Linear sketching for Functions over Boolean Hypercube

Grigory Yaroslavtsev
(Indiana University, Bloomington)

http://grigory.us

with Sampath Kannan (U. Pennsylvania),
Elchanan Mossel (MIT) and Swagato Sanyal (NUS)
\( \mathbb{F}_2 \)-Sketching

- Input \( x \in \{0,1\}^n \)
- Parity = Linear function over \( \mathbb{GF}_2 : \bigoplus_{i \in S} x_i \)
- Deterministic linear sketch: set of \( k \) parities:
  \[
  \ell(x) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \ldots; \bigoplus_{i_k \in S_k} x_{i_k}
  \]
  E.g. \( x_4 \oplus x_2 \oplus x_{42}; x_{239} \oplus x_{30}; x_{566}; \ldots \)
- Randomized linear sketch: distribution over \( k \) parities (random \( S_1, S_2, \ldots, S_k \)):
  \[
  \ell(x) = \bigoplus_{i_1 \in S_1} x_{i_1}; \bigoplus_{i_2 \in S_2} x_{i_2}; \ldots; \bigoplus_{i_k \in S_k} x_{i_k}
  \]
Linear sketching over $\mathbb{F}_2$

- Given $f(x): \{0,1\}^n \rightarrow \{0,1\}$
- **Question:** Can one recover $f(x)$ from a small ($k \ll n$) linear sketch over $\mathbb{F}_2$?

- Allow randomized computation (99% success)
  - Probability over choice of random sets
  - Sets are known at recovery time
  - Recovery is deterministic (w.l.o.g)
Motivation: Distributed Computing

- Distributed computation among $M$ machines:
  - $x = (x_1, x_2, ..., x_M)$ (more generally $x = \bigoplus_{i=1}^{M} x_i$)
  - $M$ machines can compute sketches locally: $\ell(x_1), ..., \ell(x_M)$
  - 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

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

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

<table>
<thead>
<tr>
<th>$x_0$</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_2$</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$x_3$</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mathbf{x}$</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

$\ell(\mathbf{x})$ allows to recover $f(\mathbf{x})$ with $k$ bits of space
Frequently Asked Questions

• **Q:** Why $\mathbb{F}_2$ updates instead of $\pm 1$?
  – AFAIK it doesn’t help if you know the sign

• **Q:** Some applications?
  – Essentially all dynamic graph streaming algorithms can be based on $L_0$-sampling
  – $L_0$-sampling can be done optimally using $\mathbb{F}_2$-sketching [Kapralov et al. FOCS’17]

• **Q:** Why not allow approximation?
  – Stay tuned
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:** \( f(x) = x_1 \lor \cdots \lor x_n \)
  – 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

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

• **Fact:** Every function $f : \{-1,1\}^n \rightarrow \{-1,1\}$ is uniquely represented as a 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 $S \equiv$ vectors in $\mathbb{F}_2^n$
• “$f$ 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

\[ f = k\text{-dimensional } \oplus \text{ “noise”} \]

- Sparse Fourier noise (via [Sanyal’15])
  - \( \hat{f} = k\text{-dim. } + \text{ “Fourier } L_0\text{-noise”} \)
  - \( \| \widehat{\text{noise}} \|_0 = \# \text{ non-zero Fourier coefficients of noise} \)
  (aka “Fourier sparsity”)
  - Linear sketch size: \( k + O(\| \widehat{\text{noise}} \|_0^{1/2}) \)

- **Our work**: can’t be improved even with randomness and even for uniform \( x \), e.g. for “addressing function”.

