortion and error $\frac{\delta}{k}$
Stronger Direct Sum Theorem

- The distribution $\mu$ of $(X,Y)$ is a **product** distribution if
  $\mu(X,Y) = \mu_X(X)\mu_Y(Y)$
- $(\mu, \nu)$ is a **mixture of product distributions**, if for every $t$
  the distribution $(\mu|\nu = t)$ is a product distribution

**Theorem:** For every communication problem $f$, mixture of product distributions $(\mu, \nu)$ and $\delta > 0$

$$IC_{\mu^k,\delta} (f^k | \nu^k) \geq \Omega(k) IC_{\mu, \frac{\delta}{10}O(k)} (f | \nu)$$

Also holds for one-way and bounded-round communication
**Theorem:** For every communication problem and product $\mu$

\[
\text{IC}_{\mu^k, \delta}(f^k) \geq \Omega(k) \text{IC}_{\mu, \frac{\delta}{10}, O\left(\frac{\delta}{k}\right)}(f)
\]

Here is a diagram illustrating the interaction:

- Alice and Bob communicate with $x_1, x_2, \ldots, x_k$ and $y_1, y_2, \ldots, y_k$.
- The function $f$ operates on pairs $(x_i, y_i)$ for $i = 1, 2, \ldots, k$.
- The result is denoted as $f^k(w)$, where $w = [x_1, x_2, \ldots, x_k, y_1, y_2, \ldots, y_k]$. 
Stronger Direct Sum Theorem

Proof: Consider protocol $\Pi$ that computes $f^k$ with prob $1 - \delta$. Want to show

$$I(\Pi(W); W) \geq \Omega(k) \text{IC}_{\mu, \frac{\delta}{10}, O\left(\frac{\delta}{k}\right)}(f)$$

1) Chain rule:

$$I(\Pi(W); W) = \sum_{i=1}^{k} I(\Pi(W); W_i|W_{<i})$$

By averaging suffices to show that for at least $\Omega(k)$ values of $i$

$$I(\Pi(W); W_i|W_{<i}) \geq \text{IC}_{\mu, \frac{\delta}{10}, O\left(\frac{\delta}{k}\right)}(f)$$

Want to obtain from $\Pi$ a protocol with abortion to solve $i$-th copy $f_i^k$ with
error prob $\frac{\delta}{k}$ and information cost cost at most

$$I(\Pi(W); W_i|W_{<i})$$
Stronger Direct Sum Theorem

2) Conditioning amplifies success: For typical $i$

- $\Pr(\Pi_{<i}(W) \neq f^k_{<i}(W)) \leq \delta$
- $\Pr(\Pi_i(W) \neq f^k_i(W) \mid \Pi_{<i}(W) = f^k_{<i}(W)) = O\left(\frac{\delta}{k}\right)$

This is because

$$1 - \delta \leq \Pr(\Pi = f^k) = \prod_{i=1..k} \Pr(\Pi_i = f^k_i \mid \Pi_{<i} = f^k_{<i})$$

3) Theorem: For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, w_{>i} \in G)$ the error probability on $W_i$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}w_iw_{>i}) \neq f^k_i(W_i) \mid \Pi_{<i}(w_{<i}w_iw_{>i}) = f^k_{<i}(w_{<i}w_iw_{>i})) = O\left(\frac{\delta}{k}\right)$
**Stronger Direct Sum Theorem**

**Theorem:** For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, w_{>i}) \in G$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}w_iw_{>i}) \neq f_i^k(w_i) | \Pi_i(w_{<i}w_iw_{>i}) = f_i^k(w_{<i}w_iw_{>i})) = O\left(\frac{\delta}{k}\right)$

**Protocol with abortion for solving $f_i^k$**

Alice

Bob

$x_2$

$y_2$

$i = 2$
Theorem: For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ and random seeds such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, (w_{>i}, r) \in G)$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}W_iw_{>i}) \neq f_{i}^{k}(W_i) \mid \Pi_i(w_{<i}W_iw_{>i}) = f_{<i}(w_{<i}W_iw_{>i})) = O\left(\frac{\delta}{k}\right)$

