

Linear sketching with parities

Grigory Yaroslavtsev

<http://grigory.us>



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

Linear sketching with parities

- Input $\mathbf{x} \in \{0,1\}^n$
- Parity = Linear function over \mathbb{GF}_2 : $\bigoplus_{i \in S} x_i$
- E.g. $x_4 \oplus x_2 \oplus x_{42}$
- **Deterministic linear sketch**: set of k parities:
$$\ell(\mathbf{x}) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \dots; \bigoplus_{i_k \in S_k} x_{i_k}$$
- **Randomized linear sketch**: **distribution** over k parities (random S_1, S_2, \dots, S_k):
$$\ell(\mathbf{x}) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \dots; \bigoplus_{i_k \in S_k} x_{i_k}$$

Linear sketching over \mathbb{GF}_2

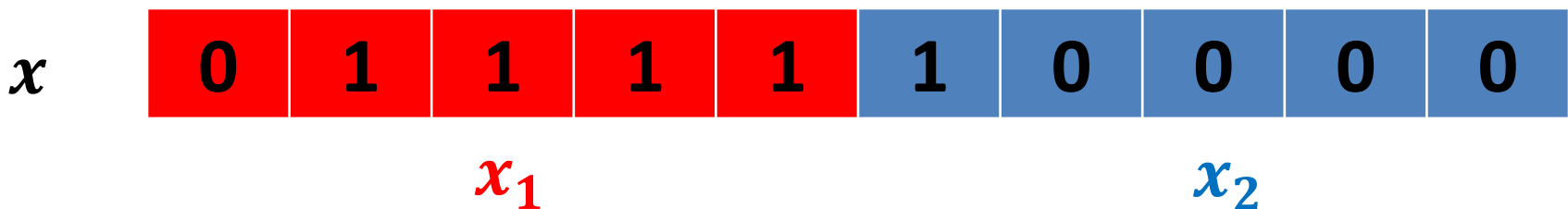
- Given $f(\mathbf{x}): \{0,1\}^n \rightarrow \{0,1\}$
- **Question:**

Can one recover $f(\mathbf{x})$ from a small ($k \ll n$) linear sketch over \mathbb{GF}_2 ?

- Allow randomized computation (99% success)
 - Probability over choice of random sets
 - Sets are known at recovery time
 - Recovery is deterministic (also consider randomized)

Motivation: Distributed Computing

- **Distributed computation among M machines:**
 - $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M)$ (more generally $\mathbf{x} = \bigoplus_{i=1}^M \mathbf{x}_i$)
 - M machines can compute sketches locally:
 $\ell(\mathbf{x}_1), \dots, \ell(\mathbf{x}_M)$
 - Send them to the coordinator who computes:
 $\ell_i(\mathbf{x}) = \ell_i(\mathbf{x}_1) \oplus \dots \oplus \ell_i(\mathbf{x}_M)$ (coordinate-wise XORs)
 - Coordinator computes $f(\mathbf{x})$ with kM communication



Motivation: Streaming

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

x_0	0	0	0	0	0	0	0	0
Updates: (1, 3, 8)								
x_1	1	0	0	0	0	0	0	0
x_2	1	0	0	0	0	0	0	0
x_3	1	0	0	0	0	1	0	0
x	1	0	0	0	0	1	0	0



$\ell(x)$ allows to recover $f(x)$ with k bits of space

Deterministic vs. Randomized

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

Fourier Analysis

- $f(x_1, \dots, x_n): \{0,1\}^n \rightarrow \{0,1\}$

- Notation switch:

 - $0 \rightarrow 1$

 - $1 \rightarrow -1$

- $f': \{-1,1\}^n \rightarrow \{-1,1\}$

- Functions as vectors form a vector space:

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

- Inner product on functions = “correlation”:

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

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

“Main Characters” are Parities

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

$$f(x_1, \dots, x_n) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x)$$

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

Fourier Dimension

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

$$\sum_{S \subseteq A_k} \hat{f}(S)^2 = 1$$

$$f(x_1, \dots, x_n) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) = \sum_{S \subseteq A_k} \hat{f}(S) \chi_S(x)$$

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

Deterministic Sketching and Noise

Suppose “noise” has a bounded norm

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

- L_0 -noise in the Fourier domain (via [Sanyal'15])
 - $\hat{\mathbf{f}} = \mathbf{k}\text{-dim.} + \text{“Fourier } L_0\text{-noise”}$
 - $\left\| \widehat{\text{noise}} \right\|_0 = \# \text{ non-zero Fourier coefficients of noise}$
(aka “Fourier sparsity”)
 - Linear sketch size: $\mathbf{k} + O\left(\left\| \widehat{\text{noise}} \right\|_0^{1/2}\right)$
 - **Our work:** can't be improved even with randomness and even for uniform x , e.g. for “addressing function”.