- Suppose “noise” has a bounded norm

\[ f = k\text{-dimensional } \oplus \text{ “noise”} \]

- Sparse Fourier noise (via [Sanyal’15])
  - \( \hat{f} = k\text{-dim. } + \text{ “Fourier } L_0\text{-noise”} \)
  - \( \| \widehat{\text{noise}} \|_0 = \# \text{ non-zero Fourier coefficients of noise} \)
  (aka “Fourier sparsity”)
  - Linear sketch size: \( k + O(\| \widehat{\text{noise}} \|_0^{1/2}) \)

- **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 original domain (via 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: $k + O(\sqrt{||\hat{noise}\|^2_1})$
  - 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 $\gamma$-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 **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**: \( \text{Maj}^{\circ_k} = \text{Maj}_3 \circ \text{Maj}_3 \circ \ldots \circ \text{Maj}_3 \)
Approximate Fourier Dimension

• Not $\gamma$-concentrated on $k$-dim. Fourier subspaces
  – $\forall$ $k$-dim. Fourier subspace $A$: $\sum_{S \in A} \hat{f}(S)^2 \geq 1 - \gamma$
  – Any $k$-dim. linear sketch makes error $\frac{1}{2}(1 - \sqrt{\gamma})$

• Definition (Approximate Fourier Dimension)
  – $\dim_{\gamma}(f) = \text{smallest } d \text{ such that } f \text{ is } \gamma$-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 \( x \).
• \( \dim_{\epsilon}(f) \)-dimensional sketch gives error \( 1 - \epsilon \):
  
  – Fix \( \dim_{\epsilon}(f) \)-dimensional \( A: \sum_{S \in A} \mathbf{f}(S)^2 \geq \epsilon \)
  
  – Output: \( \mathbf{g}(x) = \text{sign} \left( \sum_{S \in A} \mathbf{f}(S) \chi_{S}(x) \right) \):
    
    \[
    \Pr_{x \sim U([-1,1]^n)} \left[ \mathbf{g}(x) = \mathbf{f}(x) \right] \geq \epsilon \Rightarrow \text{error} \ 1 - \epsilon
    \]

• We show a basic refinement \( \Rightarrow \) error \( \frac{1-\epsilon}{2} \):
  
  – Pick \( \theta \) from a carefully chosen distribution
  
  – Output: \( \mathbf{g}_{\theta}(x) = \text{sign} \left( \sum_{S \in A} \mathbf{f}(S) \chi_{S}(x) - \theta \right) \)
1-way Communication Complexity of XOR-functions

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

- $D^1(f) = \min |M|$ so that Bob is always correct
- [Montanaro-Osborne’09]: $D^1(f) = D^{lin}(f)$
- $D^{lin}(f^+) = \text{deterministic lin. sketch complexity of } f^+$
- $D^1(f) = D^{lin}(f^+) = \text{Fourier dimension of } f$
1-way Communication Complexity of XOR-functions

Shared randomness

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

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

$M(x)$

$f(x \oplus y)$

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

As we show holds for:

- Majority, Tribes, recursive majority, addressing function
- Linear threshold functions
- (Almost all) symmetric functions
- Degree-$d$ $\mathbb{F}_2$-polynomials:

\[ R_{5\epsilon}^{lin} (f) = O(d R_\epsilon^1 (f^+)) \]

Analogous question for 2-way is wide open:

[HHL’16] \[ Q_\epsilon^{\oplus -dt} (f) = poly(R_\epsilon (f^+)) \]?
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 $\mathcal{D}^{1,U}_{1-\epsilon}(f^+) \geq \frac{d}{6}$.

- 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:
  $$\mathcal{D}^{1,U}_\delta(f) \geq d,$$
  where $\delta = (\epsilon_2 - \epsilon_1)/4$.

**Corollary:** If $\tilde{f}(\emptyset) < C$ for $C < 1$ then there exists $d$:
  $$\mathcal{D}^{1,U}_{\Theta(\frac{1}{n})}(f) \geq d.$$

- Tight for the Majority function, etc.
Thm: If $\epsilon_2 > \epsilon_1 > 0$, $\dim_{\epsilon_1}(f) = \dim_{\epsilon_2}(f) = d - 1$ then:

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

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

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

$$M(x) = \begin{bmatrix} 00 & 01 & 10 & 11 \end{bmatrix}$$
\[ \mathcal{D}^{1,U}_\varepsilon \] 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 \(\varepsilon_2\)-concentrated on \(A_d\)
  2) \(f\) is not \(\varepsilon_1\)-concentrated on any \((d - 1)\)-dim. subspace

\[ \Rightarrow \Pr_{z \sim U(\{-1,1\}^n)} [f_{x_1}(z) \neq f_{x_2}(z)] \geq \varepsilon_2 - \varepsilon_1 \]
$\mathcal{D}_{\epsilon}^{1,U}$ 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 $\delta = (\epsilon_2 - \epsilon_1)/4$.

