Sketching and Streaming Matrix Norms

David Woodruff

IBM Almaden

Based on joint works with Yi Li and Huy Nguyen

Turnstile Streaming Model

- Underlying n-dimensional vector x initialized to 0^n
- Long stream of updates $x_i \leftarrow x_i + \Delta_i$ for Δ_i in $\{-1,1\}$
- At end of the stream, x is promised to be in the set $\{-M, -M+1, \ldots, M-1, M\}^n$ for some bound $M \leq poly(n)$
- Output an approximation to f(x) whp
- Goal: use as little space (in bits) as possible

Example Problem: Norms

- Suppose you want $|x|_p^p = \Sigma_{i=1}^n |x_i|^p$
- Want Z for which $(1-\varepsilon) |x|_p^p \leq Z \leq (1+\varepsilon) |x|_p^p$
- p = 1 is Manhattan norm
  - Distances between distributions, network monitoring
- p = 2 is (squared) Euclidean norm
  - Geometry, linear algebra
- p = ∞ is max norm: $|x|_p = \max_i |x_i|$
  - denial of service attacks, etc.

Space Complexity of Norms

- For $1 \leq p \leq 2$ and constant approximation, can get log n space
- For p > 2, the space is $\tilde{\Theta}(n^{1-\frac{2}{p}})$
- Lower bound: k-party disjointness
  - k vectors $x_1, \ldots, x_k \in \{0,1\}^n$ which have disjoint supports or uniquely intersect
  - x = $\sum_i x_i$ presented in the stream in the following order: $x_1, \ldots, x_k$
  - x = (0, 1, 0, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0, 0), or
  - x = (0, 1, 0, 0, 1, 0, k, 0, 0, 1, 1, 1, 0, 1, 0, 0)
  - Set k = $2n^{1/p}$. Disjointness $\Omega(\frac{n}{k})$ communication bound gives $\Omega(\frac{n}{k^2})$ stream memory bound

Matrix Norms

- We understand vector norms very well

- Recent interest in estimating *matrix* norms

- Stream of updates to an n x n matrix A

- A initialized to $0^{n \times n}$, see updates $A_{i,j} \leftarrow A_{i,j} + \Delta_{i,j}$ for $\Delta_{i,j}$ in {-1,1}
  - Entries of A bounded in absolute value by poly(n)

- Every matrix $A = U\,\Sigma V^T$ in its singular value decomposition, where U, V have orthonormal columns and Σ is a non-negative diagonal matrix

- Schatten p-norm $|A|_p^p = \sum_i \sigma_i{}^p$ where $\sigma_i = \Sigma_{i,i}$

Matrix Norms

- Schatten p-norm $|A|_p^p = \sum_i \sigma_i{}^p$ where $\sigma_i = \Sigma_{i,i}$
  - p = 0 is the rank
  - p = 1 is the trace norm $\sum_i \sigma_i$
  - p = 2 is the Frobenius norm $\sum_{i,j} A_{i,j}^2$
  - p = ∞ is the operator norm $\sup_x \frac{|Ax|_2}{|x|_2}$

- What is the complexity of approximating $|A|_p^p = \sum_i \sigma_i{}^p$ up to a constant factor?

- For one value of p, this is easy…
  - p = 2 norm can be estimated in log n bits of space

- What about other values of p?

Matrix Norm Results

- Thoughts? Conjectures?
- An important special case: suppose A is sparse, i.e., has O(1) non-zero entries per row and per column
- There is an $\tilde{O}(n)$ upper bound for every $0 \leq p \leq \infty$
- Anything better for $p \neq 2$?