Protocol with abortion for solving $f_{i}^{k}$

- Fix the prefix $w_{<i}$ and sample $W_{>i}$
**Stronger Direct Sum Theorem**

**Theorem:** For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ and random seeds such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, W_{>i}, r) \in G$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}W_{i}w_{>i}) \neq f_i^k(W_i) | \Pi_{<i}(w_{<i}W_{i}w_{>i}) = f_{<i}^k(w_{<i}W_{i}w_{>i})) = O\left(\frac{\delta}{k}\right)$

**Protocol** with abortion for solving $f_i^k$

- Fix the prefix $w_{<i}$ and sample $W_{>i}$
- Run $\Pi$

---

**Diagram:**

- Alice
  - $1$ to $x_2$
- Bob
  - $0$ to $y_2$
- $i = 2$
**Stronger Direct Sum Theorem**

**Theorem:** For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ and random seeds such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, (w_{>i}, r)) \in G$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}W_iw_{>i}) \neq f^k_i(W_i) \mid \Pi_i(w_{<i}W_iw_{>i}) = f^k_i(w_{<i}W_iw_{>i})) = O\left(\frac{\delta}{k}\right)$

**Protocol** with abortion for solving $f_i^k$

- Fix the prefix $w_{<i}$ and sample $W_{>i}$
- Run $\Pi$
- Verify if error on some copy 1,2, ... $i - 1$
  - If so, “abort”
**Stronger Direct Sum Theorem**

**Theorem:** For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ and random seeds such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, (w_{>i}, r)) \in G$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i} w_i w_{>i}) \neq f_{i}^{k}(W_i) | \Pi_i(w_{<i} w_i w_{>i}) = f_{<i}^{k}(w_{<i} W_i w_{>i})) = O\left(\frac{\delta}{k}\right)$

**Protocol** with abortion for solving $f_{i}^{k}$

- Fix a typical prefix $w_{<i}$ and sample $W_{>i}$
- Run $\Pi$
- Verify if error on some copy $1, 2, \ldots i - 1$
  - If so, “abort”
  - Else report $i$-th output
**Stronger Direct Sum Theorem**

**Theorem:** For a typical $i$ there exists a prefix $w_{<i} \in X^{i-1} \times Y^{i-1}$ and a set $G$ of fixings of the suffix of $\Pi$ and random seeds such that:

1. Information cost only changes by a constant factor after fixing $w_{<i}$
2. $G$ is a constant fraction of all fixings
3. For every fixing $(w_{<i}, (w_{>i}, r)) \in G$ the error probability on $w_{<i}$ is $\leq \frac{\delta}{10}$
4. $\Pr(\Pi_i(w_{<i}w_{>i}) \neq f_i^k(W_i) \mid \Pi_i(w_{<i}w_{>i}) = f_i^k(w_{<i}w_{>i})) = O\left(\frac{\delta}{k}\right)$

**Protocol** with abortion for solving $f_i^k$

- Fix a typical prefix $w_{<i}$ and sample $W_{>i}$
- Run $\Pi$
- Verify if error on some copy $1, 2, \ldots, i - 1$
  - If so, “abort”
  - Else report $i$-th output

Protocol $\left(\frac{\delta}{10}, O\left(\frac{\delta}{k}\right)\right)$-computes $f_i^k$ and has information cost exactly $I(\Pi(w_{<i}W_{\geq i}); W_i)$
Recap

**Theorem:** For every communication problem

\[ \text{IC}_{\mu^k, \delta}(f^k | \nu^k) \geq \Omega(k) \text{ IC}_{\mu, \frac{1}{20}, \frac{\delta}{10}, O(\frac{\delta}{k})} (f | \nu) \]

**Corollary**  For equality-type problems

\[ D_{\mu^k, \delta}(f^k) \geq \Omega(k) \text{ IC}_{\mu, \frac{1}{20}, \frac{\delta}{10}, O(\frac{\delta}{k})} (f | \nu) \geq \Omega(k) D_{\mu, \frac{\delta}{k}} (f) \]
Protocols with abortion

- A protocol \((\alpha, \beta, \delta)\)-computes \(f\) if with probability \(\geq 1 - \alpha\) over its randomness
  - It aborts with probability \(\leq \beta\)
  - Conditioned on non-abortion is correct w.p. \(\geq 1 - \delta\)

- \((\mu, \nu)\) is a convex combination of product distributions over \(((X \times Y) \times D)\) (marginals: \(\mu\) over \((X \times Y)\) and \(\nu\) over \(D\))

- \(IC_{\mu,\alpha,\beta,\delta}(f | \nu)\) = minimum information cost of a protocol which \((\alpha, \beta, \delta)\)-computes over \((\mu, \nu)\).

