Fast Fourier Sparsity Testing over the Boolean Hypercube

Grigory Yaroslavtsev

http://grigory.us

Penn
University of Pennsylvania

Joint with Andrew Arnold (Waterloo), Arturs Backurs (MIT), Eric Blais (Waterloo) and Krzysztof Onak (IBM).

Fourier Analysis

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

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

- Inner product on functions = "correlation":

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

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

Fourier Analysis

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

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

- $\hat{f}(S) \equiv$ Fourier coefficient of f on $S = \langle f, \chi_S \rangle$
- **Parseval's Thm:** For any $f: \{-1,1\}^n \to \mathbb{R}$

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

PAC-style learning

- **PAC**-learning under uniform distribution: for a class of functions $\boldsymbol{C}$, given access to $\boldsymbol{f} \in \boldsymbol{C}$ and $\boldsymbol{\epsilon}$ find a hypothesis $\boldsymbol{h}$ such that $\Pr_{x \sim \{-1,1\}^n}[\boldsymbol{f}(x) \neq \boldsymbol{h}(x)] \leq \boldsymbol{\epsilon}$
- Query model :
  – $(x, \boldsymbol{f}(x))$, for any $x \in \{-1,1\}^n$
- Fourier analysis helps because of **sparsity** in Fourier spectrum
  – Low-degree concentration
  – Concentration on a small number of Fourier coefficients

Fourier Analysis and Learning

Def (Fourier Concentration): Fourier spectrum of $\boldsymbol{f}: \{-1,1\}^n \to \mathbb{R}$ is $\boldsymbol{\epsilon}$-concentrated on a collection of subsets $\mathbb{F}$ if:

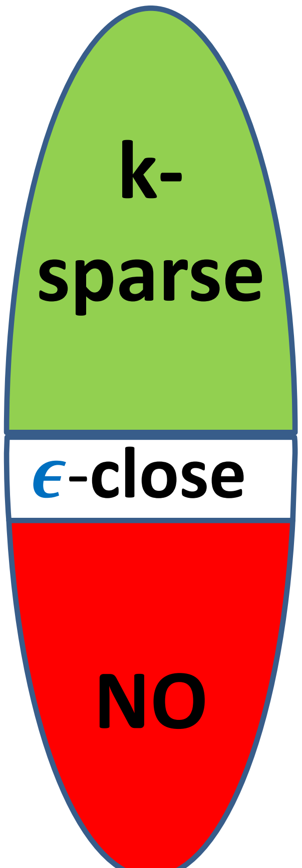
$$\sum_{\boldsymbol{S} \subseteq [n], \boldsymbol{S} \in \mathbb{F}} \hat{\boldsymbol{f}}(\boldsymbol{S})^2 \geq 1 - \boldsymbol{\epsilon}$$

Sparse Fourier Transform [Goldreich-Levin/Kushilevitz-Mansour]: Class C which is $\boldsymbol{\epsilon}$-concentrated on $\boldsymbol{k}$ sets can be PAC-learned with $\boldsymbol{k} n \, poly\left(\frac{1}{\epsilon}\right)$ queries:

$$\text{dist}(\boldsymbol{f}, \boldsymbol{h}) = \left|\left|\boldsymbol{f} - \boldsymbol{h}\right|\right|_2 = \sqrt{\mathbb{E}_{x \sim \{-1,1\}^n}[(\boldsymbol{f} - \boldsymbol{h})^2(x)]} \leq \boldsymbol{\epsilon}$$

