An Optimal Lower Bound for the Gap-Hamming-Distance Problem

Amit Chakrabarti

Dartmouth College

Joint work with Oded Regev, Tel Aviv University

ICM Satellite Conference, Bangalore, Aug 2010

The Gap-Hamming-Distance Problem

Input: Alice gets $x \in \{0,1\}^n$, Bob gets $y \in \{0,1\}^n$.

Output:

- $\text{GHD}(x,y) = 1$ if $\Delta(x,y) > \frac{n}{2} + \sqrt{n}$
- $\text{GHD}(x,y) = 0$ if $\Delta(x,y) < \frac{n}{2} - \sqrt{n}$

Want: randomized, constant error protocol

Cost: Worst case number of bits communicated

$x =$	0	1	0	0	1	0	1	1	0	0	0	1
$y =$	0	0	0	0	0	0	1	1	1	0	0	1

$n = 12; \quad \Delta(x,y) = 3 \in [6 - \sqrt{12}, 6 + \sqrt{12}]$

Data Stream Lower Bounds

Data streams: two broad application scenarios

- **Networks:** Busy router, packets whizzing by
  - Web traffic statistics
  - Intrusion detection
- **Databases:** Huge DB, linear scan cheaper than random access
  - Query optimisation: join size estimation
  - Log analysis

Data Stream Lower Bounds

Data streams: two broad application scenarios

- **Networks:** Busy router, packets whizzing by
  - Web traffic statistics
  - Intrusion detection
- **Databases:** Huge DB, linear scan cheaper than random access
  - Query optimisation: join size estimation
  - Log analysis
- DB setting: Multiple passes meaningful

GHD Motivation: Obtain pass/space tradeoffs for some basic data stream problems [Indyk-Woodruff'03], [Woodruff'04], [C.-Cormode-McGregor'07]

Data Stream Model

- Formally: input stream $= n$ tokens, each token $\in [m]$
  - Assume $\log m = \Theta(\log n)$
- Compute some function of stream, using
  - Small space, $s \ll m, n$ ... ideally, $s = O(\log n)$
  - Small number of passes, p

Data Stream Model

- Formally: input stream $= n$ tokens, each token $\in [m]$
  - Assume $\log m = \Theta(\log n)$
- Compute some function of stream, using
  - Small space, $s \ll m, n$ ... ideally, $s = O(\log n)$
  - Small number of passes, p
- Give ε-approx:

$$\Pr\left[\left|\frac{\text{output}}{\text{answer}} - 1\right| \le \varepsilon\right] \;\ge\; \frac{2}{3}$$

Problems of Interest

- Distinct elements
- Frequency moments
- Empirical entropy

Problems of Interest

- Distinct elements , F_0
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$

Problems of Interest

- Distinct elements , F_0
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$
- **Key question:** Want ε-approx; then $s = ??$
  - Upper bounds: $O(\varepsilon^{-2}\,\mathrm{polylog}(m, n))$, using 1 pass
  - Showing $\mathrm{R}(\textsc{ghd}) = \Omega(n^c)$ would imply $s = \Omega(\varepsilon^{-2c})$

Problems of Interest

- Distinct elements , F_0
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$
- **Key question:** Want ε-approx; then $s = ??$
  - Upper bounds: $O(\varepsilon^{-2}\,\mathrm{polylog}(m,n))$, using 1 pass
  - Showing $\mathrm{R}(\textsc{ghd}) = \Omega(n^c)$ would imply $s = \Omega(\varepsilon^{-2c})$
  - Showing $\mathrm{R}_{\max}^{2p-1}(\textsc{ghd}) = \Omega(n^c)$ would imply the same for p-pass algorithms
  - In particular, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) \longrightarrow$ 1-pass algorithms

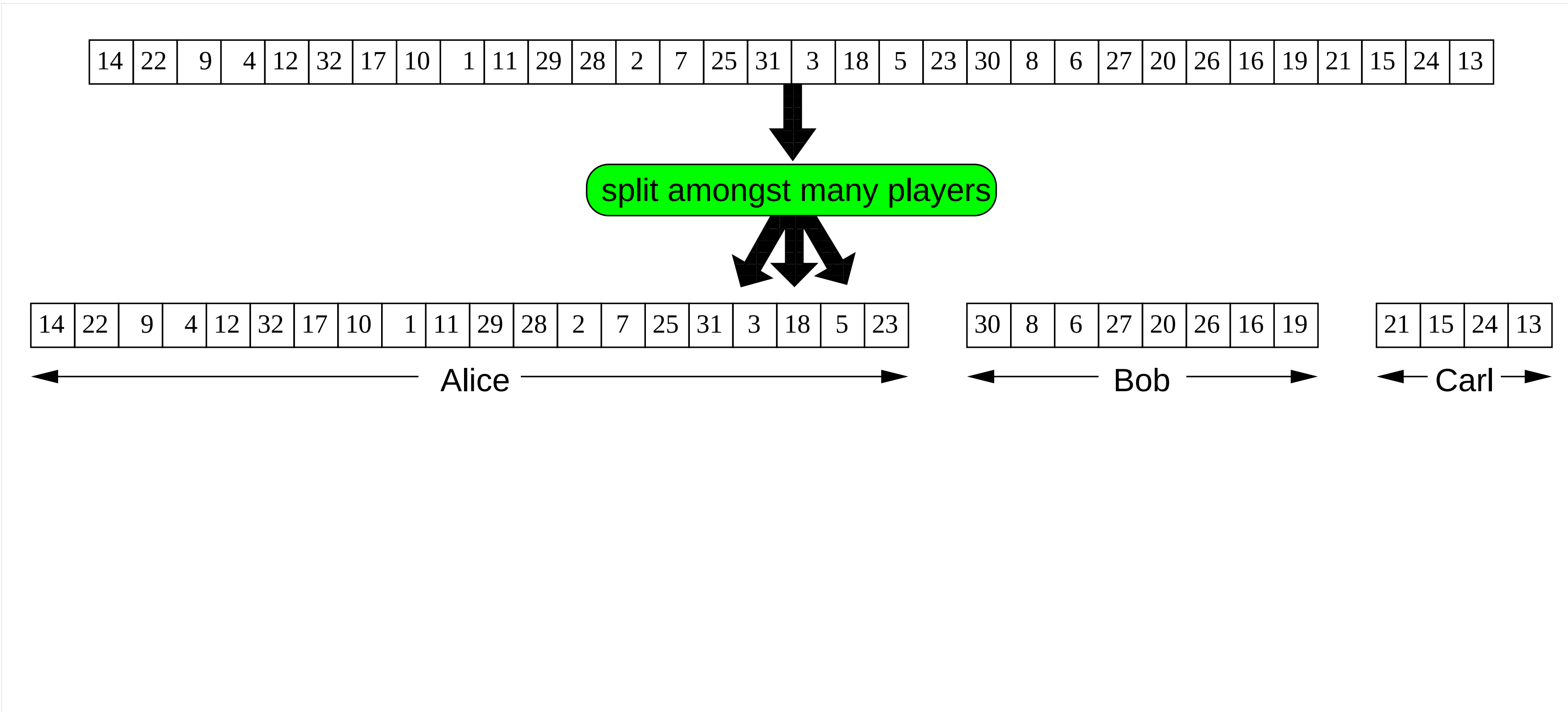
Problems of Interest

- Distinct elements , F_0
- Frequency moments , $F_k = \sum_{i=1}^{m} \mathrm{freq}(i)^k$
- Empirical entropy , $H = \sum_{i=1}^{m} (\mathrm{freq}(i)/m) \cdot \log(m/\mathrm{freq}(i))$
- **Key question:** Want ε-approx; then $s = ??$
  - Upper bounds: $O(\varepsilon^{-2}\,\mathrm{polylog}(m,n))$, using 1 pass
  - Showing $\mathrm{R}(\textsc{ghd}) = \Omega(n^c)$ would imply $s = \Omega(\varepsilon^{-2c})$
  - Showing $\mathrm{R}^{2p-1}_{\max}(\textsc{ghd}) = \Omega(n^c)$ would imply the same for p-pass algorithms
  - In particular, $\mathrm{R}^{\rightarrow}(\textsc{ghd}) \longrightarrow$ 1-pass algorithms
  - Dependence of s on n: [A-M-S'96]; [C.-Khot-Sun'03]; [Gronemeier'09]

Method: Reduce from Communication Complexity

14	22	9	4	12	32	17	10	1	11	29	28	2	7	25	31	3	18	5	23	30	8	6	27	20	26	16	19	21	15	24	13

