Linear sketching with parities

Grigory Yaroslavtsev

http://grigory.us

with Sampath Kannan (Penn) and Elchanan Mossel (MIT)

Linear sketching with parities

- Input $\boldsymbol{x} \in \{0,1\}^n$
- Parity = Linear function over $\mathbb{G}F_2$: $\bigoplus_{i \in S} x_i$
- E.g. $x_4 \oplus x_2 \oplus x_{42}$
- **Deterministic linear sketch**: set of $\boldsymbol{k}$ parities:

$$\ell(\boldsymbol{x}) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \ldots; \bigoplus_{i_{\boldsymbol{k}} \in S_{\boldsymbol{k}}} x_{i_{\boldsymbol{k}}}$$

- **Randomized linear sketch**: **distribution** over $\boldsymbol{k}$ parities (random $S_1, S_2, \ldots, S_{\boldsymbol{k}}$):

$$\ell(\boldsymbol{x}) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \ldots; \bigoplus_{i_{\boldsymbol{k}} \in S_{\boldsymbol{k}}} x_{i_{\boldsymbol{k}}}$$

Linear sketching over $\mathbb{G}F_2$

- Given $\boldsymbol{f}(\boldsymbol{x}): \{0,1\}^n \to \{0,1\}$
- **Question:**

Can one compute $\boldsymbol{f}(\boldsymbol{x})$ from a small ($\boldsymbol{k} \ll n$) linear sketch over $\mathbb{G}F_2$?

- Allow randomized computation (99% success)

Motivation: Distributed Computing

- **Distributed computation among M machines:**
  - $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_M)$ (more generally $\boldsymbol{x} = \bigoplus_{i=1}^{M} \boldsymbol{x}_i$)
  - M machines can compute sketches locally:

    $$\ell(\boldsymbol{x}_1), \dots, \ell(\boldsymbol{x}_M)$$

  - Send them to the coordinator who computes:

    $\ell_i(\boldsymbol{x}) = \ell_i(\boldsymbol{x}_1) \oplus \cdots \oplus \ell_i(\boldsymbol{x}_M)$ (coordinate-wise XORs)
  - Coordinator computes $\boldsymbol{f}(\boldsymbol{x})$ with $\boldsymbol{kM}$ communication

$\boldsymbol{x}$	0	1	1	1	1	1	0	0	0	0
			$\boldsymbol{x}_1$					$\boldsymbol{x}_2$		

Motivation: Streaming

- $\boldsymbol{x}$ generated through a sequence of updates
- Updates $i_1, \dots, i_m$: update i_t flips bit at position i_t

$\boldsymbol{x_0}$	0	0	0	[illegible]	[illegible]	[illegible]	0	0	0	0

Updates: (1, 3, 8

$\boldsymbol{x_1}$	1	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
$\boldsymbol{x_2}$	1	0	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0
$\boldsymbol{x_3}$	1	0	[illegible]	[illegible]	[illegible]	[illegible]	0	1	0	0
$\boldsymbol{x}$	1	0	0	[illegible]	[illegible]	[illegible]	0	1	0	0

$\ell(\boldsymbol{x})$ allows to compute $\boldsymbol{f}(\boldsymbol{x})$ with $\boldsymbol{k}$ bits of space

Deterministic vs. Randomized

- **Fact**: $\boldsymbol{f}$ has a deterministic sketch if and only if
  - $\boldsymbol{f} = \boldsymbol{g}(\bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \dots; \bigoplus_{i_k \in S_k} x_{i_k})$
  - Equivalent to "$\boldsymbol{f}$ has Fourier dimension $\boldsymbol{k}$"
- **Randomization can help:**
  - **OR**: $\boldsymbol{f}(\boldsymbol{x}) = x_1 \vee \cdots \vee x_n$
  - Has "Fourier dimension" $= n$
  - Pick $\boldsymbol{t} = \log 1/\boldsymbol{\delta}$ random sets $S_1, \dots, S_{\boldsymbol{t}}$
  - If there is j such that $\bigoplus_{i \in S_j} x_i = 1$ output 1, otherwise output 0
  - Error probability $\boldsymbol{\delta}$