- \(IC_{\mu,\delta}(f | \nu) = IC_{\mu,0,0,\delta}(f | \nu)\)
**Strong direct sum**

- **Strong direct sum**: For every function $f$ and a convex combination of product distributions $(\mu, \nu)$
  \[
  IC_{\mu^k, \delta}(f^k | \nu^k) \geq \Omega(k) \ IC_{\mu, \frac{1}{20}, \frac{1}{10}, \delta}(f | \nu)
  \]

- **Strong** because of high success probability $(1 - \frac{\delta}{k})$

- Gives an extra $\log k$ in the lower bound as compared to a weak direct sum [Bar-Yossef, Jayram, Kumar, Sivakumar]
  \[
  IC_{\mu^k, \delta}(f^k | \nu^k) \geq \Omega(k) \ IC_{\mu, \delta}(f | \nu)
  \]
One-way Equality with abortion

- $EQ^\ell(x,y) = 1$ iff $x = y$, where $x, y \in \{0,1\}^\ell$

- **Theorem:** For $\ell = \log(1/20\delta)$ there exists $(\mu, \nu)$:
  \[ IC \xrightarrow{\mu, \frac{1}{20\cdot10\cdot\delta}} (EQ^\ell | \nu) = \Omega(\log(1/\delta)) \]

- **Corollary:** solving $k$ copies of Equality with constant probability requires one-way communication $\Omega(k \log k)$ (for sufficiently long strings $x_i, y_i$)

- Hard distribution ($(X Y) D_0 D$)
  - Random variable for conditioning: $(D_0 D) \sim U(\{0,1\}^{\ell+1})$
  - If $D_0 = 0$ then $(X Y) \sim U(\{0,1\}^\ell) \times U(\{0,1\}^\ell)$
  - If $D_0 = 1$ then $X = Y = D$
Equality with abortion

- Hard distribution($\mathbf{X}, \mathbf{Y}, \mathbf{D}$):
  - $(\mathbf{D}, \mathbf{D}') \sim U(\{0,1\}^{\ell+1})$
  - If $D_0 = 0$ then $(\mathbf{X}, \mathbf{Y}) \sim U(\{0,1\}^\ell) \times U(\{0,1\}^\ell)$
  - If $D_0 = 1$ then $\mathbf{X} = \mathbf{Y} = \mathbf{D}$ ($\geq \frac{1}{2}$ of the mass on the diagonal)

For $\ell = \log \frac{1}{20\delta}$
- $p(x,x) = 200\delta^2 + 10\delta$
- $p(x,y) = 200\delta^2$

$y \in \{0,1\}^\ell$

$x \in \{0,1\}^\ell$
Equality with abortion

- $IC_{\mu, \ldots, v} (f | v) = \min_M I(M(X); X, Y | D_0 D) = \min_M I(M(X); X | D_0 D)$
- $I(M(X); X | D_0 D) = H(X | D_0 D) - H(X | M(X), D_0 D)$
- $H(X | D_0 D) \geq \Pr[D_0 = 0] \cdot H(X | D_0 = 0, D)) = 1/2 \log (1/20\delta)$
- By Fano’s inequality ([Cover, Thomas]):
  $H(X | M(X), D_0 D) \leq 1 + p_e \log(|supp(X)|) = 1 + p_e \log (1/20\delta)$,
  where $p_e = \min_g \Pr[g(M(X, D_0 D)) \neq X]$ and $g$ is a deterministic function
- Suffices to show that there exists a predictor with error $p_e \leq \frac{2}{5} < \frac{1}{2}$
Predictor for Equality

• A row $x$ is **good** if the protocol $\Pi(x, y) = 1$ iff $x = y$

• If the $\Pi \left( 0, \frac{1}{10}, \delta \right)$-computes $EQ^\ell$ then $\leq \frac{3}{10}$ fraction of rows is not **good**:
  - Fraction of rows with an abortion on the diagonal $(x, x)$ is $\leq 1/5$
  - Fraction of rows with an error is at most $\leq 1/10$ $\forall y \in \{0,1\}^\ell$

• **Predictor**: If the row is **good** then Bob can simulate $\Pi$ for every $y$ and recover $x$!