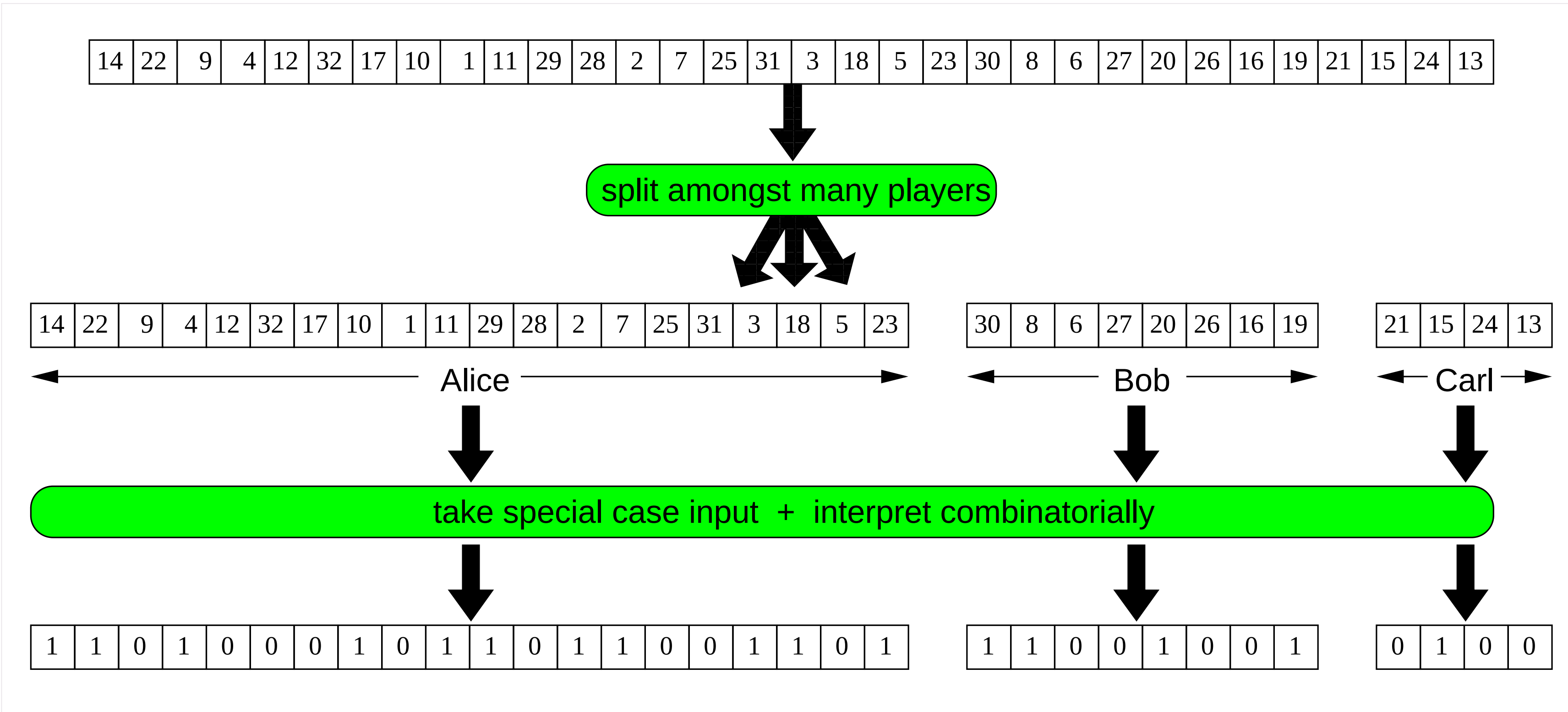
Communication vs Data Stream



p-pass streaming algorithm $\Longrightarrow$ $\Theta(p)$-round communication protocol

messages $=$ memory contents of streaming algorithm

Communication vs Data Stream



p-pass streaming algorithm $\Longrightarrow$ $\Theta(p)$-round communication protocol

messages $=$ memory contents of streaming algorithm

The Reductions

E.g., Distinct Elements (Other problems: similar)

x =	0	1	0	0	1	0	1	1	0	0	0	1
σ :	(1,0)	(2,1)	(3,0)	(4,0)	(5,1)	(6,0)	(7,1)	(8,1)	(9,0)	(10,0)	(11,0)	(12,1)
y =	0	0	0	0	0	0	1	1	1	0	0	1
τ :	(1,0)	(2,0)	(3,0)	(4,0)	(5,0)	(6,0)	(7,1)	(8,1)	(9,1)	(10,0)	(11,0)	(12,1)

Alice: $x \longmapsto \sigma = \langle (1, x_1), (2, x_2), \ldots, (n, x_n) \rangle$

Bob: $y \longmapsto \tau = \langle (1, y_1), (2, y_2), \ldots, (n, y_n) \rangle$

Notice: $F_0(\sigma \circ \tau) = n + \Delta(x, y) = \begin{cases} < \frac{3n}{2} - \sqrt{n}, & \text{or} \\ > \frac{3n}{2} + \sqrt{n}. \end{cases}$ Set $\varepsilon = \frac{1}{\sqrt{n}}$.

History

History

FOCS 2003: $\mathrm{R}^{\rightarrow}$(not-quite-GHD) $=$ "$\Omega(n)$" [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

History

FOCS 2003: $\mathrm{R}^{\rightarrow}(\text{not-quite-}\textsc{ghd}) = \text{“}\Omega(n)\text{”}$ [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

SODA 2004: $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

Very intricate combinatorial proof

History

FOCS 2003: $\mathrm{R}^{\rightarrow}(\text{not-quite-}\textsc{ghd}) =$ "$\Omega(n)$" [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

SODA 2004: $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

Very intricate combinatorial proof

GVIN 2005: Simplification of proof [Jayram-Kumar-Sivakumar]

Nice geometric intuition + reduction from INDEX

History

FOCS 2003: $\mathrm{R}^{\rightarrow}$(not-quite-$\textsc{ghd}$) $=$ "$\Omega(n)$" [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

SODA 2004: $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

Very intricate combinatorial proof

GVIN 2005: Simplification of proof [Jayram-Kumar-Sivakumar]

Nice geometric intuition + reduction from $\textsc{index}$

CCC 2009: $R^k_{\max}(\textsc{ghd}) = n/2^{O(k^2)}$ [Brody-C.'09]

Round elimination, combinatorial proof

Plus, direct combinatorial 1-round proof

History

FOCS 2003: $\mathrm{R}^{\to}(\text{not-quite-}\textsc{ghd}) =$ "$\Omega(n)$" [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

SODA 2004: $\mathrm{R}^{\to}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

Very intricate combinatorial proof

GVIN 2005: Simplification of proof [Jayram-Kumar-Sivakumar]

Nice geometric intuition + reduction from INDEX

CCC 2009: $R^k_{\max}(\textsc{ghd}) = n/2^{O(k^2)}$ [Brody-C.'09]

Round elimination, combinatorial proof

Plus, direct combinatorial 1-round proof

ICDT 2009: $\mathrm{D}^{\to}_{\mathrm{unif}}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

History

FOCS 2003: $\mathrm{R}^{\rightarrow}$(not-quite-$\textsc{ghd}$) $=$ "$\Omega(n)$" [Indyk-Woodruff]

Messy problem; gave $\Omega(\varepsilon^{-2})$ streaming bound for limited ε

SODA 2004: $\mathrm{R}^{\rightarrow}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

Very intricate combinatorial proof

GVIN 2005: Simplification of proof [Jayram-Kumar-Sivakumar]

Nice geometric intuition + reduction from $\textsc{index}$

CCC 2009: $R^k_{\max}(\textsc{ghd}) = n/2^{O(k^2)}$ [Brody-C.'09]

Round elimination, combinatorial proof

Plus, direct combinatorial 1-round proof

ICDT 2009: $\mathrm{D}^{\rightarrow}_{\mathrm{unif}}(\textsc{ghd}) = \Omega(n)$ [Woodruff]

RND 2010: $R^k_{\max}(\textsc{ghd}) = \widetilde{\Omega}(n/k^2)$ [Brody-C.-Regev-Vidick-deWolf]

Better round elimination, geometric proof

Main Theorem

And now, we show:

$$\mathrm{R}(\textsc{ghd}) = \Omega(n)$$

GHD Revisited

For $x, y \in \{0,1\}^n$, define

$$\mathrm{bias}(x,y) \;=\; \frac{n/2 - \Delta(x,y)}{\sqrt{n}}$$

Then,

$$\textsc{ghd}(x,y) \;=\; \begin{cases} 0\,, & \text{if } \mathrm{bias}(x,y) > 1\,, \\ 1\,, & \text{if } \mathrm{bias}(x,y) < -1\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

Alternative view (useful later): map $b \in \{0,1\} \longmapsto (-1)^b/\sqrt{n}$

This maps $x \in \{0,1\}^n$ into $\widetilde{x} \in \mathbb{S}^{n-1}$ (unit sphere in $\mathbb{R}^n$)