Fourier Analysis

- $\boldsymbol{f}(x_1, \ldots, x_n): \{0,1\}^n \to \{0,1\}$
- Notation switch:
  - $0 \to 1$
  - $1 \to -1$
- $\boldsymbol{f}': \{-1,1\}^n \to \{-1,1\}$
- Functions as vectors form a vector space:

$$\boldsymbol{f}: \{-1,1\}^n \to \{-1,1\} \Leftrightarrow \boldsymbol{f} \in \{-1,1\}^{2^n}$$

- Inner product on functions = "correlation":

$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = 2^{-n} \sum_{x \in \{-1,1\}^n} \boldsymbol{f}(x)\boldsymbol{g}(x) = \mathbb{E}_{x \sim \{-1,1\}^n}[\boldsymbol{f}(x)\boldsymbol{g}(x)]$$

$$||\boldsymbol{f}||_2 = \sqrt{\langle f, f \rangle} = \sqrt{\mathbb{E}_{x \sim \{-1,1\}^n}[\boldsymbol{f}^2(x)]} = 1 \text{ (for Boolean only)}$$

"Main Characters" are Parities

- For $\boldsymbol{S} \subseteq [n]$ let **character** $\chi_{\boldsymbol{S}}(x) = \prod_{i \in \boldsymbol{S}} x_i$
- **Fact**: Every function $\boldsymbol{f}: \{-1,1\}^n \to \{-1,1\}$ **uniquely** represented as multilinear polynomial

$$\boldsymbol{f}(x_1, \ldots, x_n) = \sum_{\boldsymbol{S} \subseteq [n]} \hat{\boldsymbol{f}}(\boldsymbol{S}) \chi_{\boldsymbol{S}}(x)$$

- $\hat{\boldsymbol{f}}(\boldsymbol{S})$ a.k.a. Fourier coefficient of $\boldsymbol{f}$ on $\boldsymbol{S}$
- $\hat{\boldsymbol{f}}(\boldsymbol{S}) \equiv \langle \boldsymbol{f}, \chi_{\boldsymbol{S}} \rangle = \mathbb{E}_{x \sim \{-1,1\}^n}[\boldsymbol{f}(x)\boldsymbol{g}(x)]$
- $\sum_{\boldsymbol{S}} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 = 1$ (Parseval)

Fourier Dimension

- Fourier sets S ≡ vectors in $\mathbb{G}F_2^n$
- "$\boldsymbol{f}$ has Fourier dimension $\boldsymbol{k}$" = a $\boldsymbol{k}$-dimensional subspace in Fourier domain has all weight

$$\sum_{\boldsymbol{S} \subseteq A_{\boldsymbol{k}}} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 = 1$$

$$\boldsymbol{f}(x_1, \dots, x_n) = \sum_{\boldsymbol{S} \subseteq [n]} \hat{\boldsymbol{f}}(\boldsymbol{S}) \boldsymbol{\chi}_{\boldsymbol{S}}(\boldsymbol{x}) = \sum_{\boldsymbol{S} \subseteq A_{\boldsymbol{k}}} \hat{\boldsymbol{f}}(\boldsymbol{S}) \boldsymbol{\chi}_{\boldsymbol{S}}(\boldsymbol{x})$$

- Pick a basis $\boldsymbol{S_1}, \dots, \boldsymbol{S_k}$ in $A_{\boldsymbol{k}}$:
  - Sketch: $\boldsymbol{\chi}_{\boldsymbol{S_1}}(\boldsymbol{x}), \dots, \boldsymbol{\chi}_{\boldsymbol{S_k}}(\boldsymbol{x})$
  - For every $\boldsymbol{S} \in A_{\boldsymbol{k}}$ there exists $\boldsymbol{Z} \subseteq [\boldsymbol{k}]$: $\boldsymbol{S} = \bigoplus_{\boldsymbol{i} \in \boldsymbol{Z}} \boldsymbol{S_i}$

$$\boldsymbol{\chi}_{\boldsymbol{S}}(\boldsymbol{x}) = \bigoplus_{\boldsymbol{i} \in \boldsymbol{Z}} \boldsymbol{\chi}_{\boldsymbol{S_i}}(\boldsymbol{x})$$

Deterministic Sketching and Noise

Suppose "noise" has a bounded norm

$$\boldsymbol{f} = \boldsymbol{k}\text{-dim.} + \text{noise}$$

- L_0-noise in the Fourier domain (via [Sanyal'15])
  - $\hat{\boldsymbol{f}} = \boldsymbol{k}$-dim. + "Fourier L_0-noise"
  - Linear sketch size: $\boldsymbol{k} + O(||\widehat{noise}||_0^{1/2})$
  - **Our work**: can't be improved even with randomness, e.g for ``addressing function''.