Testing Sparsity in ℓ_2

Property Tester

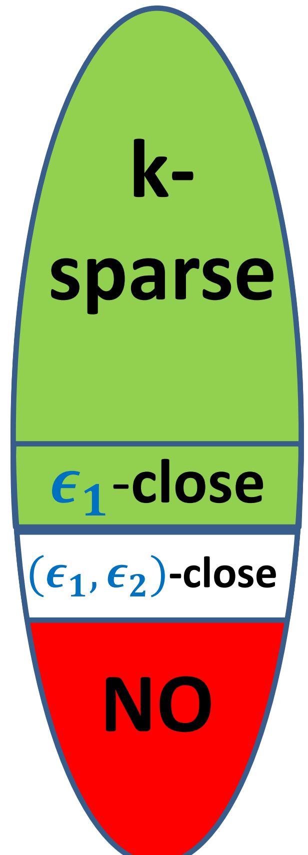


$\Rightarrow$ **Accept with probability $\geq \frac{2}{3}$**

$\Rightarrow$ **Don't care**

$\Rightarrow$ **Reject with probability $\geq \frac{2}{3}$**

Tolerant Property Tester



$\Rightarrow$ **Accept with probability $\geq \frac{2}{3}$**

$\Rightarrow$ **Don't care**

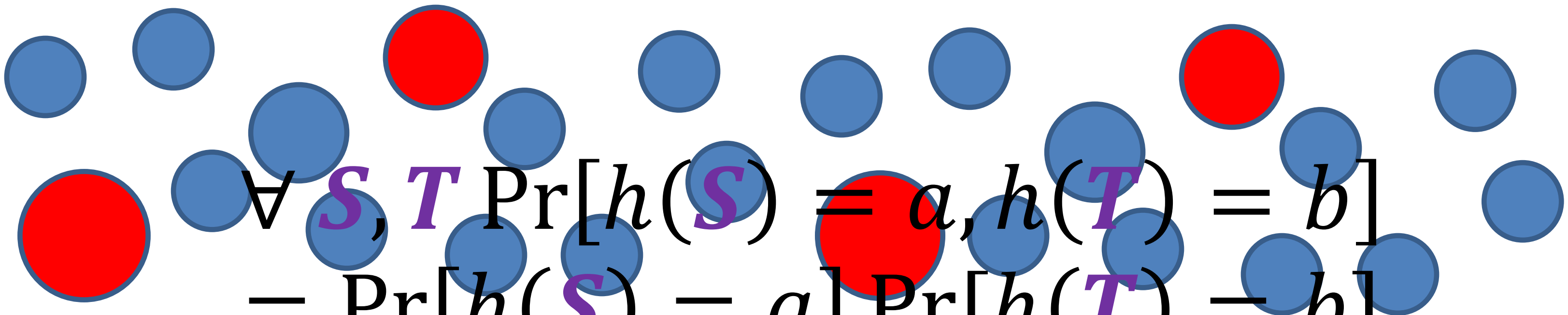
$\Rightarrow$ **Reject with probability $\geq \frac{2}{3}$**

ϵ-close : $\text{dist}(\boldsymbol{f}, \text{k-sparse}) = \inf_{g \in \text{k-sparse}} dist(\boldsymbol{f}, \boldsymbol{g}) \leq \epsilon$

Previous work under Hamming

- Testing sparsity of Boolean functions under Hamming distance
  - [Gopalan,O'Donnell,Servedio,Shiplka,Wimmer'11
    - Non-tolerant test
    - Complexity $O\left(k^{14}\log k + \frac{k^6}{\epsilon^2 \log k}\right)$
    - Reduction to testing under ℓ_2
    - Lower bound $\Omega(\sqrt{k})$
  - [Yoshida, Wimmer'13]
    - Tolerant test
    - Complexity $poly\left(k, \frac{1}{\epsilon}\right)$
- Our results give a tolerant test with almost quadratic improvement on [GOSSW'11]

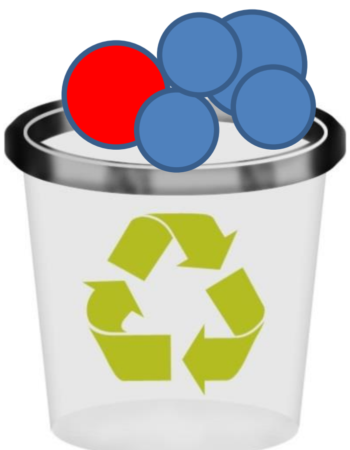
Pairwise Independent Hashing



$$\forall\, \boldsymbol{S}, \boldsymbol{T}\ \Pr[h(\boldsymbol{S}) = a, h(\boldsymbol{T}) = b]$$
$$= \Pr[h(\boldsymbol{S}) = a]\Pr[h(\boldsymbol{T}) = b]$$

Pairwise Fourier Hashing [FGKP'09]

= Cosets of a random linear subspace of $\boldsymbol{F_2^n}$



= $\boldsymbol{f_b} \equiv$ Projection of $\boldsymbol{f}$ on the coset

$$\text{"Energy"} = ||\boldsymbol{f}||_2^2 = \sum_b ||\boldsymbol{f}_b||_2^2$$

Testing k-sparsity [GOSSW'11]

\#  $= O(k^2) \Rightarrow$ 

- **Fact**: $O\left(\frac{\log\frac{1}{\delta}}{\epsilon}\right)$ random samples from $\boldsymbol{f}$ suffice to estimate $||\boldsymbol{f_b}||_2^2$ up to $\pm\boldsymbol{\epsilon}$ with prob. $\geq 1 - \boldsymbol{\delta}$
- **Algorithm:** Estimate all projections up to $\pm\,\boldsymbol{\epsilon}^2/k^4$ with probability 1 – $O\left(\frac{1}{k^2}\right)$
- **Complexity**: $O\left(\frac{k^6 \log k}{\boldsymbol{\epsilon}^4}\right)$, only non-tolerant