How Randomization Handles Noise

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

Randomized Sketching: Hardness

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

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

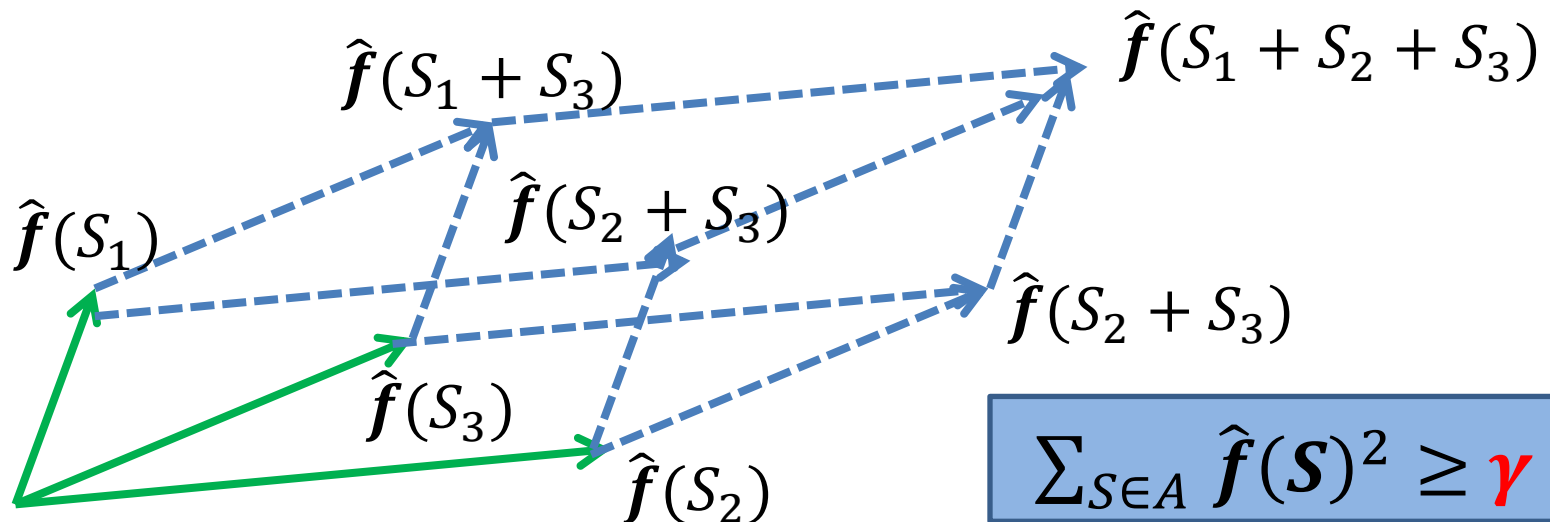
- Any k -dim. linear sketch makes error $\frac{1-\sqrt{\gamma}}{2}$
- Converse doesn't hold, i.e. concentration is not enough

Randomized Sketching: Hardness

- Not γ -concentrated on $o(n)$ -dim. Fourier subspaces:
 - Almost all **symmetric functions**, i.e. $f(x) = h(\sum_i x_i)$
 - 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**: $Maj^{\circ k} = Maj_3 \circ Maj_3 \dots \circ Maj_3$

Approximate Fourier Dimension

- Not γ -concentrated on k -dim. Fourier subspaces
 - $\forall k$ -dim. Fourier subspace A : $\sum_{S \notin A} \hat{f}(S)^2 \geq 1 - \gamma$
 - Any k -dim. linear sketch makes error $\frac{1}{2}(1 - \sqrt{\gamma})$
- **Definition** (Approximate Fourier Dimension)
 - $\dim_{\gamma}(f) =$ smallest d such that f is γ -concentrated on some Fourier subspace of dimension d