How Randomization Handles Noise

- L_0-noise in the original domain (hashing a la OR)
  - $\boldsymbol{f} = \boldsymbol{k}$-dim. + “$L_0$-noise”
  - Linear sketch size: $\boldsymbol{k}$ + O(log $||noise||_0$)
  - Optimal (but only existentially, i.e. $\exists \boldsymbol{f}$: ...)
- L_1-noise in the Fourier domain (via [Grolmusz’97])
  - $\hat{\boldsymbol{f}} = \boldsymbol{k}$-dim. + “Fourier L_1-noise”
  - Linear sketch size: $\boldsymbol{k} + O(||\widehat{noise}||_1^2)$
  - Example = $\boldsymbol{k}$-dim. + small decision tree / DNF / etc.

Randomized Sketching: Hardness

- $\boldsymbol{k}$ -dimensional **affine extractors** require $\boldsymbol{k}$:
  - $\boldsymbol{f}$ is an **affine-extractor** for dim. $\boldsymbol{k}$ if any restriction on a $\boldsymbol{k}$-dim. affine subspace has values 0/1 w/prob. ≥ 0.1 each
  - Example (inner product): $\boldsymbol{f}(\boldsymbol{x}) = \bigoplus_{i=1}^{n/2} x_{2i-1}x_{2i}$
- Not $\boldsymbol{\gamma}$-concentrated on $\boldsymbol{k}$ -dim. Fourier subspaces
  - For $\forall$ $\boldsymbol{k}$ -dim. Fourier subspace A :

$$\sum_{S \notin A} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 \geq 1 - \boldsymbol{\gamma}$$

  - Any $\boldsymbol{k}$ -dim. linear sketch makes error $\frac{1}{2} - \frac{\sqrt{\boldsymbol{\gamma}}}{2}$
  - Converse doesn't hold, i.e. concentration is not enough

Randomized Sketching: Hardness

- Not $\boldsymbol{\gamma}$-concentrated on $o(n)$-dim. Fourier subspaces:
  - Almost all **symmetric functions**, i.e. $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{h}(\sum_i x_i)$
    - If not Fourier-close to constant or $\bigoplus_{i=1}^{n} x_i$
    - E.g. Majority (not an extractor even for $O(\sqrt{n})$)
  - **Tribes** (balanced DNF)
  - **Recursive majority**: $Maj^{\circ k} = Maj_3 \circ Maj_3 \ldots \circ Maj_3$
  - Composition theorem (under some conditions):
    - Ambainis' "Sort function": recursive $Sort_4$
    - Kushilevitz's "Icosahedron Function": recursive Hex_6

Uniform Distribution + Approx. Dimension

- Not $\boldsymbol{\gamma}$-concentrated on $\boldsymbol{k}$ -dim. Fourier subspaces
  - $\forall$ $\boldsymbol{k}$ -dim. Fourier subspace A: $\sum_{S\notin A} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 \geq 1-\boldsymbol{\gamma}$
  - Any $\boldsymbol{k}$ -dim. linear sketch makes error ½$(1-\sqrt{\boldsymbol{\gamma}})$
- **Definition** (Approximate Fourier Dimension)
  - $\dim_{\boldsymbol{\gamma}}(\boldsymbol{f}) =$ largest $\boldsymbol{d}$ such that $\boldsymbol{f}$ is not $\boldsymbol{\gamma}$-concentrated on any Fourier subspace of dimension $\boldsymbol{d}$
- Over **uniform distribution** $(\dim_{\mathbf{1}-\boldsymbol{\epsilon}}(\boldsymbol{f})+1)$-dimensional sketch is enough for error $\leq \boldsymbol{\epsilon}$:
  - Fix $(\dim_{1-\boldsymbol{\epsilon}}(\boldsymbol{f})+1)$-dimensional A: $\sum_{S\in A} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 \geq 1-\boldsymbol{\epsilon}$
  - Output: $\boldsymbol{g}(x) = \mathrm{sign}\left(\sum_{\boldsymbol{S}\in A} \hat{\boldsymbol{f}}(\boldsymbol{S})\chi_{\boldsymbol{S}}(x)\right)$:

$$\Pr_{x\sim U(\{-1,1\}^n)}[\boldsymbol{g}(x)=\boldsymbol{f}(x)] \geq 1-\boldsymbol{\epsilon}$$