• If $\Pi \left( \frac{1}{20}, \frac{1}{10}, \delta \right)$-computes $EQ^\ell$ then $p_e \leq \frac{3}{10} + \frac{1}{20} < \frac{2}{5}$
Augmented indexing

- Augmented indexing over large alphabet ($x_i, y \in [m]$)

\[
\begin{align*}
\text{Alice} (x_1, \ldots, x_N) & \quad M(x) \quad \text{Bob} (i, x_1, \ldots, x_{i-1}, y) \quad x_i = y? \\
\end{align*}
\]

- **Theorem:** For sufficiently large $m$ there exists $(\mu, \nu)$:

\[
\begin{align*}
IC_{\mu,1,1,1}^{\rightarrow_{\Omega \frac{1}{20}, \frac{1}{10}, \frac{1}{m}}} (\text{Augmented Indexing} | \nu) = \Omega(N \log m)
\end{align*}
\]

- **Corollary:** Solving $k$ copies of Augmented Indexing (with const. prob.) requires one-way communication \(\Omega(Nk \log k)\) (for sufficiently large alphabet size)
Application: JL-transform of \( n \) vectors

- Let \( S \) be a distribution over \( k \times d \) matrices, such that for any \( \mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^d \) with prob. \( \geq 1 - \delta \)
  \[
  \|S\mathbf{v}_i - S\mathbf{v}_j\|_2 = (1 \pm \epsilon) \|\mathbf{v}_i - \mathbf{v}_j\|_2
  \]

- \( k = \# \) rows in \( S \geq \frac{1}{\epsilon^2} \log \left( \frac{n}{\delta} \right) \), dependence on \( n \) is new
- Even if \( S \) is allowed to depend on the first \( n/2 \) points
- Any encodinging \( \phi(\mathbf{v}_1), \ldots, \phi(\mathbf{v}_n) \) that allows pairwise \( \ell_p \)-distance estimation for \( p \in \{1,2\} \) requires
  \[
  \Omega \left( n \epsilon^{-2} \log \frac{n}{\delta} (\log d + \log M) \right) \text{ bits}
  \]
  \((M = \max \text{ abs. value in } \mathbf{v}_i)\)
Other applications

• Sketching matrix products
  – Minimum number of columns in a $n \times k$ matrix $S$ such that $C = A S S^T B$ is a good approximation for $A B$, where $A, B$ are $n \times n$ matrices?
    $\begin{align*}
    |(AB)_{i,j} - C_{i,j}| &\leq \epsilon |A_i|_2 |B_j|_2 \Rightarrow k = O(\epsilon^{-2} \log \frac{n}{\delta}) \quad [\text{Sarlos}] \\
    ||A B - C||_F &\leq \epsilon ||A||_F ||B||_F \Rightarrow k = O(\epsilon^{-2} \log \frac{1}{\delta}) \quad [\text{Clarkson,Woodruff}] \\
    \text{Our result: entry-wise guarantee indeed requires } k = \Omega(\epsilon^{-2} \log \frac{n}{\delta})
    \end{align*}$

• Optimality of database sketching [Alon, Gibbons, Matias, Szegedy] and mergeable summaries
Open problems

• **Strong direct sum:** For every function $f$ and a convex combination of product distributions $(\mu, \nu)$

$$IC_{\mu^k, \delta}(f^k | \nu^k) \geq \Omega(k) IC_{\mu, \frac{1}{20}, \frac{1}{10}, \delta}(f | \nu)$$

• More problems with low-error one-way lower bounds?
• Natural problems for low-error 2-way lower bounds (disjointness doesn’t work)?
• Applications of direct sums to property testing? [Blais, Brody, Matulef ’11, Goldreich ‘13]
• Strong direct sum for predicates $g(f(x_1, y_1), \ldots, f(x_k, y_k))$?
  For OR-EQUALITY ($g = \lor, f = EQ^\ell$) there is a direct sum [Brody, Chakrabarti, Kondapally’12, Saglam, Tardos’13]