Sketching over Uniform Distribution + Approximate Fourier Dimension

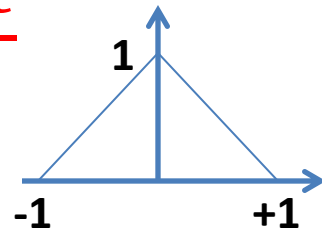
- Sketching error over **uniform distribution of x** .
- $\dim_{\epsilon}(\mathbf{f})$ -dimensional sketch gives error **$1 - \epsilon$** :
 - Fix $\dim_{\epsilon}(\mathbf{f})$ -dimensional $A: \sum_{S \in A} \hat{\mathbf{f}}(S)^2 \geq \epsilon$
 - Output: $\mathbf{g}(x) = \text{sign}(\sum_{S \in A} \hat{\mathbf{f}}(S) \chi_S(x))$:

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

- We show a basic refinement \Rightarrow error **$\frac{1-\epsilon}{2}$**

– Pick θ from a carefully chosen distribution

– Output: $\mathbf{g}_{\theta}(x) = \text{sign}(\sum_{S \in A} \hat{\mathbf{f}}(S) \chi_S(x) - \theta)$



Sketching over Uniform Distribution

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

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 :

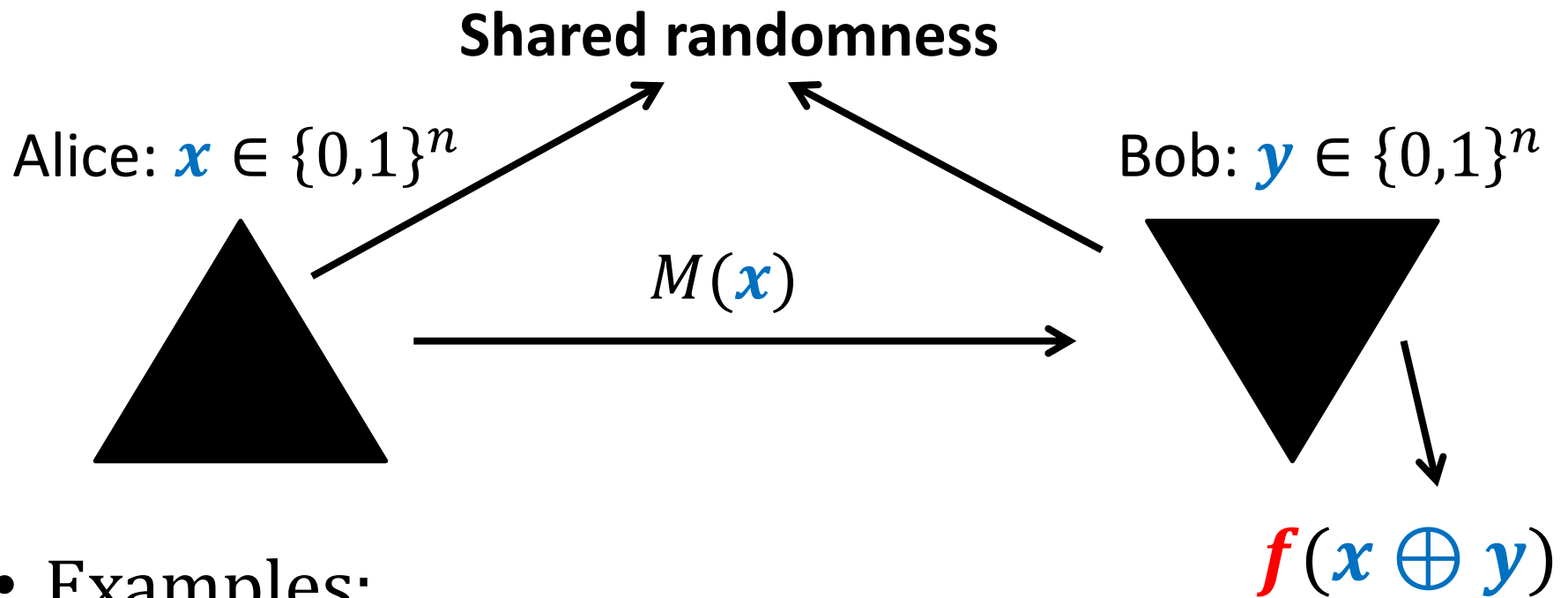
$$\mathfrak{D}_{\Theta\left(\frac{1}{n}\right)}^{1,U}(f) \geq d.$$

- Optimal up to error as d -dim. linear sketch has error $\frac{1-\epsilon_2}{2}$
- Tight for the Majority function

Application: Random Streams

- $\mathbf{x} \in \{0,1\}^n$ generated via a stream of updates
- Each update randomly flips a **random coordinate**
- **Goal:** maintain $f(\mathbf{x})$ during the stream (error ϵ)
- **Question:** how much space necessary?
- **Answer:** $\mathcal{D}_{\epsilon}^{1,U}$ and best algorithm is linear sketch
 - After first $O(n \log n)$ updates input \mathbf{x} is uniform
- **Big open question:**
 - Is the same true if \mathbf{x} is not uniform?
 - True for **VERY LONG** ($2^{2^{2^{\Omega(n)}}}$) streams (via [LNW'14])
 - How about short ones?