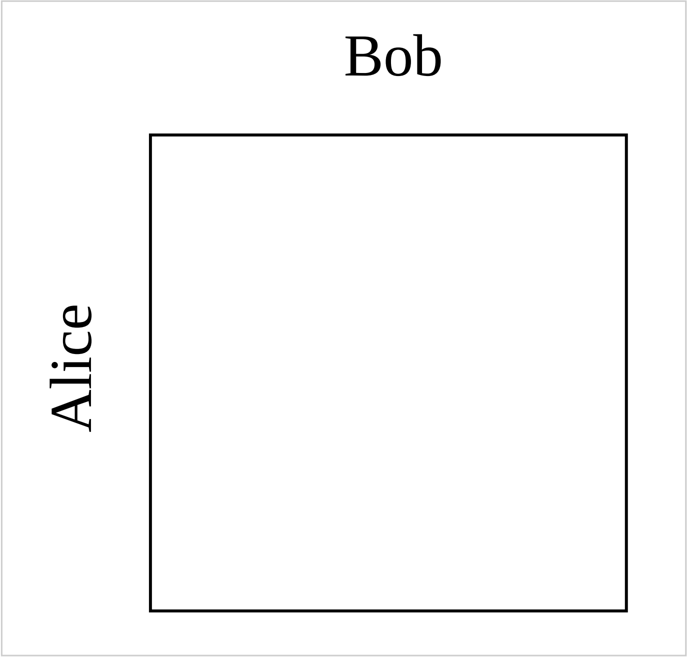
$$\mathrm{bias}(x,y) \;=\; \langle \widetilde{x}, \widetilde{y} \rangle \cdot \sqrt{n}/2$$

The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles
(sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)

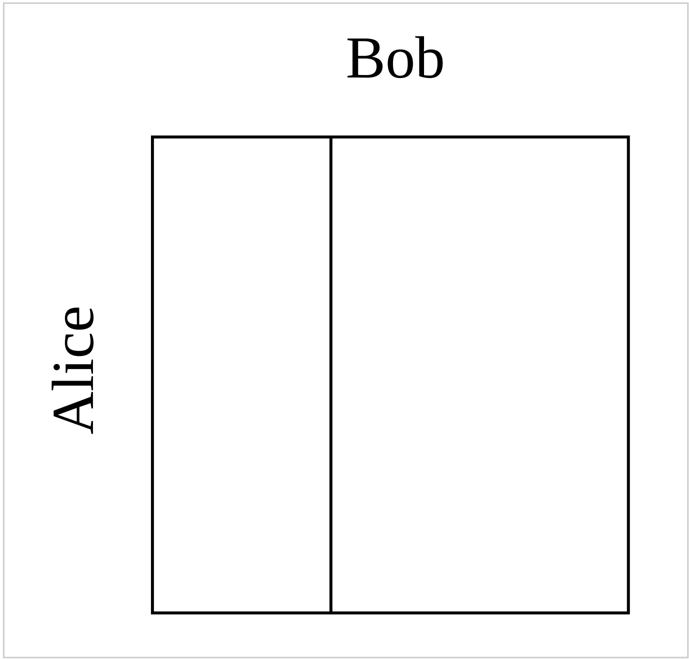


The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles
(sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)

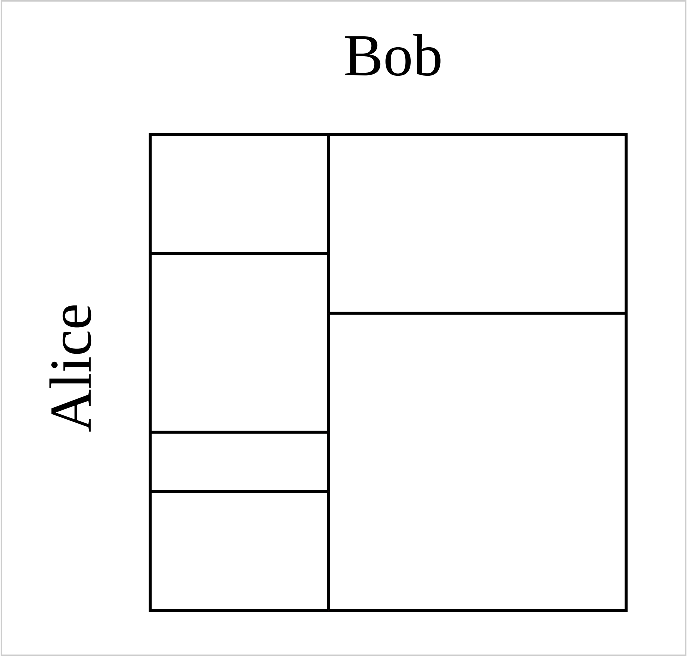


The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles (sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)

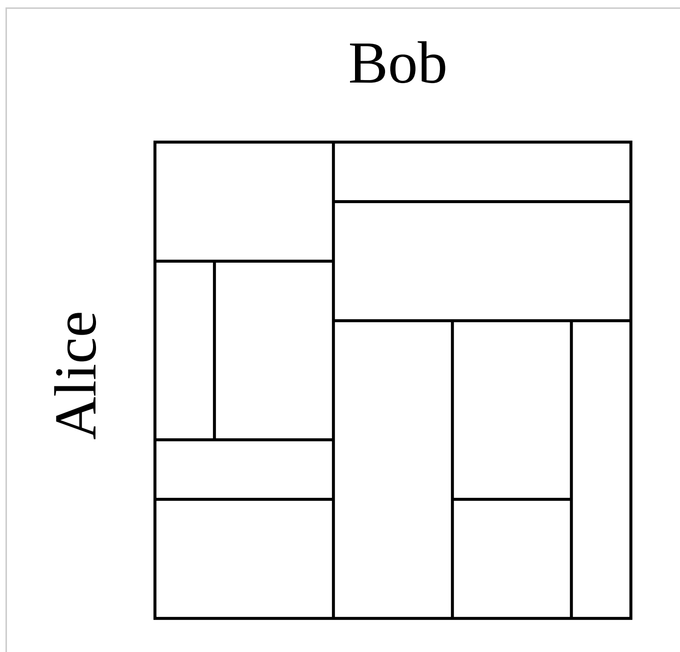


The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles
(sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)

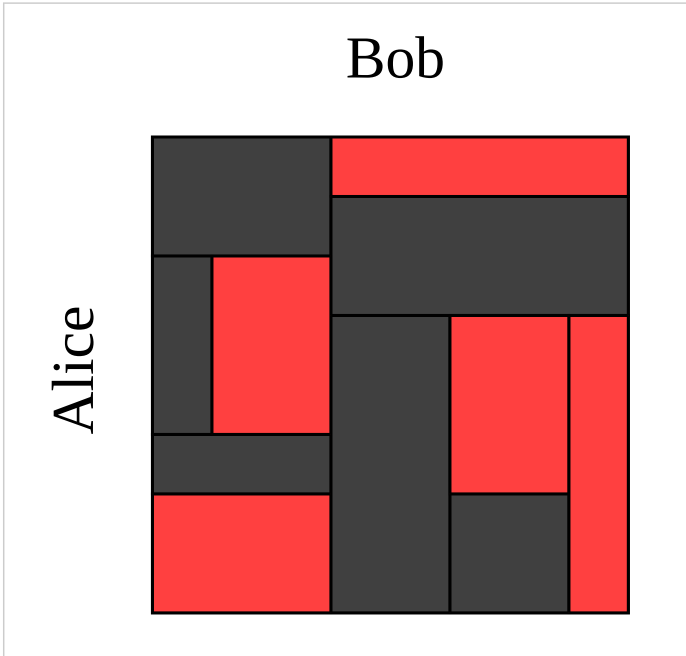


The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles
(sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)

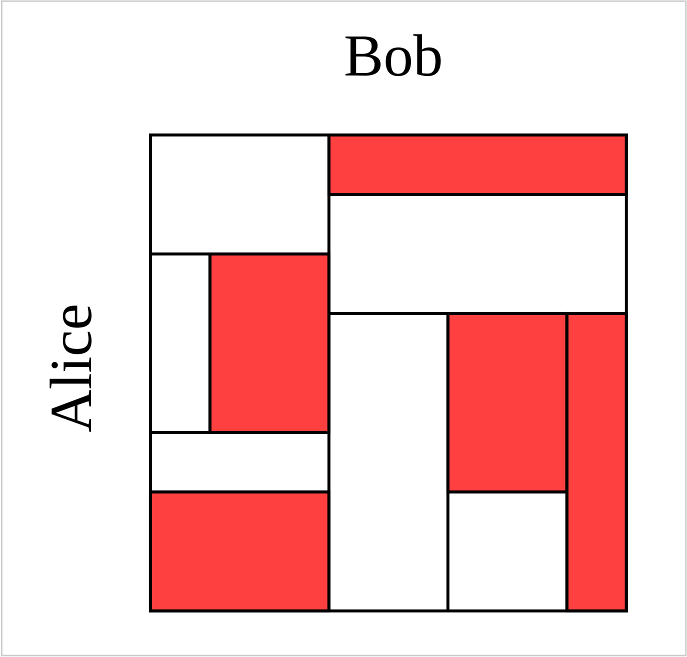


The Rectangle Property

Let $U = \{0,1\}^n \times \{0,1\}^n$ (input universe for Alice + Bob)

