Linear sketching over $\mathbb{F}_2$

Grigory Yaroslavtsev

http://grigory.us

with Sampath Kannan (Penn),
Elchanan Mossel (MIT) and Swagato Sanyal (NUS)
Linear sketching with parities

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

• Given $f(x) : \{0,1\}^n \to \{0,1\}$

• Question:
  Can one recover $f(x)$ 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:
  
  – $x = (x_1, x_2, \ldots, x_M)$ (more generally $x = \bigoplus_{i=1}^{M} x_i$)
  
  – $M$ machines can compute sketches locally:
    
    $\ell(x_1), \ldots, \ell(x_M)$
    
  – Send them to the coordinator who computes:
    
    $\ell_i(x) = \ell_i(x_1) \oplus \cdots \oplus \ell_i(x_M)$ (coordinate-wise XORs)
    
  – Coordinator computes $f(x)$ with $kM$ communication

\[
\begin{array}{cccccccc}
  x & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
  x_1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
  x_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}
\]
Motivation: Streaming

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

$\ell(x)$ allows to recover $f(x)$ with $k$ bits of space
Deterministic vs. Randomized

• **Fact:** $f$ has a deterministic sketch if and only if
  
  $f = g(\bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \ldots; \bigoplus_{i_k \in S_k} x_{i_k})$

  – Equivalent to "$f$ has Fourier dimension $k$"

• **Randomization can help:**
  
  – OR: $f(x) = x_1 \lor \ldots \lor x_n$
  
  – Has "Fourier dimension" = $n$
  
  – Pick $t = \log 1/\delta$ random sets $S_1, \ldots, S_t$
  
  – If there is $j$ such that $\bigoplus_{i \in S_j} x_i = 1$ output 1,
    otherwise output 0

  – Error probability $\delta$
Fourier Analysis

- \( f(x_1, \ldots, x_n): \{0,1\}^n \to \{0,1\} \)
- Notation switch:
  - 0 → 1
  - 1 → -1
- \( f': \{-1,1\}^n \to \{-1,1\} \)
- Functions as vectors form a vector space:
  \( f: \{-1,1\}^n \to \{-1,1\} \iff f \in \{-1,1\}^{2^n} \)
- 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 \textbf{character} $\chi_S(x) = \prod_{i \in S} x_i$

• \textbf{Fact:} Every function $f : \{-1,1\}^n \rightarrow \{-1,1\}$ uniquely represented as multilinear polynomial

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

• $\hat{f}(S)$ 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 $S \equiv \text{vectors in } \mathbb{G} F_2^n$
- "$f$ has Fourier dimension $k$" = a $k$-dimensional subspace in Fourier domain has all weight
  \[
  \sum_{S \subseteq A_k} \hat{f}(S)^2 = 1
  \]
  \[
  f(x_1, \ldots, x_n) = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S(x) = \sum_{S \subseteq A_k} \hat{f}(S) \chi_S(x)
  \]
- Pick a basis $S_1, \ldots, S_k$ in $A_k$:
  - Sketch: $\chi_{S_1}(x), \ldots, \chi_{S_k}(x)$
  - 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

\[ f = k \text{-dim.} + \text{noise} \]

- \( L_0 \)-noise in the Fourier domain (via [Sanyal’15])
  - \( \hat{f} = k \text{-dim.} + \text{“Fourier } L_0 \text{-noise”} \)
  - \( \| \text{noise} \|_0 \) = \# non-zero Fourier coefficients of noise (aka “Fourier sparsity”)

- Linear sketch size: \( k + O(\| \text{noise} \|_0^{1/2}) \)

- **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: $k + O(\log \left| |noise| \right|_0)$
  – 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: $k + O(\left| |\hat{noise}| \right|_1^2)$
  – 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. $\geq 0.1$ each
  - Example (inner product): $f(x) = \bigoplus_{i=1}^{n/2} x_{2i-1}x_{2i}$

- 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 $\gamma$-concentrated on $o(n)$-dim. Fourier subspaces:
  – Almost all \textbf{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 $O(\sqrt{n})$)
  – \textbf{Tribes} (balanced DNF)
  – \textbf{Recursive majority}: $\text{Maj}^{\circ k} = \text{Maj}_3 \circ \text{Maj}_3 \ldots \circ \text{Maj}_3$
Approximate Fourier Dimension

• Not \( \gamma \)-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)
  – \( \text{dim}_\gamma(f) = \) smallest \( d \) such that \( f \) is \( \gamma \)-concentrated on some Fourier subspace of dimension \( d \)

\[
\sum_{S \in A} \hat{f}(S)^2 \geq \gamma
\]
Sketching over Uniform Distribution + Approximate Fourier Dimension

• Sketching error over **uniform distribution of** \(x\).

• \(\dim_\varepsilon (f)\)-dimensional sketch gives error \(1 - \varepsilon\):
  
  – Fix \(\dim_\varepsilon (f)\)-dimensional \(A: \sum_{S \in A} \hat{f}(S)^2 \geq \varepsilon\)
  
  – Output: \(g(x) = \text{sign}(\sum_{S \in A} \hat{f}(S)\chi_S(x))\):

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

• We show a basic refinement \(\Rightarrow\) error \(\frac{1 - \varepsilon}{2}\):

  – Pick \(\theta\) from a carefully chosen distribution

  – Output: \(g_\theta(x) = \text{sign}(\sum_{S \in A} \hat{f}(S)\chi_S(x) - \theta)\)
1-way Communication Complexity of XOR-functions