Bit lower bound for Schatten norms

	Previous lower bounds	Lower bounds in [LW16]
$p \in (2, \infty) \cap 2\mathbb{Z}$	$n^{1-2/p}$	??
$p \in (2, \infty) \setminus 2\mathbb{Z}$	$n^{1-2/p}$	$n^{1-g(\epsilon)}$
$p \in [1, 2)$	$n^{1/p-1/2}/\log n$ [AKP15]	$n^{1-g(\epsilon)}$
$p \in (0, 1)$	$\log n$ [KNP10]	$n^{1-g(\epsilon)}$
$p = 0$	$n^{1-g(\epsilon)}$ [BS15]	

- $g(\epsilon) \to 0$ as $\epsilon \to 0$
- The near-linear lower bound is tight for sparse matrices (each row and column contains $O(1)$ non-zero entries)

What about even integers p? [LW16]

- Show an $\widetilde{O}(n^{1-\frac{2}{p}})$ upper bound for every even integer p
- Matches the lower bound for vectors
- The even integer p-norms are the only norms with non-trivial space!

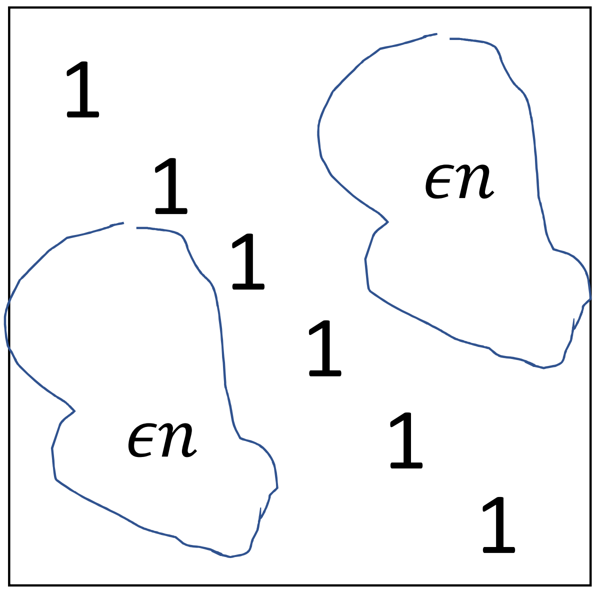
Upper Bound Intuition for p = 4

- $|A|_4^4 = \left|AA^T\right|_F^2 = \sum_{i,j} < A_i, A_j >^2$, where A_i are the rows of A

- $< A_i, A_j >^2 \leq |A_i|_2^2 \cdot \left|A_j\right|_2^2 \leq \max_{i,j} < A_i, A_i >^2$

- If $|A_i|_2^2 = 1$ for all i, then
  (1) $< A_i, A_j >^2 \leq 1$ for all i and j
  (2) if $\sum_{i \neq j} < A_i, A_j >^2 \geq \epsilon \sum_i < A_i, A_i >^2 \geq \epsilon n$

- Implies uniformly sampling $\tilde{O}(n)$ terms $< A_i, A_j >^2$ for $i \neq$ j suffices for estimating $\sum_{i \neq j} < A_i, A_j >^2$