Take P deterministic protocol, communicating $\leq c$ bits

Then P partitions U into $\leq 2^c$ combinatorial rectangles
(sets $A \times B$, where $A, B \subseteq \{0,1\}^n$)



If P computes $f : U \to \{0,1\}$, then $f^{-1}(0) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

Discrepancy and Corruption

We had: $f^{-1}(1) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

If P is a correct protocol, matrix of f contains 0-rectangle of size $\geq 2^{2n-c}$

Basic method for lower bounding $\mathrm{D}(f)$:
Show that f does *not* contain large 0-rectangle

To lower bound $\mathrm{R}(f)$, apply Yao's minimax principle

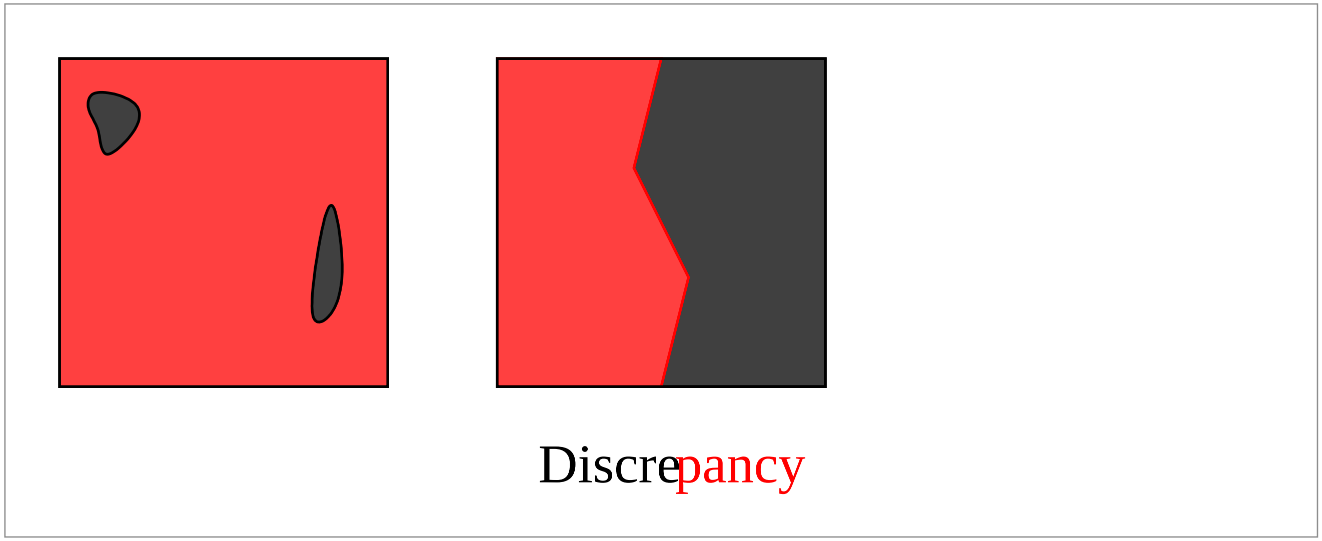
Discrepancy and Corruption

We had: $f^{-1}(1) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

If P is a correct protocol, matrix of f contains 0-rectangle of size $\geq 2^{2n-c}$

Basic method for lower bounding $\mathrm{D}(f)$:
Show that f does *not* contain large 0-rectangle

To lower bound $\mathrm{R}(f)$, apply Yao's minimax principle



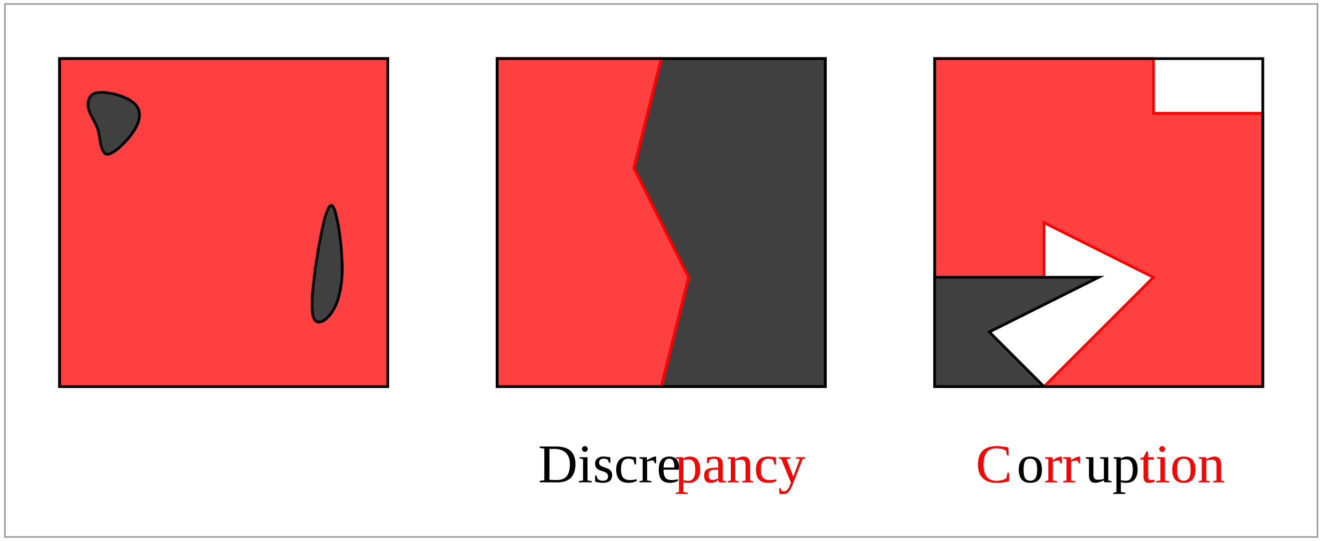
Discrepancy and Corruption

We had: $f^{-1}(1) = R_1 \cup R_2 \cup \cdots \cup R_{2^c}$

If P is a correct protocol, matrix of f contains 0-rectangle of size $\geq 2^{2n-c}$

Basic method for lower bounding $\mathrm{D}(f)$:
Show that f does *not* contain large 0-rectangle

To lower bound $\mathrm{R}(f)$, apply Yao's minimax principle



The Trouble with Corruption for GHD

There exist very large "uncorrupted" rectangles!

Consider:

$$A = B = \{0^{100\sqrt{n}}x : x \in \{0,1\}^{n-100\sqrt{n}}\}$$

Then, $A \times B$ has size $2^{2n-200\sqrt{n}}$ and is essentially a 0-rectangle!

The Trouble with Corruption for GHD

There exist very large "uncorrupted" rectangles!

Consider:

$$A = B = \{0^{100\sqrt{n}} x : x \in \{0,1\}^{n-100\sqrt{n}}\}$$

Then, $A \times B$ has size $2^{2n-200\sqrt{n}}$ and is essentially a 0-rectangle!

$$\Pr_{(x,y)\in_R A\times B}[\mathrm{GHD}(x,y) = 0] \;=\; \Pr_{(x,y)\in_R A\times B}[\mathrm{bias}(x,y) > 1] \;=\; 1 - 2^{-\Omega(n)}$$

Need a new technique?