Sketching over Uniform Distribution

$\mathfrak{D}_{\delta}^{1,U}(f)$ = bit-complexity of the **best compression scheme** allowing to compute f with err. δ over uniform distribution

$\dim_{\gamma}(f)$ = largest d such that f is not γ-concentrated on any Fourier subspace of dimension d

Thm: If $\epsilon_2 > \epsilon_1 > 0, \dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1$ then:

$$\mathfrak{D}_{\delta}^{1,U}(f) \geq d,$$

where $\delta = (\epsilon_2 - \epsilon_1)/4.$

Corollary: If $\hat{f}(\emptyset) < C$ for $C < 1$ then there exists d:

$$\mathfrak{D}_{\Theta\left(\frac{1}{n}\right)}^{1,U}(f) \geq d.$$

This is optimal up to error as d-dim. sketch has error $1 - \epsilon_2$

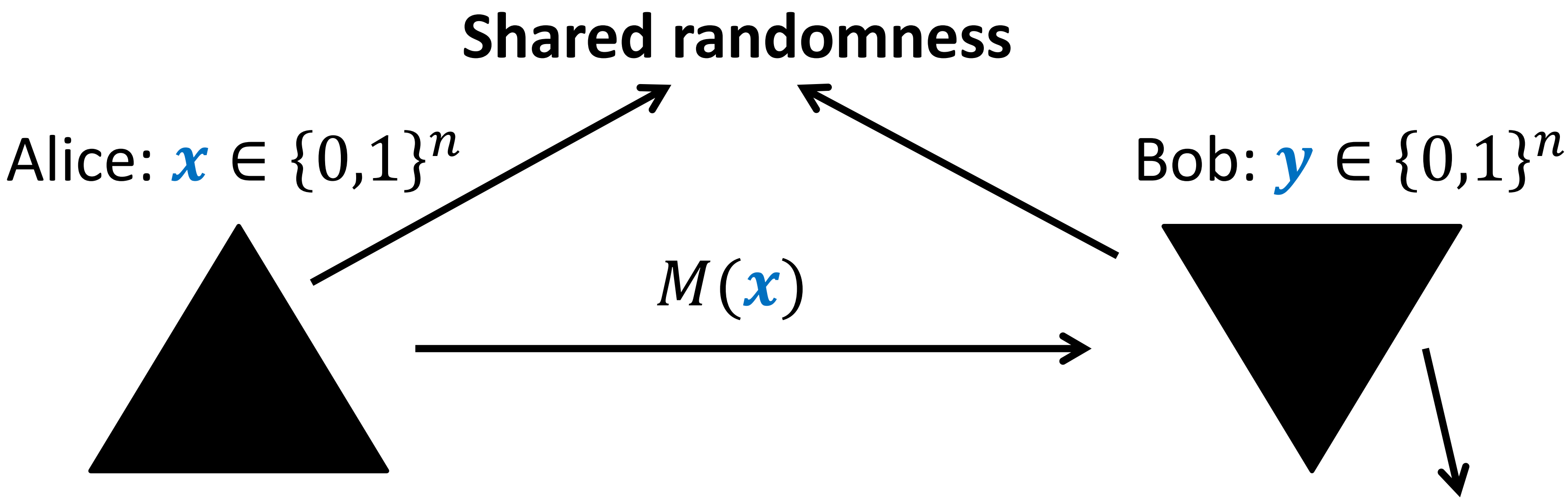
Example: Majority

- Majority function:
$$\boldsymbol{Maj_n}(z_1, \dots, z_n) \equiv \sum_{i=1}^{n} z_i \geq n/2$$
- $\widehat{\boldsymbol{Maj_n}}(\boldsymbol{S})$ only depends on $|\boldsymbol{S}|$
- $\widehat{\boldsymbol{Maj_n}}(\boldsymbol{S}) = 0$ if $|\boldsymbol{S}|$ is odd
- $W^{\boldsymbol{k}}(\boldsymbol{Maj_n}) = \sum_{\boldsymbol{S}:|\boldsymbol{S}|=\boldsymbol{k}} \widehat{\boldsymbol{Maj_n}}(\boldsymbol{S}) = \alpha \boldsymbol{k}^{-\frac{3}{2}}\left(1 \pm O\left(\frac{1}{\boldsymbol{k}}\right)\right)$
- $(n-1)$-dimensional subspace with most weight:
$$\boldsymbol{A_{n-1}} = span(\{1\},\{2\},\dots,\{n-1\})$$