1-way Communication Complexity of XOR-functions



- Examples:

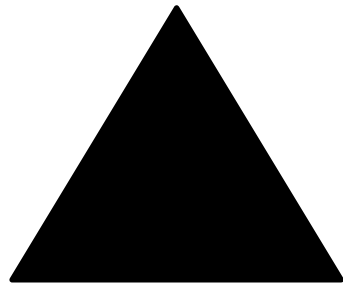
- $f(z) = OR_{i=1}^n(z_i) \Rightarrow$ (not) Equality
- $f(z) = (||z||_0 > d) \Rightarrow$ Hamming Distance $> d$
- $R_\epsilon^1(f) = \min. |M|$ so that Bob's error prob. ϵ

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

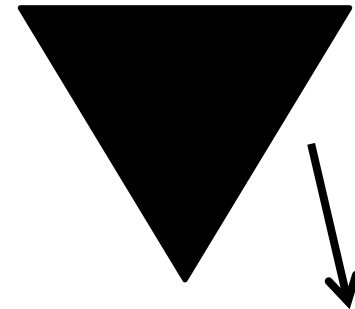
Alice: $\mathbf{x} \in \{0,1\}^n$



$M(\mathbf{x})$



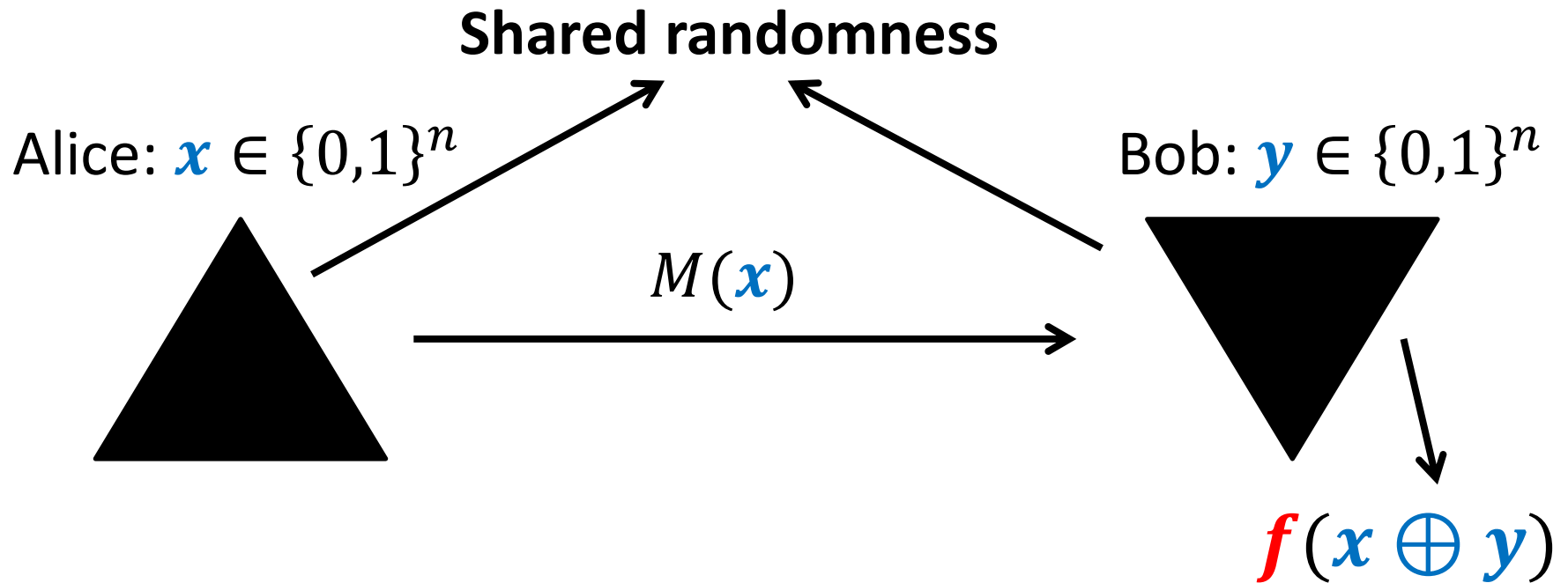
Bob: $\mathbf{y} \in \{0,1\}^n$



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

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

1-way Communication Complexity of XOR-functions

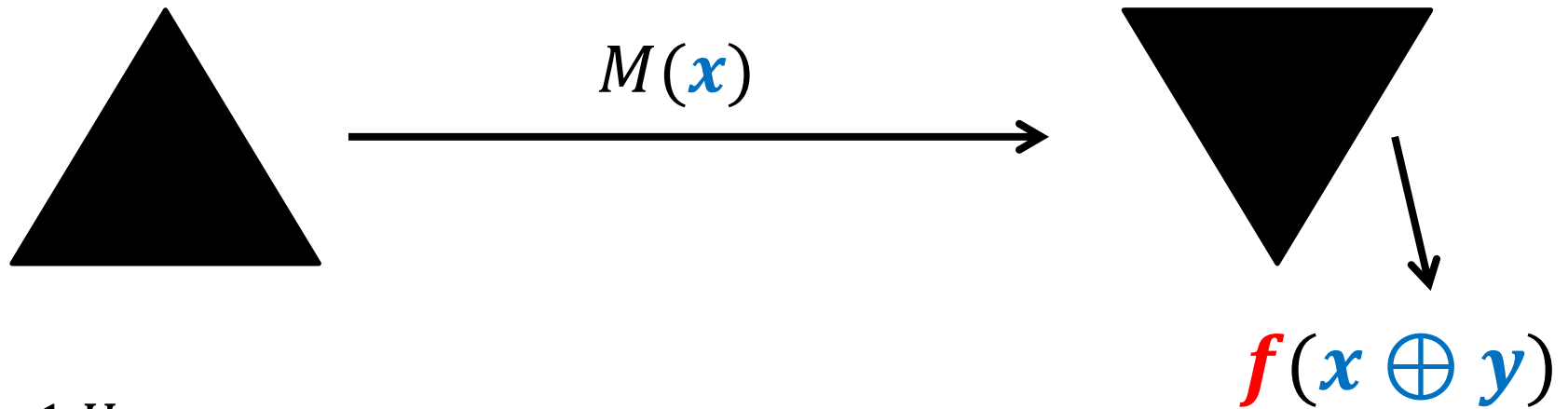


- $R_\epsilon^1(f) = \min. |M|$ so that Bob's error prob. ϵ