The Corruption Method: A Closer Look

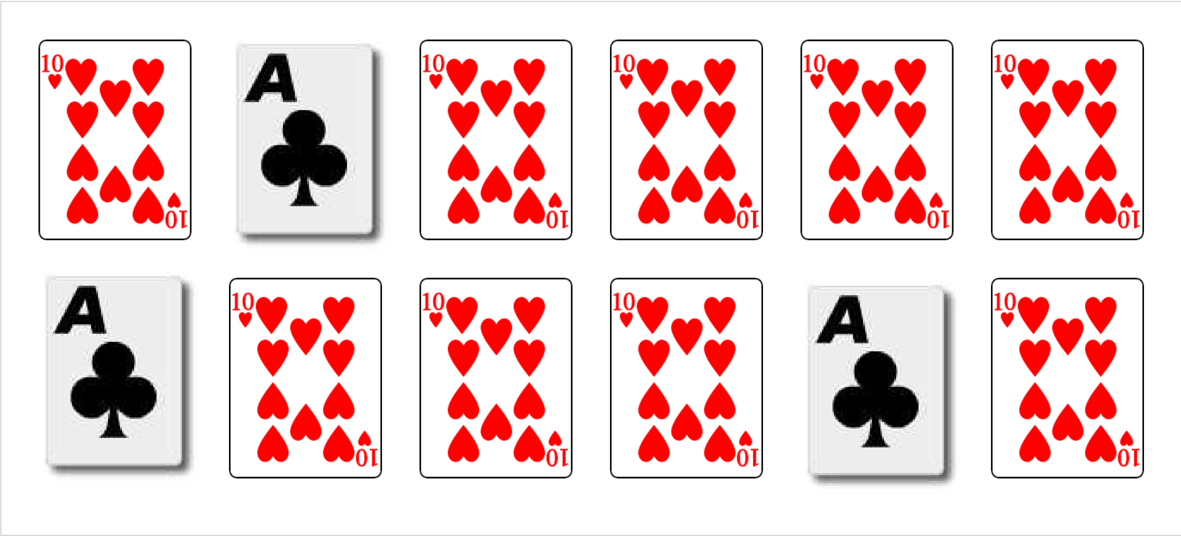
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$

Argue that for all large rectangles R, we have

$$\mu_1(R) \;\geq\; \alpha\,\mu_0(R)$$

Sum this over all 0-rectangles R; if protocol P is good for μ_0, μ_1:

$$\mu_1\big(\{P \text{ outputs } 0\}\big) \;\geq\; \alpha \cdot \mu_0\big(\{P \text{ outputs } 0\}\big)$$



The Corruption Method: A Closer Look

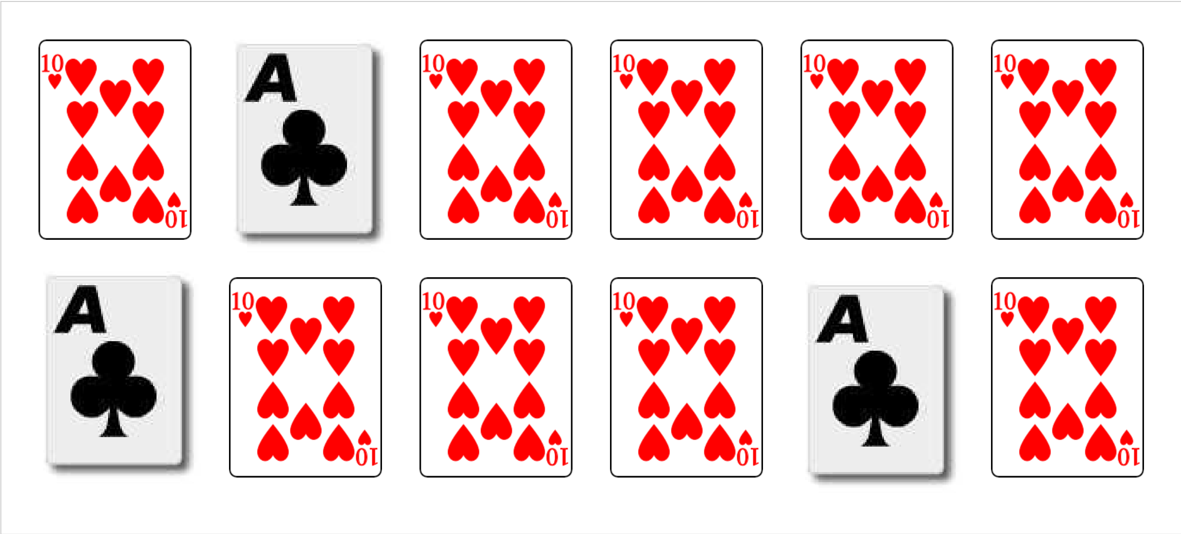
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$

Argue that for all large rectangles R, we have

$$\mu_1(R) \;\geq\; \alpha\, \mu_0(R)$$

Sum this over all 0-rectangles R; if protocol P is good for μ_0, μ_1:

$$\mu_1\big(\{P \text{ outputs } 0\}\big) \;\geq\; \alpha \cdot \mu_0\big(\{P \text{ outputs } 0\}\big) \;\geq\; \alpha(1-\varepsilon)$$



The Corruption Method: A Closer Look

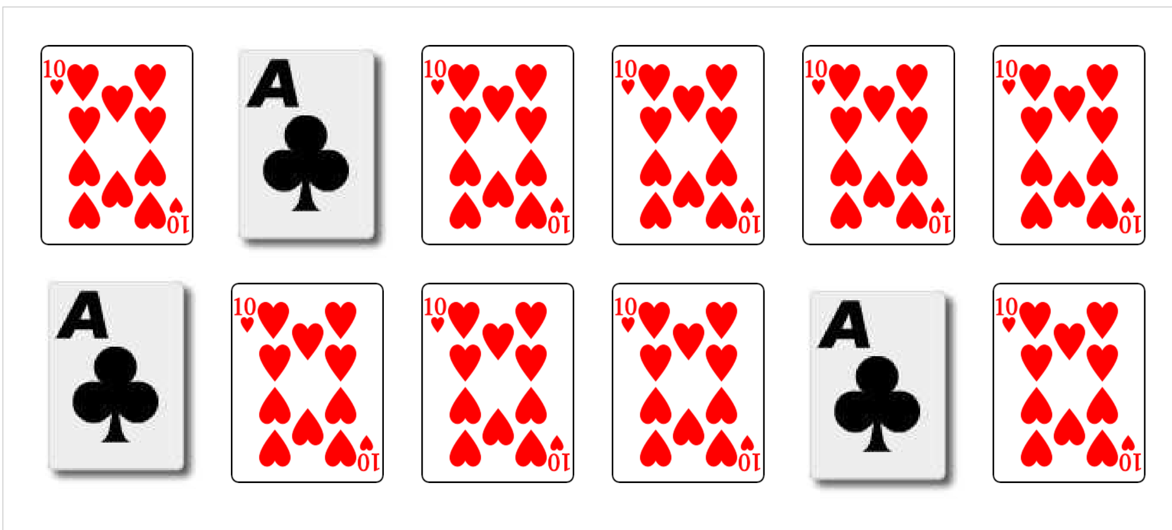
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$

Argue that for all large rectangles R, we have

$$\mu_1(R) \;\geq\; \alpha\, \mu_0(R)$$

Sum this over all 0-rectangles R; if protocol P is good for μ_0, μ_1:

$$\varepsilon \;\geq\; \mu_1\big(\{P \text{ outputs } 0\}\big) \;\geq\; \alpha \cdot \mu_0\big(\{P \text{ outputs } 0\}\big) \;\geq\; \alpha(1-\varepsilon)$$



Jokers

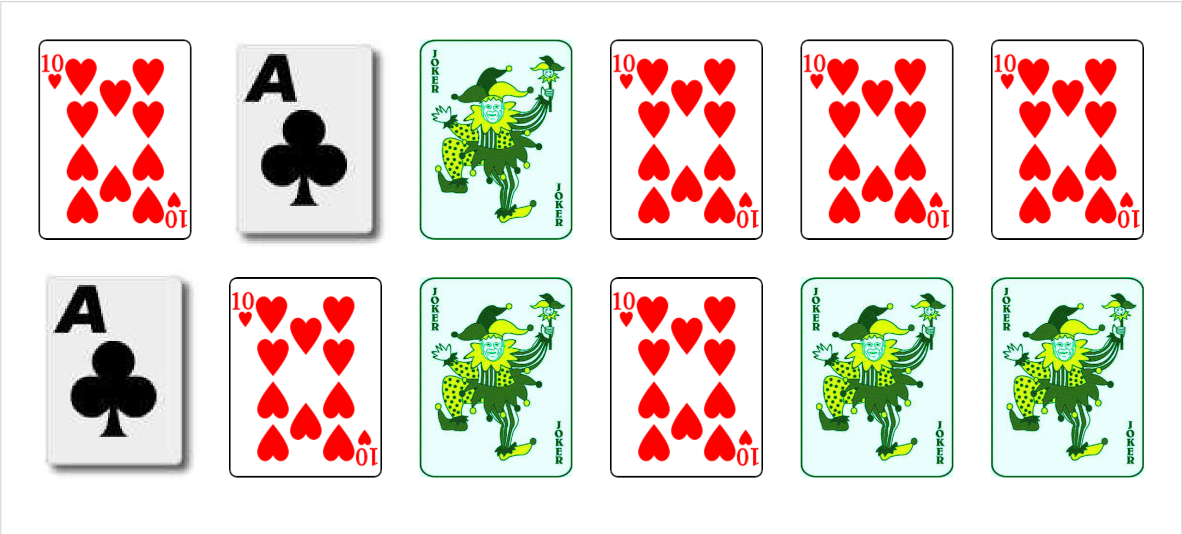
Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\,\mu_\star(R) \;\geq\; \alpha\,\mu_0(R) \qquad (\alpha > \beta)$$

Sum this over all 0-rectangles R; if protocol P is good for μ_0, μ_1:

$$\mu_1(P_0) + \beta\,\mu_\star(P_0) \;\geq\; \alpha\,\mu_0(P_0) \;\geq\; \alpha(1 - \varepsilon)$$