- $\sum_{\boldsymbol{S} \in \boldsymbol{A_{n-1}}} \widehat{\boldsymbol{Maj_n}}(\boldsymbol{S}) = 1 - \frac{\gamma}{\sqrt{n}} \pm O\left(n^{-3/2}\right)$
- Set $\boldsymbol{\epsilon_2} = 1 - O(n^{-3/2})$, $\boldsymbol{\epsilon_1} = 1 - \frac{\gamma}{\sqrt{n}} + O\left(n^{-3/2}\right)$
$$\mathfrak{D}^{1,U}_{O(1/\sqrt{n})}(\boldsymbol{Maj_n}) \geq \boldsymbol{n}$$

Application: Random Streams

- $\boldsymbol{x} \in \{0,1\}^n$ generated via a stream of updates
- Each update randomly flips a **random coordinate**
- **Goal**: maintain $\boldsymbol{g}(\boldsymbol{x})$ during the stream (error $\boldsymbol{\epsilon}$)
- **Question**: how much space necessary?
- **Answer:** $\mathfrak{D}_{\epsilon}^{1,U}$ and best algorithm is linear sketch
  - $\boldsymbol{x}$ after first O($n \log n$) updates input is uniform
- **Big open question:**
  - Is the same true if distribution is not uniform?
  - True for very long streams, how about short ones?

1-way Communication Complexity of XOR-functions

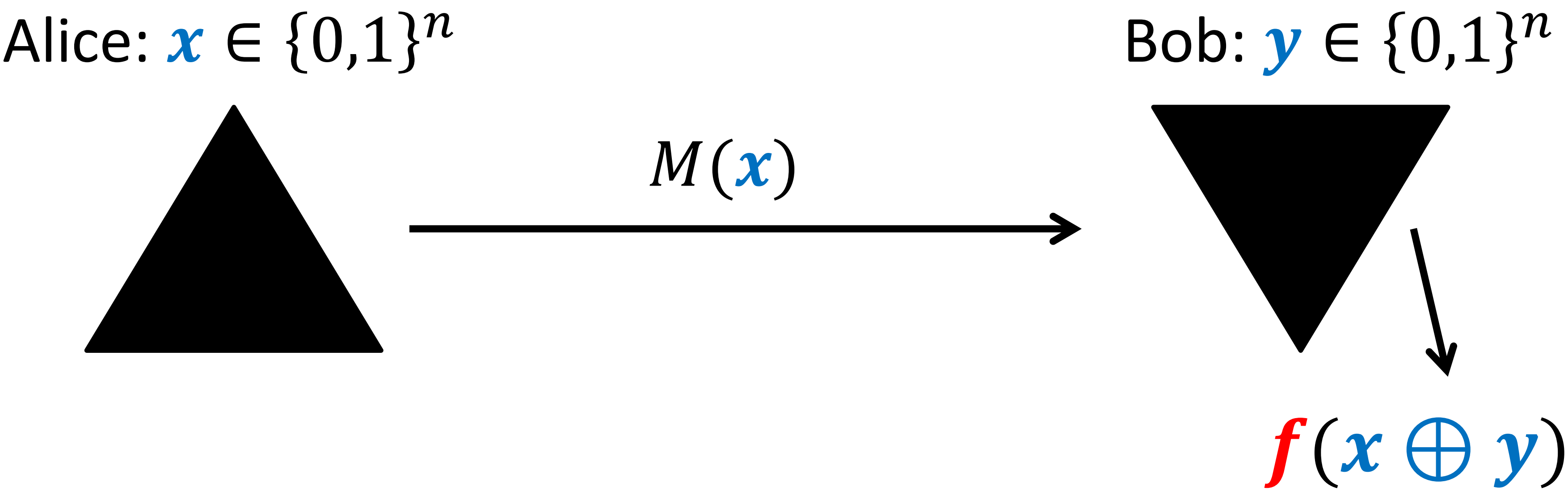


- Examples:
  - $\boldsymbol{f}(\mathrm{z}) = OR_{i=1}^{n}(z_i) \Rightarrow$ (not) Equality
  - $\boldsymbol{f}(\mathrm{z}) = (\left\|z\right\|_0 > \mathrm{d}) \Rightarrow$ Hamming Distance $> \mathrm{d}$
- $R_\epsilon^1(\boldsymbol{f})$ = min.|M| so that Bob's error prob. ϵ