$(1) < A_i, A_j >^2 \leq 1$ for all i,j

$$(2) \sum_{i \neq j} < A_i, A_j >^2 \geq \epsilon n$$

These conditions imply uniformly sampling $\widetilde{O}(n)$ entries works

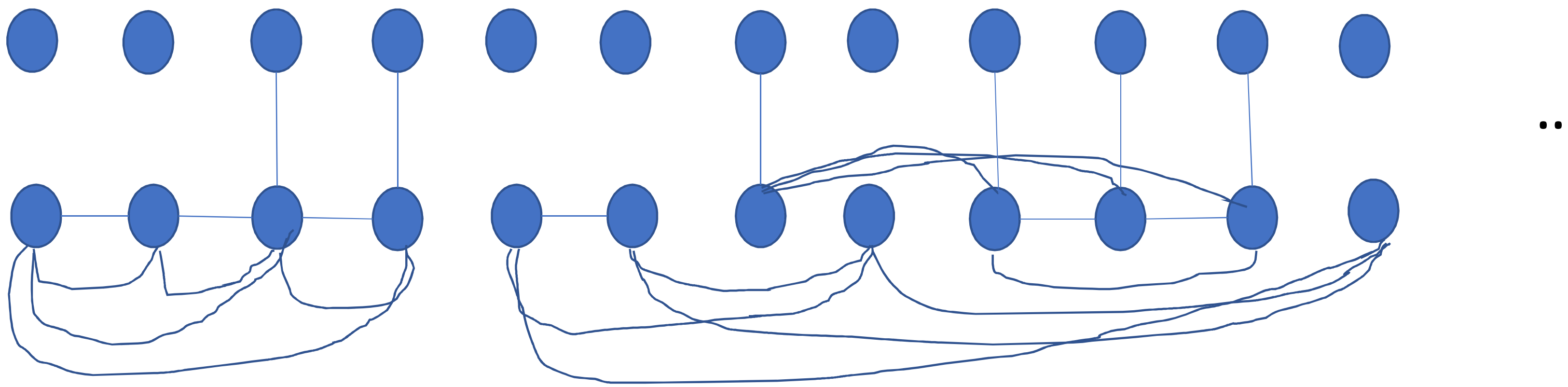
- To sample $\widetilde{O}(n)$ entries, we sample $\widetilde{O}(\sqrt{n})$ rows in their entirety (can approximately do this in a stream)
  - Can store all sampled rows using $\widetilde{O}(\sqrt{n})$ space given O(1) non-zero entries per row
  - Estimate (2) using all pairwise inner products in the sampled rows (some slight dependence issues)
- When $|A_i|_2 \neq 1$ for all i, instead sample rows proportional to $|A_i|_2^2$

Beyond p = 4

- For even integers p, let q = p/2. Then,
- $|A|_p^p = \sum_{1 \le i_1, i_2, \ldots, i_q \le n} \prod_{j=1,\ldots,q} < A_{i_j}, A_{i_{j+1}} >$, where $i_{q+1} = i_1$
- Sample $\widetilde{O}(n^{1-\frac{2}{p}})$ rows in their entirety proportional to their squared norm
- Approximate above sum by summing over all q-tuples from your sample
- For non-even integers p and p = 0, no such expression for $|A|_p^p$ exists!

Lower bound for $p = 0$ [BS15]

- Hidden Boolean Hypermatching Problem ([VY11], [BS15])
  - Alice has a boolean vector $x \in \{0,1\}^n$ such that $w_H(x) = n/2$
  - Bob has a perfect t-hypermatching M of n/t edges, each edge has t nodes
  - Determine whether $Mx := \left(\bigoplus_{i=1}^{t} x_{M_{1,i}}, \ldots, \bigoplus_{i=1}^{t} x_{M_{n/t,i}}\right)$ is $\mathbf{1}$ or $\mathbf{0}$
- $\Omega(n^{1-1/t})$ bits for one-way communication



- 2n nodes
- Create a t-clique for each hyperedge in Bob's input
- Add 'tentacles' according to Alice's input x
- Determine whether all cliques have an even or odd number of tentacles
- Maximum matching size different by a constant factor in the cases
- If clique size is t, then with r tentacles, block matching size is r + $\lfloor\frac{t-r}{2}\rfloor$
- Matching size is 3n/4 if r are all even, Matching size is 3n/4-n/(2t) if r are all odd

Connection with Matrices

- Consider the Tutte matrix A of the graph
  - $A_{i,j}$ = 0 if {i,j} is not an edge
  - $A_{i,j}$ = $y_{i,j}$ if {i,j} is an edge and i < j
  - $A_{i,j}$ = $-y_{i,j}$ if {i,j} is an edge and j < i

- rank(A), under random assignment to the $y_{i,j}$, is twice the maximum matching size, with high probability

- $\Omega(n^{1-\frac{1}{t}})$ lower bound for $(1+\Theta\left(\frac{1}{t}\right))$-approximation

Distributional BHH Problem

- Distributional BHH [VY11]: Alice get a uniformly random x in $\{0,1\}^n$, and Bob an independent, uniformly random perfect t-hyper-matching M on the n coordinates and a binary string w in $\{0,1\}^{n/t}$. Promise: $Mx \oplus w = 1^{n/t}$ or $Mx \oplus w = 0^{n/t}$