Jokers

Pick distribs μ_0, μ_1 on $f^{-1}(0), f^{-1}(1)$, and another distrib $\mu_\star$

Argue that for all large rectangles R, we have

$$\mu_1(R) + \beta\,\mu_\star(R) \;\geq\; \alpha\,\mu_0(R) \qquad\qquad (\alpha > \beta)$$

Sum this over all 0-rectangles R; if protocol P is good for μ_0, μ_1:

$$\varepsilon + \beta \;\geq\; \mu_1(P_0) + \beta\,\mu_\star(P_0) \;\geq\; \alpha\,\mu_0(P_0) \;\geq\; \alpha(1-\varepsilon)$$

The Distributions: Zeroes, Ones, Jokers

Consider slightly "shifted" version (doesn't really change anything)

$$\textsc{ghd}'(x,y) \;=\; \begin{cases} 0\,, & \text{if } \mathrm{bias}(x,y) > -4\,, \\ 1\,, & \text{if } \mathrm{bias}(x,y) < -6\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

The Distributions: Zeroes, Ones, Jokers

Consider slightly "shifted" version (doesn't really change anything)

$$\textsc{ghd}'(x,y) \;=\; \begin{cases} 0\,, & \text{if } \mathrm{bias}(x,y) > -4\,, \\ 1\,, & \text{if } \mathrm{bias}(x,y) < -6\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

Let

$$\begin{aligned} \mu_0 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 0\} \\ \mu_1 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = -10\} \\ \mu_\star &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 10\} \end{aligned}$$

The Distributions: Zeroes, Ones, Jokers

Consider slightly "shifted" version (doesn't really change anything)

$$\text{GHD}'(x,y) \;=\; \begin{cases} 0\,, & \text{if } \mathrm{bias}(x,y) > -4\,, \\ 1\,, & \text{if } \mathrm{bias}(x,y) < -6\,, \\ \star\,, & \text{otherwise.} \end{cases}$$

Let

$$\begin{aligned} \mu_0 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 0\} \\ \mu_1 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = -10\} \\ \mu_\star &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 10\} \end{aligned}$$

The Key Inequality: For rectangles R of size $\geq 2^{2n-0.01n}$

$$\boxed{\tfrac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\geq\; 0.9\mu_0(R)}$$

Interpretation: An Anti-Concentration Inequality

$$\begin{aligned}
\mu_0 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 0\} \\
\mu_1 &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = -10\} \\
\mu_\star &= \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 10\}
\end{aligned}$$

Key Inequality: $|R| \geq 2^{1.99n} \Longrightarrow \boxed{\frac{1}{2}(\mu_1(R) + \mu_\star(R)) \ \geq \ 0.9\mu_0(R)}$

Distrib of biases in large rectangle can't be too concentrated around zero

Interpretation: An Anti-Concentration Inequality

$$\mu_0 \;=\; \text{Uniform on } \{(x,y) : \text{bias}(x,y) = 0\}$$
$$\mu_1 \;=\; \text{Uniform on } \{(x,y) : \text{bias}(x,y) = -10\}$$
$$\mu_\star \;=\; \text{Uniform on } \{(x,y) : \text{bias}(x,y) = 10\}$$

Key Inequality: $|R| \geq 2^{1.99n} \Longrightarrow \boxed{\frac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\geq\; 0.9\mu_0(R)}$

Distrib of biases in large rectangle can't be too concentrated around zero

Rectangularity crucial:

$$S \;=\; \{(x,y) \in U : \text{bias}(x,y) = 0\} \quad \text{has} \quad |S| \approx 2^{2n}/\sqrt{n}$$

Interpretation: An Anti-Concentration Inequality

$$\mu_0 \;=\; \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 0\}$$
$$\mu_1 \;=\; \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = -10\}$$
$$\mu_\star \;=\; \text{Uniform on } \{(x,y) : \mathrm{bias}(x,y) = 10\}$$

Key Inequality: $|R| \ge 2^{1.99n} \Longrightarrow \boxed{\frac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\ge\; 0.9\mu_0(R)}$

Distrib of biases in large rectangle can't be too concentrated around zero

Rectangularity crucial:

$$S \;=\; \{(x,y) \in U : \mathrm{bias}(x,y) = 0\} \quad \text{has} \quad |S| \approx 2^{2n}/\sqrt{n}$$

Largeness crucial:

$$A \;=\; \{x \in \{0,1\}^{n/2}, |x| = n/4\}; \quad R \;=\; (0^{n/2}\cdot A) \times (A\cdot 0^{n/2})$$

Interpretation: An Anti-Concentration Inequality

$$\mu_0 \;=\; \text{Uniform on } \{(x,y) : \operatorname{bias}(x,y) = 0\}$$
$$\mu_1 \;=\; \text{Uniform on } \{(x,y) : \operatorname{bias}(x,y) = -10\}$$
$$\mu_\star \;=\; \text{Uniform on } \{(x,y) : \operatorname{bias}(x,y) = 10\}$$

Key Inequality: $|R| \geq 2^{1.99n} \Longrightarrow \boxed{\tfrac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\geq\; 0.9\mu_0(R)}$

Distrib of biases in large rectangle can't be too concentrated around zero

Rectangularity crucial:

$$S \;=\; \{(x,y) \in U : \operatorname{bias}(x,y) = 0\} \quad \text{has} \quad |S| \approx 2^{2n}/\sqrt{n}$$

Largeness crucial:

$$A \;=\; \{x \in \{0,1\}^{n/2}, |x| = n/4\}; \quad R \;=\; (0^{n/2}\cdot A) \times (A \cdot 0^{n/2})$$
$$\text{then} \quad \forall (x,y) \in R : \operatorname{bias}(x,y) = 0 \quad \text{and} \quad |R| \approx 2^n/\sqrt{n}$$

Sleight of Hand?

Did we pull a new (joker) distribution out of a hat?

Do these jokers have any "meaning"?

Sleight of Hand?

Did we pull a new (joker) distribution out of a hat?

Do these jokers have any "meaning"?

- Yes! What we did here can be understood more deeply by studying a linear program (and its dual)
- Careful study of this type of generalization: "smooth rectangle bound" and "partition bound"

[Klauck'10] [Jain-Klauck'10]

The Inequality: A Gaussian Version

Original inequality: $|R| \geq 2^{1.99n} \implies \frac{1}{2}(\mu_1(R) + \mu_\star(R)) \geq 0.9\mu_0(R)$

Apply map from $\{0,1\}^n$ to unit sphere $\mathbb{S}^{n-1}$

Let $\gamma = n$-dimensional Gaussian distrib

Analogous inequality:

$\gamma(A), \gamma(B) \geq 2^{-n/100},\ x \leftarrow A,\ y \leftarrow B \implies$
distrib of $\langle x, y\rangle/\sqrt{n}$ is "spread out" like $N(0,1)$

The Inequality: A Gaussian Version

Original inequality: $|R| \geq 2^{1.99n} \implies \frac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\geq\; 0.9\mu_0(R)$

Apply map from $\{0,1\}^n$ to unit sphere $\mathbb{S}^{n-1}$

Let $\gamma = n$-dimensional Gaussian distrib

Analogous inequality:
$\gamma(A), \gamma(B) \geq 2^{-n/100},\; x \leftarrow A,\; y \leftarrow B \implies$
distrib of $\langle x, y\rangle/\sqrt{n}$ is "spread out" like $N(0,1)$

[Can't just *fix* a direction $x \in A$: what if $\mathrm{proj}(B, x)$ sharply concentrated?]