Communication Complexity of XOR-functions

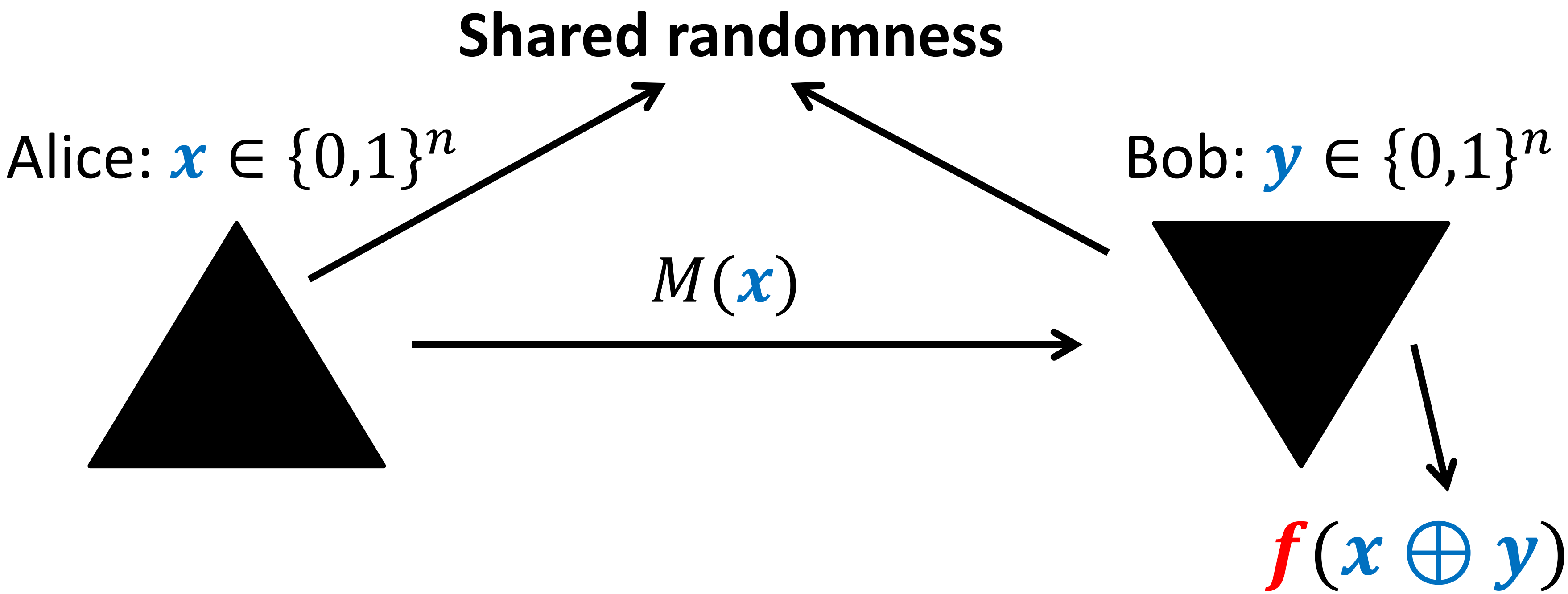
- Well-studied (often for 2-way communication):
  - [Montanaro,Osborne], ArXiv'09
  - [Shi, Zhang], QIC'09,
  - [Tsang, Wong, Xie, Zhang], FOCS'13
  - [O'Donnell, Wright, Zhao,Sun,Tan], CCC'14
  - [Hatami, Hosseini, Lovett], FOCS'16
- Connections to log-rank conjecture [Lovett'14]:
  - Even special case for XOR-functions still open

Deterministic 1-way Communication Complexity of XOR-functions



- $D^1(\boldsymbol{f})$ = min.|M| so that Bob is always correct
- [Montanaro-Osborne'09]: $D^1(\boldsymbol{f}) = D^{lin}(\boldsymbol{f})$
- $D^{lin}(\boldsymbol{f}) =$ deterministic lin. sketch complexity of $\boldsymbol{f}$
- $D^1(\boldsymbol{f}) = D^{lin}(\boldsymbol{f}) =$ "Fourier dimension of $\boldsymbol{f}$"

