Lifting Linear Sketches: Optimal Bounds and Adversarial Robustness

Elena Gribelyuk[1], Honghao Lin[2], David P. Woodruff[2], Huacheng Yu[1], Samson Zhou[3]

[1] Princeton University, [2] Carnegie Mellon University, [3] Texas A&M University

Standard Streaming Model

- **Input:** Elements of an underlying frequency vector $x \in \mathbb{Z}^n$, which arrive sequentially one at a time (*worst-case*, fixed in advance).
- **Output:** At the end of the stream, $\mathcal{A}$ outputs an approximation of a given function of the stream.
- **Goal:** $\mathcal{A}$ should use space *sublinear* in the length m of the input stream and universe size n.

Linear Sketches

- A *linear sketch* is an algorithm that
  1. Samples a sketching matrix $A \in \mathbb{R}^{r\times n}$ and maintains Ax throughout the stream (typically $r \ll n$).
  2. Returns $f(Ax)$ for some estimator f.
- Lower bounds are often proven by selecting a pair of hard distributions $\mathfrak{D}_1$ and $\mathfrak{D}_2$ which exhibit a "gap" for the problem of interest.
- Then, show that $d_{TV}(Ax, Ay)$ is small when A has r rows.

Dimension Lower Bounds

- For many problems, (e.g. operator norm, norm estimation, etc), the hard distributions $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are chosen to be Gaussians (or somewhat "near" Gaussian).
- Example: For the problem of estimating $||x||_2^2$, pick $\mathfrak{D}_1 = \mathcal{N}(0, I_n)$ and $\mathfrak{D}_2 = \mathcal{N}(0, (1+\epsilon)I_n)$. WLOG, A has orthonormal rows. Then, $Ax \sim \mathcal{N}(0, I_r)$, $Ay \sim \mathcal{N}(0, (1+\epsilon)I_r)$, so A must have $r = \Omega\left(\frac{1}{\epsilon^2}\log 1/\delta)\right)$ rows to distinguish these.
- Unfortunately, none of these lower bounds translate to the streaming model!
- **Question:** *Is it possible to lift linear sketch lower bounds for continuous inputs to obtain linear sketch lower bounds for discrete inputs?*

Adversarially Robust Streaming

- **Input:** Elements of a stream, which arrive sequentially and *adversarially.*
- **Output:** At each time t, $\mathcal{A}$ receives an update u_t, updates its internal state, and returns a *current estimate* r_t , which is recorded by the *adversary*.

"Future updates may depend on previous estimates"

Adaptive Attack for Linear Sketches

- Linear sketches for F_p estimation ($p > 0$) are "not robust" to adversarial attacks, i.e. require $\Omega(n)$ dimension [HW13].
- **High-level intuition:** suppose the adversary knows the sketch matrix A: then, a hard distribution is to query $x \in \ker(A)$ or $x = 0^n$, each with probability ½.
- Thus, the adversary will aim to learn the approximate *rowspace* $R(A)$.
- Start with $V_1 = \emptyset$.
  1. Correlation finding: Find vectors weakly correlated with A orthogonal to V_{i-1}.
  2. Boosting: Use these vectors to find strongly correlated vector v.
  3. Progress: Set $V_i = \text{span}(V_{i-1}, v)$.
- Drawback: All queries are drawn from (continuous) Gaussian distributions with appropriate covariance, and the analysis heavily relies on rotational invariance. This lower bound does not directly translate to the adversarial streaming setting!
- **Question:** *Does there exist a sublinear space adversarially robust F_p estimation linear sketch in a finite precision stream?*

Main Results

Theorem (Lifting Framework): Suppose that

- $X \sim D(0, S^TS)$ and $Y \sim N(0, S^TS)$, Z is an arbitrary integer distribution
- f satisfies $\Pr_{x\sim X+Z, y\sim Y+Z}[f(x) \neq f(y)] \leq \frac{\delta}{3}$.
- $g(Ax) = f(x)$ for $x \sim X + Z$ w.p. $1 - \frac{\delta}{3}$
- $A \in \mathbb{R}^{r\times n}$ has polynomially-bounded integer entries and the singular values of S^TS are sufficiently large

Then, there is $A' \in \mathbb{R}^{4r\times n}$ and estimator h such that $h(A'y) = f(y)$ w.p. $1 - \delta$ for $y \sim Y + Z$.

Overview of our Approach

- Let $\mathcal{D}_{L,S}$ denote the discrete Gaussian distribution on support L and with covariance matrix S^TS.
- Let $x \sim \mathcal{D}_{\mathbb{Z}^n, S}$, $y \sim \mathcal{D}_{A\mathbb{Z}^n, SA^T}$, $z \sim N(0, S^TS)$

1. Show $\mathbf{d_{TV}(Ax, y)}$ is small on $A\mathbb{Z}^n$.

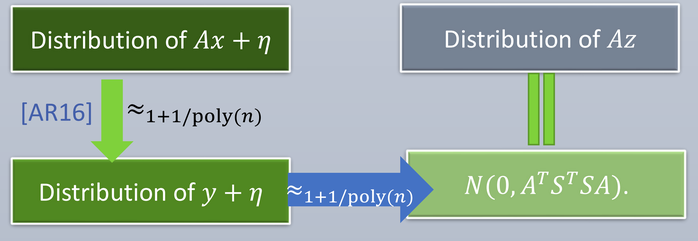
Theorem. (Sufficient condition for small distributional gap [AR16]) Suppose that $\sigma_n(S) > \lambda_{\max}(\mathcal{L}^{\perp}(A))\sqrt{\frac{\ln\left(2n\left(1+\frac{1}{\varepsilon}\right)\right)}{\pi}}$, then

$$1 - 2\varepsilon \leq \frac{\rho_{Ax}(s)}{\rho_y(s)} \leq 1 + 2\varepsilon,$$

where ρ denotes the corresponding PMF.

Design a pre-processing for A to satisfy this condition by adding $O(r)$ rows

2. Let $\boldsymbol{\eta}$ be uniform noise in unit cell of $\boldsymbol{A\mathbb{Z}^n}$



- WLOG algorithm sees $Ax + \eta$ (round in post-processing), so should also work on Az.