The Inequality: A Gaussian Version

Original inequality: $|R| \geq 2^{1.99n} \implies \frac{1}{2}(\mu_1(R) + \mu_\star(R)) \;\geq\; 0.9\mu_0(R)$

Apply map from $\{0,1\}^n$ to unit sphere $\mathbb{S}^{n-1}$

Let $\gamma = n$-dimensional Gaussian distrib

Analogous inequality:

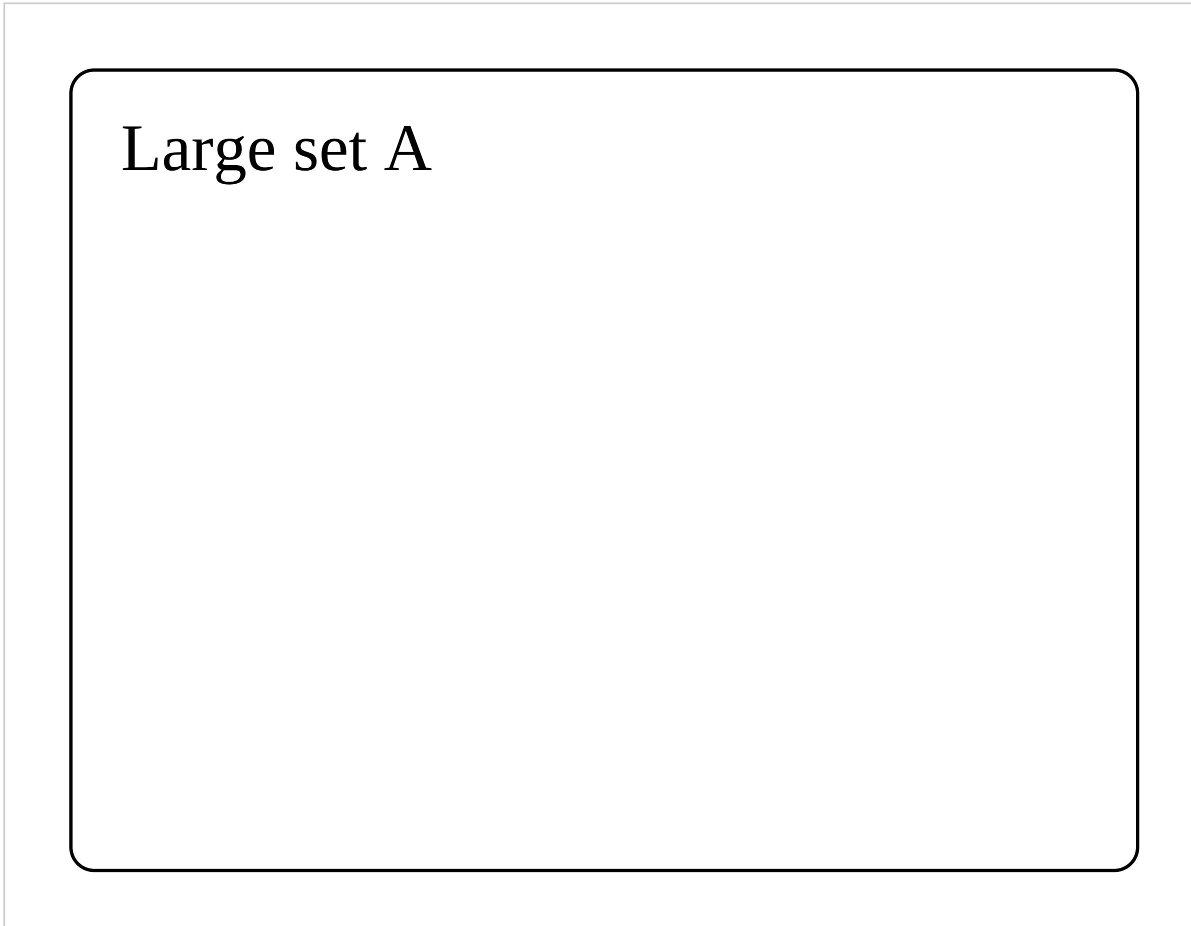
$\gamma(A), \gamma(B) \geq 2^{-n/100},\; x \leftarrow A,\; y \leftarrow B \implies$
distrib of $\langle x, y \rangle / \sqrt{n}$ is "spread out" like $N(0,1)$

[Can't just *fix* a direction $x \in A$: what if $\mathrm{proj}(B, x)$ sharply concentrated?]

A Stronger Statement

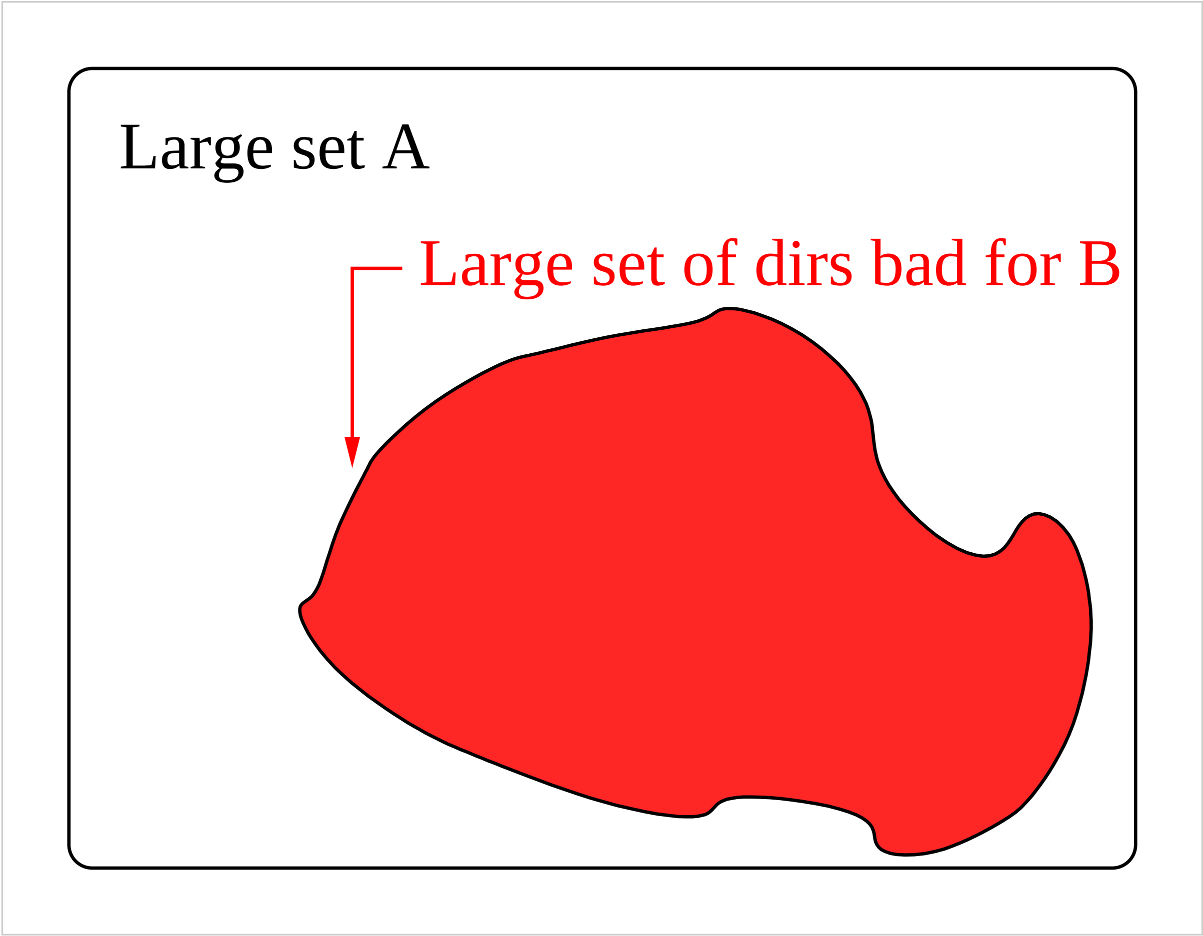
$\gamma(B) \geq 2^{-n/100} \implies$ projection of B on all but $2^{-n/50}$ of directions distributed like $N(0,1) + Z$ (i.e., mixture of normals with variance 1)

Proof Overview



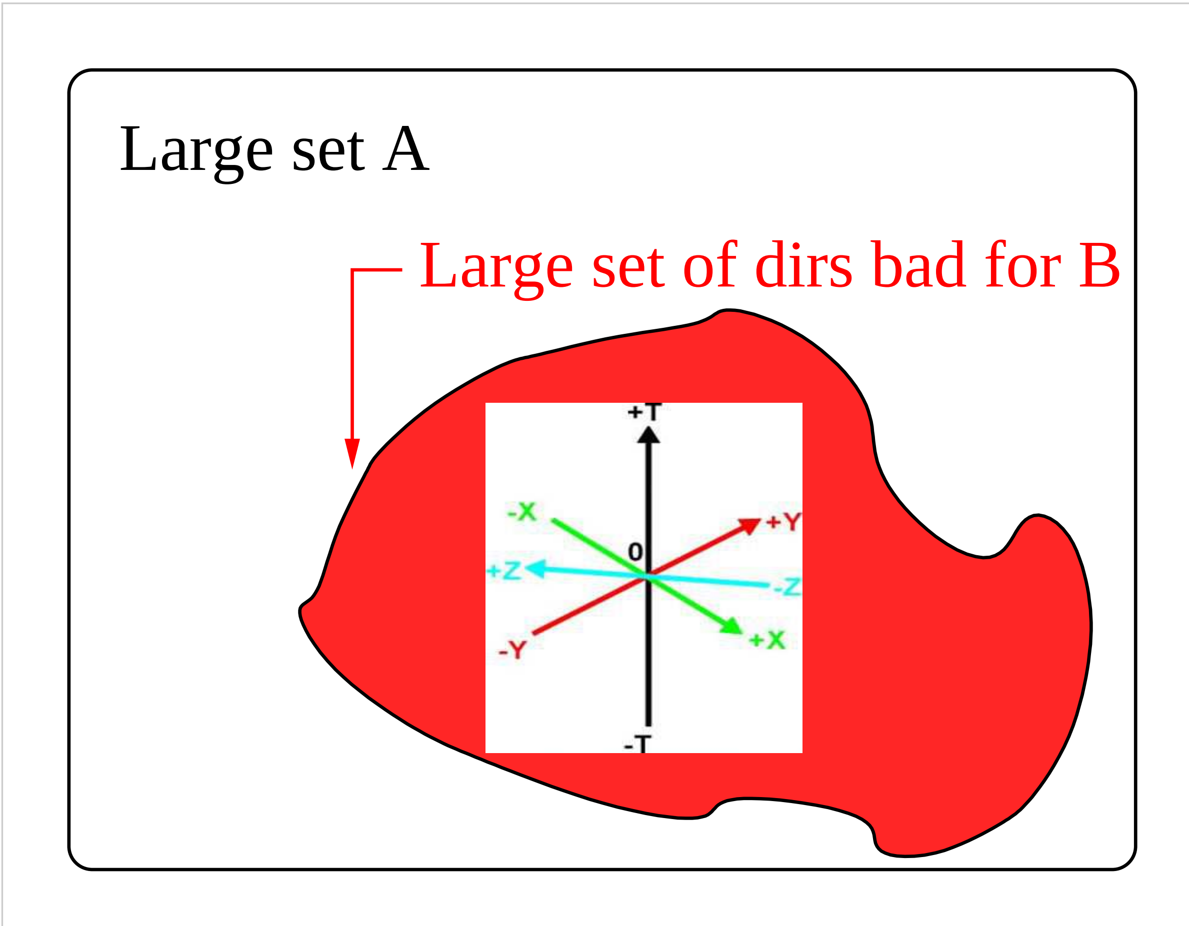
Large set of bad dirs $\supseteq$ many orthogonal bad dirs