Our Algorithm

- Take # cosets $\mathbf{B} = O\left(\frac{k}{\epsilon^8}\right)$
- Let $\boldsymbol{m_b^i}$ be a random sample from $\boldsymbol{f_b}$
- For a coset $\boldsymbol{b}$ let $\boldsymbol{z_b} = \text{median}(\boldsymbol{m_b^1}, \ldots, \boldsymbol{m_b^u})$, where $\boldsymbol{u} = O(\log \boldsymbol{B})$
- Output $\max_{S \subseteq [t], |S| = k} \sum_{b \in S} \boldsymbol{z_b}$

- **Complexity:** $O\left(\frac{k}{\epsilon^8} \log \frac{k}{\epsilon}\right)$
- **Fact:** The "median estimators" suffice to estimate all $||\boldsymbol{f_b}||_2^2$ up to $\pm\boldsymbol{\epsilon}$

Analysis

- Take # cosets $\mathbf{B} = O\left(\frac{k}{\epsilon^8}\right)$
- Let $\boldsymbol{m_b^i}$ be a random sample from $\boldsymbol{f_b}$
- For a coset $\boldsymbol{b}$ let $\boldsymbol{z_b} = \text{median}(\boldsymbol{m_b^1}, \ldots, \boldsymbol{m_b^u})$, where $\boldsymbol{u} = O(\log \boldsymbol{t})$
- Output $\max_{S \subseteq [t], |S| = k} \sum_{b \in S} \boldsymbol{z_b}$

- Two main challenges
  - Top-k coefficients may collide
  - Noise from non top-k coefficients

Analysis: Large coefficients

Lemma: Fix $\tau = \frac{\zeta}{8k}$. If all coefficients are $\geq \tau$ then for $O\left(\frac{k}{\zeta^2}\right)$ buckets the weight in buckets with collisions $\leq \frac{\zeta}{2}$

Proof:

- \# coefficients $\leq 1/\tau$
- Pr[coefficient i collides] $\leq \frac{1}{B\tau} \leq \frac{\zeta}{4}$
- By Markov w.p. $\frac{1}{2}$ the colliding weight $\leq \frac{\zeta}{2}$

Analysis: Small coefficients

Lemma: Fix $\tau = \frac{\zeta}{8k}$. If all coefficients are $\leq \tau$ then for $O\left(\frac{k}{\zeta^2}\right)$ buckets the weight in any subset of size k is $\leq \frac{\zeta}{2}$

- "Light buckets" with weight $\leq 2\tau$ contribute $\leq \zeta/4$
- "Heavy buckets" contribute $Z = \sum_{j \in [k']} Z_j$:
  - Weighted # collisions $\mathrm{W} = \sum_b \sum_{i \neq i' \in b} w_i w_{i'}$
  - $\mathbb{E}[W] = B \sum_{i \neq i'} \frac{w_i w_{i'}}{B^2} \leq \frac{1}{B} (\sum w_i)^2 \leq \frac{1}{B}$
  - Each w_j in a "heavy bucket" Z_i contributes $\geq \frac{Z_i}{2}$ to W
  - Overall: $W \geq \frac{k'}{2} \left(\frac{Z}{k'}\right)^2 \geq \frac{Z^2}{2k} \Rightarrow Z \leq \sqrt{\frac{2k}{B}}$

Analysis: Putting it together

Lemma: If the previous two lemmas hold then the ℓ_2^2-error of the algorithm is at most $\sqrt{\zeta}$

- $\sqrt{\zeta}$ instead of ζ because of error in singleton heavy coefficients
- Crude bound because of pairwise independence + Cauchy-Schwarz

If $\zeta = O(\epsilon^4) \Rightarrow$ **B** = $O(k/\boldsymbol{\epsilon}^8)$ and ℓ_2^2-error $\boldsymbol{\epsilon}^2$

Other results + Open Problems

- Our result: $O\left(\frac{k}{\epsilon^2}\log k + \frac{1}{\epsilon^4}\right)$ non-tolerant test
  - Using BLR-test to check linearity of projections
- Lower bound of [GOSSW'11] is $\Omega(\sqrt{k})$
- Extensions to other domains
  - Sparse FFT on the line [Hassanieh, Indyk,Katabi,Price'12]
  - Sparse FFT in d dimensions [Indyk, Kapralov'14]
- Other properties that can be tested in ℓ_2?
  - Monotonicity, Lipschitzness, convexity [Berman, Raskhodnikova, Y. '14]