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

<table>
<thead>
<tr>
<th>$g_{x_1}$</th>
<th>0</th>
<th>1</th>
<th>1</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_{x_2}$</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

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

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

$R =$ “typical rectangle”
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 prob. $\epsilon$)
- **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** $2^{2^{\Omega(n)}}$ streams (via [LNW'14])
- How about short ones?
- Answer would follow from our conjecture if true
Approximate $\mathbb{F}_2$-Sketching [Y.’17]

- $f(x_1, \ldots, x_n): \{0,1\}^n \to \mathbb{R}$
- Normalize: $\|f\|_2 = 1$
- Question:

Can one compute $f': \mathbb{E}[(f(x) - f'(x))^2 \leq \epsilon]$ from a small ($k \ll n$) linear sketch over $\mathbb{F}_2$?
Approximate $\mathbb{F}_2$-Sketching [Y.’17]

Interesting facts:

• All results under the **uniform distribution** generalize directly to approximate sketching

• $L_1$-sampling has optimal dependence on parameters:

  \[
  \frac{\|\hat{f}\|_1^2}{\epsilon}
  \]

  – **Optimal dependence**: $O\left(\frac{\|\hat{f}\|_1^2}{\epsilon}\right)$

  – **Open problem**: Is $L_1$-sampling optimal for Boolean functions?
\(F_2\)-Sketching of Valuation Functions \([Y. '17]\)

- Additive \(\left( \sum_{i=1}^{n} w_i x_i \right)\):
  - \(\Theta \left( \min \left( \frac{\|w\|_1^2}{\epsilon}, n \right) \right)\) (optimal via weighted Gap Hamming)
- Budget-additive \(\left( \min \left( b, \sum_{i=1}^{n} w_i x_i \right) \right)\):
  - \(\Theta \left( \min \left( \frac{\|w\|_1^2}{\epsilon}, n \right) \right)\)
- Coverage:
  - Optimal \(\Theta \left( \frac{1}{\epsilon} \right)\) (via \(L_1\)-Sampling)
- Matroid rank
- \(\alpha\)-Lipschitz submodular functions:
  - \(\Omega(n)\) communication lower bound for \(\alpha = \Omega(1/n)\)
  - Uses a large family of matroids from [Balcan, Harvey’10]
Thanks! Questions?

• Other stuff [Karpov, Y.]:
  – Linear Threshold Functions: \( \Theta \left( \frac{\theta}{m} \log \frac{\theta}{m} \right) \)
    • Resolves a communication conjecture of [MO’09]
  – Simple neural nets: LTF(ORs), LTF(LTFs)

• Blog post: [http://grigory.us/blog/the-binary-sketchman](http://grigory.us/blog/the-binary-sketchman)
Example: Majority

- Majority function:
  \[ \text{Maj}_n(z_1, ..., z_n) \equiv \sum_{i=1}^{n} z_i \geq n/2 \]
- \( \text{Maj}_n(S) \) 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:
  \[ A_{n-1} = \text{span} \{ \{1\}, \{2\}, ..., \{n-1\} \} \]
- \( \sum_{S \in A_{n-1}} \text{Maj}_n(S) = 1 - \frac{\gamma}{\sqrt{n}} \pm O(n^{-3/2}) \)
- Set \( \epsilon_2 = 1 - O(n^{-3/2}), \epsilon_1 = 1 - \frac{\gamma}{\sqrt{n}} + O(n^{-3/2}) \)
  \[ \mathcal{D}^{1,U}_{O(1/\sqrt{n})} (\text{Maj}_n) \geq n \]