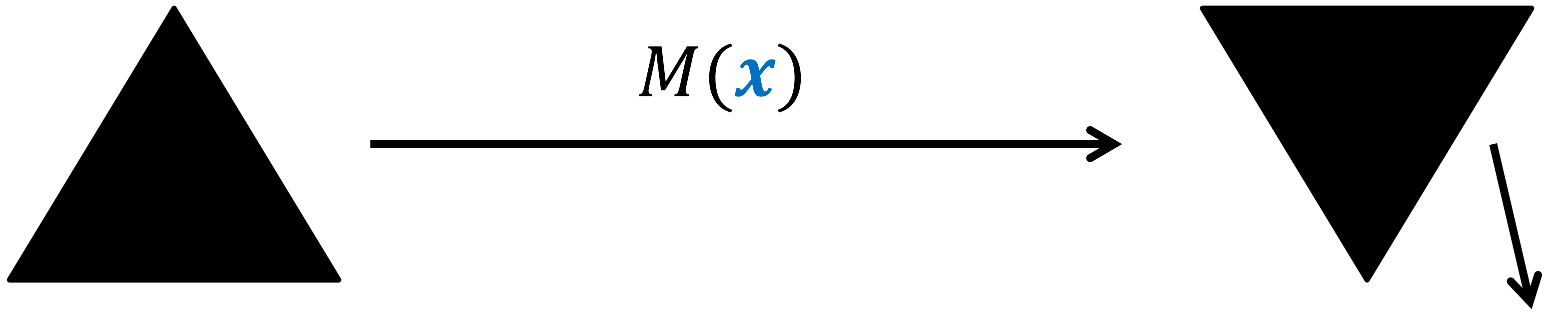
1-way Communication Complexity of XOR-functions



- $R_\epsilon^1(f)$ = min. |M| so that Bob's error prob. ϵ
- $R_\epsilon^{lin}(f) =$ rand. lin. sketch complexity (error ϵ)
- $R_\epsilon^1(f) \leq R_\epsilon^{lin}(f)$
- **Question:** $R_\epsilon^1(f) \approx R_\epsilon^{lin}(f)$? (true for symmetric)

Distributional 1-way Communication under Uniform Distribution

Alice: $\boldsymbol{x} \sim U(\{0,1\}^n)$ Bob: $\boldsymbol{y} \sim U(\{0,1\}^n)$



$\boldsymbol{f}(\boldsymbol{x} \oplus \boldsymbol{y})$

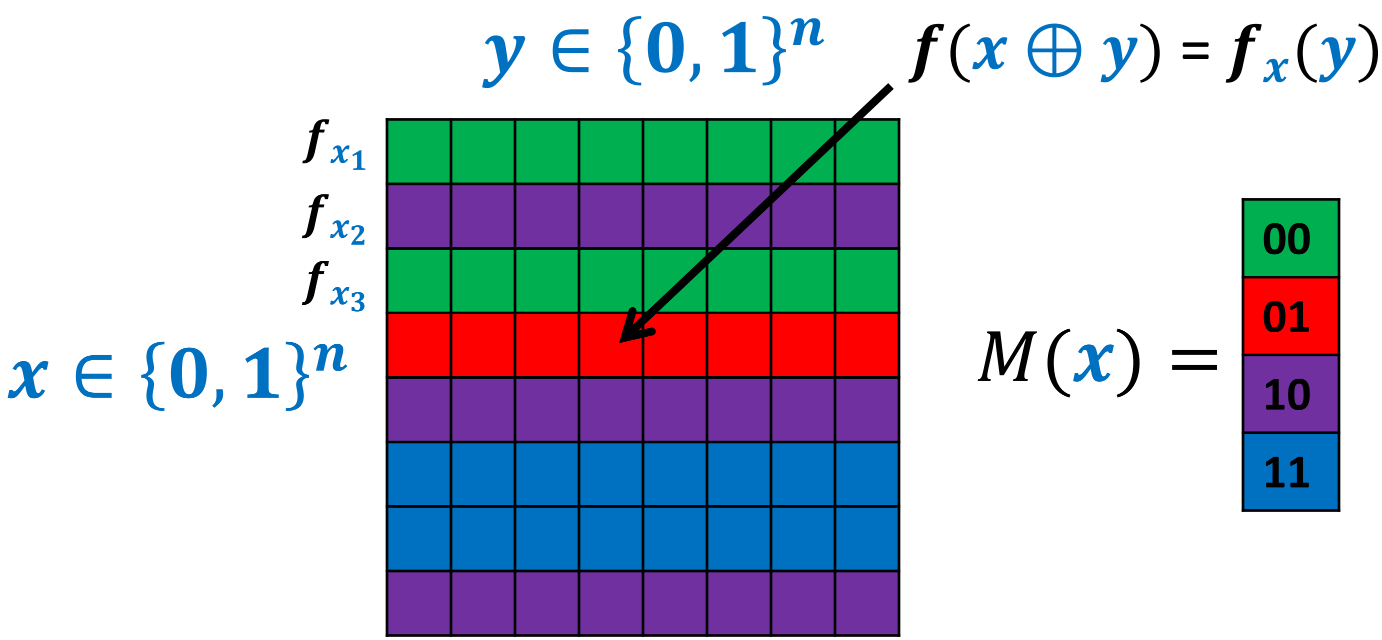
- $\mathfrak{D}_{\epsilon}^{1,U}(\boldsymbol{f})$ = min.|M| so that Bob's error prob. ϵ is over the uniform distribution over $(\boldsymbol{x}, \boldsymbol{y})$
- Enough to consider deterministic messages only
- Motivation: streaming/distributed with random input
- $R_{\epsilon}^{1}(\boldsymbol{f}) = \sup_{D} \mathfrak{D}_{\epsilon}^{1,D}(\boldsymbol{f})$