Proof Overview



Large set of bad dirs $\supseteq$ many orthogonal bad dirs

Proof Overview



Large set of bad dirs $\supseteq$ many orthogonal bad dirs

Finding Orthogonal Bad Directions

Want to show that A doesn't have many bad directions

We'll show: if it does, then $\exists$ many *nearly orthogonal* bad directions

Finding Orthogonal Bad Directions

Want to show that A doesn't have many bad directions

We'll show: if it does, then $\exists$ many *nearly orthogonal* bad directions

A lemma from Raz: [Raz'99]

> Any set $A' \subseteq \mathbb{S}^{n-1}$ of at least $2^{-n/50}$ directions contains a set of $\frac{1}{10}$-near-orthogonal vectors $x_1, \ldots, x_{n/2}$, i.e.,
>
> $$\| \operatorname{proj}(x_i, \operatorname{span}(x_1, \ldots, x_{i-1})) \| \ \leq \ 1/10$$

Proof via isoperimetric inequality

Can't Have Orthogonal Bad Directions

Lemma 1: Suppose $B \subseteq \mathbb{R}^n$ is s.t. $\gamma(B) \geq 2^{-n/100}$. Let $y \leftarrow B$. Let directions $x_1, \ldots, x_{n/2}$ be orthogonal. Then all of $\langle y, x_1 \rangle, \ldots, \langle y, x_{n/2} \rangle$ cannot be sharply concentrated.

Precise statement: At least one (in fact, most) projection $\langle y, x_k \rangle$ is close to $N(0, 1)$ (even when conditioned on $\langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle$)

Can't Have Orthogonal Bad Directions

Lemma 1: Suppose $B \subseteq \mathbb{R}^n$ is s.t. $\gamma(B) \geq 2^{-n/100}$. Let $y \leftarrow B$. Let directions $x_1, \ldots, x_{n/2}$ be orthogonal. Then all of $\langle y, x_1 \rangle, \ldots, \langle y, x_{n/2} \rangle$ cannot be sharply concentrated.

Precise statement: At least one (in fact, most) projection $\langle y, x_k \rangle$ is close to $N(0,1)$ (even when conditioned on $\langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle$)

Proof Idea: Complete to orthonormal basis: $\{x_1, \ldots, x_n\}$

Then y is determined by $\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle$. Wave hands as follows:

$$0.99n \;\leq\; \mathrm{H}(y) \;\leq\; \mathrm{H}(\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle)$$

Can't Have Orthogonal Bad Directions

Lemma 1: Suppose $B \subseteq \mathbb{R}^n$ is s.t. $\gamma(B) \geq 2^{-n/100}$. Let $y \leftarrow B$. Let directions $x_1, \ldots, x_{n/2}$ be orthogonal. Then all of $\langle y, x_1 \rangle, \ldots, \langle y, x_{n/2} \rangle$ cannot be sharply concentrated.

Precise statement: At least one (in fact, most) projection $\langle y, x_k \rangle$ is close to $N(0,1)$ (even when conditioned on $\langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle$)

Proof Idea: Complete to orthonormal basis: $\{x_1, \ldots, x_n\}$

Then y is determined by $\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle$. Wave hands as follows:

$$\begin{aligned}
0.99n \;\leq\; \mathrm{H}(y) \;&\leq\; \mathrm{H}(\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle) \\
&= \textstyle\sum_{k=1}^{n/2} \mathrm{H}(\langle y, x_k \rangle \mid \langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle) \\
&\quad + \textstyle\sum_{k=n/2+1}^{n} \mathrm{H}(\langle y, x_k \rangle \mid \langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle)
\end{aligned}$$

Can't Have Orthogonal Bad Directions

Lemma 1: Suppose $B \subseteq \mathbb{R}^n$ is s.t. $\gamma(B) \geq 2^{-n/100}$. Let $y \leftarrow B$. Let directions $x_1, \ldots, x_{n/2}$ be orthogonal. Then all of $\langle y, x_1 \rangle, \ldots, \langle y, x_{n/2} \rangle$ cannot be sharply concentrated.

Precise statement: At least one (in fact, most) projection $\langle y, x_k \rangle$ is close to $N(0,1)$ (even when conditioned on $\langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle$)

Proof Idea: Complete to orthonormal basis: $\{x_1, \ldots, x_n\}$

Then y is determined by $\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle$. Wave hands as follows:

$$\begin{aligned}
0.99n \;\leq\; \mathrm{H}(y) \;&\leq\; \mathrm{H}(\langle y, x_1 \rangle, \ldots, \langle y, x_n \rangle) \\
&= \textstyle\sum_{k=1}^{n/2} \mathrm{H}(\langle y, x_k \rangle \mid \langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle) \\
&\quad + \textstyle\sum_{k=n/2+1}^{n} \mathrm{H}(\langle y, x_k \rangle \mid \langle y, x_1 \rangle, \ldots, \langle y, x_{k-1} \rangle) \\
&\leq \textstyle\sum_{k=1}^{n/2} 0.7 + \sum_{k=n/2+1}^{n} 1 \;=\; 0.85n
\end{aligned}$$

Finishing the Proof

Theorem: $\gamma(B) \geq 2^{-n/100} \implies$ projection of B on all but $2^{-n/50}$ of directions distributed like $N(0,1) + Z$

Proof Sketch:

- Let $A' = \{\text{bad directions}\}$; suppose to the contrary that its measure is $\geq 2^{-n/50}$
- Get near-orthogonal bad dirs $x_1, \ldots, x_{n/2} \in A'$ by Raz's Lemma
- If these vectors were orthogonal, by Lemma 1, $\exists\, k$ s.t. $\langle B, x_k \rangle$ is close to $N(0,1)$. So x_k is not bad. Contradiction.

Finishing the Proof

Theorem: $\gamma(B) \geq 2^{-n/100} \implies$ projection of B on all but $2^{-n/50}$ of directions distributed like $N(0,1) + Z$

Proof Sketch:

- Let $A' = \{\text{bad directions}\}$; suppose to the contrary that its measure is $\geq 2^{-n/50}$
- Get near-orthogonal bad dirs $x_1, \ldots, x_{n/2} \in A'$ by Raz's Lemma
- If these vectors were orthogonal, by Lemma 1, $\exists\, k$ s.t. $\langle B, x_k \rangle$ is close to $N(0,1)$. So x_k is not bad. Contradiction.
- Since they are only $\frac{1}{10}$-near-orthogonal, we instead get that $\langle B, x_k \rangle$ is distributed like $N(0,1) + Z$. Still a contradiction.

Conclusions

- Settled communication complexity of GHD, proving a long-conjectured $\Omega(n)$ bound
- As a result, understood multi-pass space complexity of a number of data stream problems, including frequency moments

Conclusions

- Settled communication complexity of GHD, proving a long-conjectured $\Omega(n)$ bound
- As a result, understood multi-pass space complexity of a number of data stream problems, including frequency moments

Open Problem

Apply the "jokers" idea (more generally, the smooth rectangle bound) to other interesting communication and query complexity problems.