- **Examples:**
  - \( f(z) = OR_{i=1}^{n}(z_i) \Rightarrow \text{(not) Equality} \)
  - \( f(z) = (||z||_0 > d) \Rightarrow \text{Hamming Distance} > d \)
  - \( R_{\epsilon}^1(f^+) = \min. |M| \) so that Bob’s error prob. \( \epsilon \)
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: \( x \in \{0,1\}^n \)

Bob: \( y \in \{0,1\}^n \)

\[ f^+ = f(x \oplus y) \]

- \( D^1(f) = \min |M| \) 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

Shared randomness

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

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

$M(x)$

$f(x \oplus y)$

- $R_\epsilon^1(f) = \min |M|$ so that Bob’s error prob. $\epsilon$
- $R_\epsilon^{lin}(f^+) = \text{rand. lin. sketch complexity (error } \epsilon \text{)}$
- $R_\epsilon^1(f^+) \leq R_\epsilon^{lin}(f)$
- **Question:** $R_\epsilon^1(f^+) \approx R_\epsilon^{lin}(f)$?
\[ R_\varepsilon^1(f^+) \approx R_\varepsilon^{lin}(f)? \]

Holds for:
- Majority, Tribes, recursive majority, addressing function
- Linear threshold functions
- (Almost all) symmetric functions
- Degree-\(d\) \(\mathbb{F}_2\)-polynomials:
  \[ R_{5\varepsilon}^{lin}(f) = O(d \ R_\varepsilon^1(f^+)) \]

Analogous question for 2-way is wide open:

\[ [HHL'16] \ Q_\varepsilon^{\oplus dt}(f) = poly(R_\varepsilon(f^+))? \]
Distributional 1-way Communication under Uniform Distribution

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

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

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

**Thm:** If $\dim_\epsilon(f) = d - 1$ then $\mathfrak{D}_{1-\epsilon}^{1,U}(f^+) \geq \frac{d}{6}$.

- Optimal up to error as $d$-dim. linear sketch has error $\frac{1-\epsilon}{2}$.

**Weaker:** If $\epsilon_2 > \epsilon_1$, $\dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1$ 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 $d$:
  \[
  \mathfrak{D}_{\emptyset}(f) \geq d.
  \]

- Tight for the Majority function, etc.
\( \mathcal{D}_{\epsilon}^{1,U} \) and Approximate Fourier Dimension

**Thm:** If \( \epsilon_2 > \epsilon_1 > 0 \), \( \dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1 \) then:
\[
\mathcal{D}_{\delta}^{1,U}(f) \geq d,
\]
where \( \delta = (\epsilon_2 - \epsilon_1)/4 \).

\( f(x \oplus y) = f_x(y) \) for \( x, y \in \{0, 1\}^n \),

\[
M(x) = \begin{cases} 
00 & \text{if } x = 00, \\
01 & \text{if } x = 01, \\
10 & \text{if } x = 10, \\
11 & \text{if } x = 11.
\end{cases}
\]
$\mathcal{D}_{\epsilon}^{1,U}$ and Approximate Fourier Dimension

- If $|M(x)| = d - 1$ average "rectangle" size $= 2^{n-d}+1$
- A subspace $A$ distinguishes $x_1$ and $x_2$ if:
  \[ \exists S \in A : \chi_S(x_1) \neq \chi_S(x_2) \]
- Lem 1: Fix a $d$-dim. subspace $A_d$: typical $x_1$ and $x_2$ in a typical "rectangle" are distinguished by $A_d$
- Lem 2: If a $d$-dim. subspace $A_d$ distinguishes $x_1$ and $x_2$ +
  1) $f$ is $\epsilon_2$-concentrated on $A_d$
  2) $f$ not $\epsilon_1$-concentrated on any $(d - 1)$-dim. subspace

\[
\Rightarrow \Pr_{z \sim U([-1,1]^n)}[f_{x_1}(z) \neq f_{x_2}(z)] \geq \epsilon_2 - \epsilon_1
\]
**$\mathbb{D}_{\epsilon}^{1,U}$** and Approximate Fourier Dimension

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

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

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

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

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

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

- $x \in \{0,1\}^n$ generated via a stream of updates
  - Each update flips a random coordinate
- Goal: maintain $f(x)$ during the stream (error $\epsilon$)
- 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 $x$ is uniform

Big open question:
- Is the same true if $x$ is not uniform?
- True for VERY LONG $\left(2^{2^{\Omega(n)}}\right)$ streams (via [LNW’14])
- How about short ones?
- Answer would follow from our conjecture if true
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](http://grigory.us/blog/the-binary-sketchman)
Example: Majority

- Majority function:
  \[ \text{Maj}_n(z_1, \ldots, z_n) \equiv \sum_{i=1}^{n} z_i \geq n/2 \]
- \( \text{Maj}_n(S) \) only depends on \(|S|\)
- \( \text{Maj}_n(S) = 0 \) if \(|S|\) is odd
- \( W^k(\text{Maj}_n) = \sum_{|S|=k} \text{Maj}_n(S) = \alpha k^{-3/2} \left( 1 \pm O \left( \frac{1}{k} \right) \right) \)
- \((n-1)\)-dimensional subspace with most weight:
  \[ A_{n-1} = \text{span} \{1, 2, \ldots, n-1\} \]
- \( \sum_{S \in A_{n-1}} \text{Maj}_n(S) = 1 - \frac{\gamma}{\sqrt{n}} \pm O(n^{-3/2}) \)
- Set \( \epsilon_2 = 1 - O(n^{-3/2}) \), \( \epsilon_1 = 1 - \frac{\gamma}{\sqrt{n}} + O(n^{-3/2}) \)
- Set \( D_{0(1/\sqrt{n})}^{1,U} (\text{Maj}_n) \geq n \)