$\mathfrak{D}_{\epsilon}^{1,U}$ and Approximate Fourier Dimension

Thm: If $\epsilon_2 > \epsilon_1 > 0, \dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1$ then:

$$\mathfrak{D}_{\delta}^{1,U}(f) \geq d,$$

where $\delta = (\epsilon_2 - \epsilon_1)/4$.



$\mathfrak{D}_{\epsilon}^{1,U}$ and Approximate Fourier Dimension

- If $|\boldsymbol{M}(\boldsymbol{x})| = \boldsymbol{d} - 1$ average "rectangle" size = $2^{\boldsymbol{n}-\boldsymbol{d}+\mathbf{1}}$
- A subspace A **distinguishes** $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ if:

$$\exists \boldsymbol{S} \in A : \chi_{\boldsymbol{S}}(\boldsymbol{x}_1) \neq \chi_{\boldsymbol{S}}(\boldsymbol{x}_2)$$

- Fix a $\boldsymbol{d}$-dim. subspace $A_{\boldsymbol{d}}$: typical $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in a typical "rectangle" are distinguished by $A_{\boldsymbol{d}}$
- **Lem:** If a $\boldsymbol{d}$-dim. subspace $A_{\boldsymbol{d}}$ distinguishes $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ +

1) $\boldsymbol{f}$ is $\boldsymbol{\epsilon}_2$-concentrated on $A_{\boldsymbol{d}}$

2) $\boldsymbol{f}$ not $\boldsymbol{\epsilon}_1$-concentrated on any $(\boldsymbol{d} - 1)$-dim. subspace

$$\Pr_{z \sim U(\{-1,1\}^n)}\left[\boldsymbol{f}_{\boldsymbol{x}_1}(z) \neq \boldsymbol{f}_{\boldsymbol{x}_2}(z)\right] \geq \boldsymbol{\epsilon}_2 - \boldsymbol{\epsilon}_1$$

$\mathfrak{D}_{\epsilon}^{1,U}$ and Approximate Fourier Dimension

Thm: If $\epsilon_2 > \epsilon_1 > 0, \dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1$ then:

$$\mathfrak{D}_{\delta}^{1,U}(f) \geq d,$$

Where $\delta = (\epsilon_2 - \epsilon_1)/4$.

$$\Pr_{z \sim U(\{-1,1\}^n)}[f_{x_1}(z) \neq f_{x_2}(z)] \geq \epsilon_2 - \epsilon_1$$

				y				
g_{x_1}	0	1	1	0	0	0	1	0
g_{x_2}	0	1	0	1	0	1	1	0
				0				

R = "typical rectangle"

Error for fixed y = min($\Pr_{x \in R}[f_x(y) = 0], \Pr_{x \in R}[f_x(y) = 1]$)

Average error for $(x, y) \in R$ = $\Omega(\epsilon_2 - \epsilon_1)$

Thanks! Questions?