- $R_\epsilon^{lin}(f) = \text{rand. lin. sketch complexity (error } \epsilon \text{)}$
- $R_\epsilon^1(f) \leq R_\epsilon^{lin}(f)$
- **Question:** $R_\epsilon^1(f) \approx R_\epsilon^{lin}(f)$? (true for symmetric)

Distributional 1-way Communication under Uniform Distribution

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

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



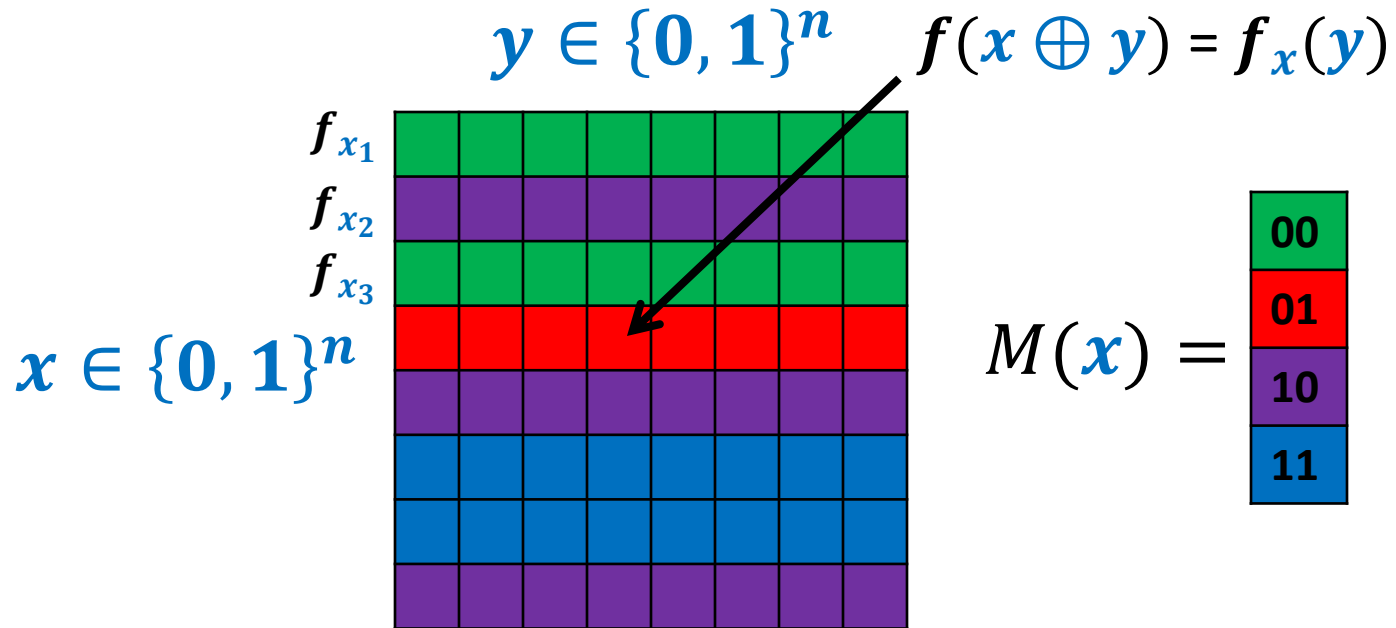
- $\mathfrak{D}_{\epsilon}^{1,U}(f) = \min. |M|$ so that Bob's error prob. ϵ is over the uniform distribution over (\mathbf{x}, \mathbf{y})
- Enough to consider deterministic messages only
- Motivation: streaming/distributed with random input
- $R_{\epsilon}^1(f) = \sup_D \mathfrak{D}_{\epsilon}^{1,D}(f)$

