Adversarial Robustness on Insertion-Deletion Streams

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

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

Streaming Model

- **Input:** m updates to an underlying frequency vector $x \in \mathbb{Z}^n$, which arrive sequentially one at a time (*worst-case*, fixed in advance).
  - ➢ Update (a_t, Δ_t): $x_{a_t} \leftarrow x_{a_t} + \Delta_t$
- **Output:** At the end of the stream, A outputs an approximation of a given function of the stream.
- **Goal:** A should use space *sublinear* in the length m of the update sequence and universe size n.

Adversarially Robust Streaming

- **Input:** updates (a_t, Δ_t) to $x \in \mathbb{R}^n$, which arrive sequentially and *adversarially.*
- **Output:** At each time t, A receives an update u_t, updates its internal state, and returns a *current estimate* r_t , which is recorded by the adversary.
- **Motivation:** database queries, optimization.

"Future updates may depend on previous estimates"

Frequency Moments

We study $\boldsymbol{F_p}$ **moment estimation** for $p \in [0,2]$:

- Given a stream of updates to coordinates $i \in [n]$, let x_i denote the frequency of element i.

$$F_p = \sum_{i \in [n]} |x_i|^p$$

- **Goal:** Given a stream of m updates to x and an accuracy parameter $\varepsilon \in (0,1)$, output a $(1+\varepsilon)$-approximation to F_p, with probability $\geq 1-\delta$.
- [AMS99]: $\tilde{O}\left(\frac{1}{\varepsilon^2}\log^2 n\right)$ space algorithm for $(1+\varepsilon)$-approximation to F_p moments for $p \in [0,2]$

Previous Approaches

Upper Bounds:

- [BJWY20]: $O(m)$ bits of space.
- If $m > n$, simply store x using $\tilde{O}(n)$ space.
- [HKMMS22]: $\tilde{O}(\sqrt{m})$ bits of space.
- [BEO21]: $O(m^{0.4})$ space for F_2, $O(m^{1/3})$ space for F_0.

Lower Bounds:

- A linear sketch is an algorithm that samples a matrix $A \in \mathbb{R}^{r \times n}$ for $r \ll n$ and maintains Ax during the stream.
- In classical streaming: All optimal turnstile streaming algorithms "might as well be linear sketches" for streams of length $m = \text{poly}(n)$. [LNW14], [JLY26]
- Any multiplicative approximation algorithm for F_p based on a linear sketch over an adaptively chosen stream requires $\Omega(n)$ space [GLWYZ25]

> **Conjecture:** $\Omega(n)$ space is necessary for any adversarially robust insertion-deletion algorithm for F_p moment estimation.

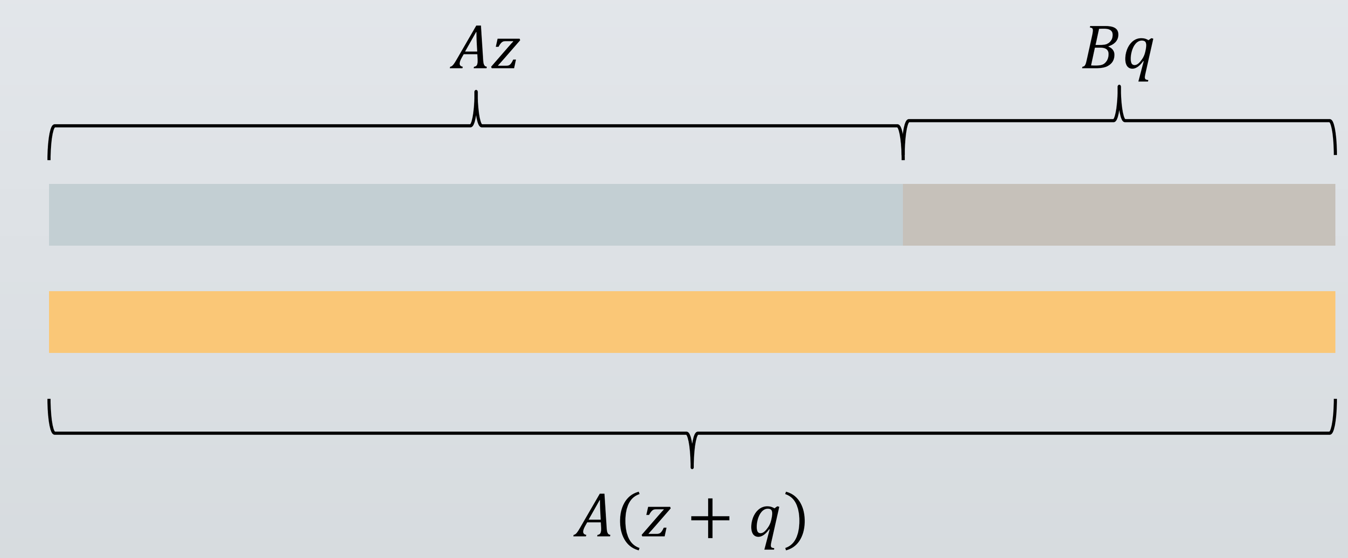
Main Contributions

- **We refute this conjecture!! We use a non-linear sketch!**
- We give the first $\text{poly}(^1/_\varepsilon, \log n)$ space algorithm for F_2 moment estimation in an adaptive insertion-deletion stream.
- Corollary: obtain $\text{poly}(^1/_\varepsilon, \log n)$ space algorithm for detecting L_2 heavy hitters, and constant factor approximation for any symmetric function with approximate triangle inequality in $o(n)$ space.
- **More recently:** we generalize our F_2 algorithm to obtain the first nearly-optimal $\text{poly}(^1/_\varepsilon, \log n)$ space algorithms for all small frequency moments $p \in [0,2]$! (In FOCS 26)
- At a high-level, F_p algorithm leverages the existence of an embedding of F_p into F_2 and *simulates* the F_2 algorithm.

Overview of our Approach

Intuition:

- To circumvent lower bounds against linear sketches, must introduce *new random bits* during the stream
- Decompose frequency vector $x = z + q$, use Az to sketch z and Bq to sketch q (A, B are independent F_2 sketches)
- Maintain Ax in the background; Goal: minimize number of adaptive interactions with sketch matrix A



Suppose $\|Az\|_2^2 + \|Bq\|_2^2$ is the estimator for $\|z+q\|_2^2$.

1. Either $\langle z, q\rangle$ is small, so that $\|x\|_2^2 \approx \|Az\|_2^2 + \|Bq\|_2^2 \approx \|z\|_2^2 + \|q\|_2^2$, or
   - ➢ Gradually learn information about Bz!
2. $\langle z, q\rangle$ is large, so q is "well-aligned" with z

Full algorithm:

- More generally, initialize $z' = 0$ and let
  $$S = \|A(z-z')\|_2^2 + \|B(z'+q)\|_2^2$$
  be the estimator. If $S \notin (1 \pm \varepsilon)\|A(z+q)\|_2^2$, update $z'' \leftarrow z' \pm \alpha(z'+q)$
- We show: when z' is updated to z'',
  $$\|z-z''\|_2^2 \leq (1-\varepsilon^2)\|z-z'\|_2^2$$

As $\|z\|_2^2 \leq \text{poly}(n)$, z' is updated $L \leq O(\frac{1}{\varepsilon^2}\log n)$ times.

- Finally, to protect B, recurse on $\|(z'+q_1) + q_2\|_2^2$ where $q = q_1 + q_2$ is the current suffix.