$\mathfrak{D}_\epsilon^{1,U}$ 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 $|M(\mathbf{x})| = d - 1$ average “rectangle” size = 2^{n-d+1}
- A subspace A **distinguishes** \mathbf{x}_1 and \mathbf{x}_2 if:
$$\exists S \in A : \chi_S(\mathbf{x}_1) \neq \chi_S(\mathbf{x}_2)$$
- Fix a d -dim. subspace A_d : typical \mathbf{x}_1 and \mathbf{x}_2 in a typical “rectangle” are distinguished by A_d
- **Lem:** If a d -dim. subspace A_d distinguishes \mathbf{x}_1 and \mathbf{x}_2 +
 - 1) f is ϵ_2 -concentrated on A_d
 - 2) f not ϵ_1 -concentrated on any $(d - 1)$ -dim. subspace

$$\Pr_{z \sim U(\{-1,1\}^n)} [f_{\mathbf{x}_1}(z) \neq f_{\mathbf{x}_2}(z)] \geq \epsilon_2 - \epsilon_1$$

$\mathfrak{D}_\epsilon^{1,U}$ 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					

$R = \text{“typical rectangle”}$

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

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

Thanks! Questions?

- Other stuff:
 - Sketching Linear Threshold Functions: $O\left(\frac{\theta}{m} \log \frac{\theta}{m}\right)$
 - Resolves a communication conjecture of [MO'09]
- Blog post: <http://grigory.us/blog/the-binary-sketchman>



Example: Majority

- Majority function:

$$\mathbf{Maj}_n(z_1, \dots, z_n) \equiv \sum_{i=1}^n z_i \geq n/2$$

- $\widehat{\mathbf{Maj}}_n(\mathcal{S})$ only depends on $|\mathcal{S}|$

- $\widehat{\mathbf{Maj}}_n(\mathcal{S}) = 0$ if $|\mathcal{S}|$ is odd

- $W^k(\mathbf{Maj}_n) = \sum_{\mathcal{S}:|\mathcal{S}|=k} \widehat{\mathbf{Maj}}_n(\mathcal{S}) = \alpha k^{-\frac{3}{2}} \left(1 \pm O\left(\frac{1}{k}\right) \right)$

- $(n - 1)$ -dimensional subspace with most weight:

$$A_{n-1} = \text{span}(\{1\}, \{2\}, \dots, \{n - 1\})$$

- $\sum_{\mathcal{S} \in A_{n-1}} \widehat{\mathbf{Maj}}_n(\mathcal{S}) = 1 - \frac{\gamma}{\sqrt{n}} \pm O(n^{-3/2})$

- Set $\epsilon_2 = 1 - O(n^{-3/2})$, $\epsilon_1 = 1 - \frac{\gamma}{\sqrt{n}} + O(n^{-3/2})$

$$\mathfrak{D}_{O(1/\sqrt{n})}^{1,U}(\mathbf{Maj}_n) \geq n$$