- Let t be even. Distributional BHH problem [BS15]:
  - Replace x with new input x ←(x, $\bar{x}$)
  - For i-th set $S = \{x_{i_1}, \ldots, x_{i_t}\} \in$ M,
    - if $w_i = 0$, include $\{x_{i_1}, \ldots, x_{i_t}\}$ and $\{\overline{x_{i_1}}, \ldots \overline{x_{i_t}}\}$ in new input M
    - if $w_i = 1$, include $\{\overline{x_{i_1}}, x_{i_2}, \ldots, x_{i_t}\}$ and $\{x_{i_1}, \overline{x_{i_2}}, \overline{x_{i_3}}, \ldots, \overline{x_{i_t}}\}$ in the new input M
  - Correctness is preserved, and $Mx = 1^{n/t}$ or $Mx = 0^{n/t}$
  - In graph, can partition t-cliques into pairs: in each pair number of tentacles is q and t-q, for a binomially distributed odd (even) integer q if $Mx = 1^{n/t}$ (if $Mx = 0^{n/t}$)

Distributional BHH Problem

- Consider Tutte matrix A with diagonal 0 and indeterminates equal to 1
- After permuting rows and columns, A is block-diagonal
- Each block is (2t) x (2t) and corresponds to a clique with tentacles
- t = 4 and the three possible blocks for an even number of tentacles:

$$B_0 = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad B_2 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & 0 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad B_4 = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 & 0 & 1 & 0 \\ -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Distribution of Singular Values

- $|A|_p^p = \sum_{\text{blocks B in A}} |B|_p^p$
- Suppose $E_{q \sim E(t)}\left[\left|B_q\right|_p^p\right] \neq E_{q \sim O(t)}[\left|B_q\right|_p^p]$
  - E(t) is distribution on even integers q with Pr[q = i] = {t choose i}/2^{t-1}
  - O(t) is distribution on odd integers q with Pr[q = i] = {t choose i}/2^{t-1}
- Since blocks B are of constant size, and pairs of blocks are independent, by Hoeffding bounds $|A|_p^p$ differs by a constant factor if $Mx = 1^{n/t}$ or if $Mx = 0^{n/t}$
- Suffices to show $E_{q \sim E(t)}\left[\left|B_q\right|_p^p\right] \neq E_{q \sim O(t)}[\left|B_q\right|_p^p]$!

$\Omega(\sqrt{n})$ lower bound for $p \neq 2$

- $\Omega(n^{1-1/t}) = \Omega(\sqrt{n}) \;\Rightarrow\; t = 2$

$$M_0 = \left(\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad M_1 = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ \hline -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right), \quad M_2 = \left(\begin{array}{cc|cc} 0 & 1 & 1 & 0 \\ -1 & 0 & 0 & 1 \\ \hline -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{array}\right)$$

$$\{1, 1\} \qquad \{\sqrt{2}, \sqrt{2}\} \qquad \{\tfrac{1+\sqrt{5}}{2}, \tfrac{1+\sqrt{5}}{2}, \tfrac{\sqrt{5}-1}{2}, \tfrac{\sqrt{5}-1}{2}\}$$

- $\mathbb{E}\|M_k\|_p^p$ in even and odd cases:

$$\frac{1}{2}\left\{2 \cdot 1^p + 2\left(\left(\frac{\sqrt{5}+1}{2}\right)^p + \left(\frac{\sqrt{5}-1}{2}\right)^p\right)\right\} \quad \neq \quad 2\sqrt{2}^p$$

- $\Omega(\sqrt{n})$ lower bound follows.

$n^{1-g(\epsilon)}$ Lower Bound for p not an Even Integer

- Just need to show $E_{q\sim E(t)}\left[\left|B_q\right|_p^p\right] \neq E_{q\sim O(t)}[\left|B_q\right|_p^p]$
- Change the definition of blocks B_q to make analysis tractable
- Singular values are either 1 or roots of a quadratic equation depending on q
- Analysis uses power series expansion of the roots and hypergeometric polynomials

Conclusions and Future Directions

- Nearly tight bounds for sparse matrices for matrix norms for every p
- For dense matrices, for p = 0 there is an $n^{2-g(\epsilon)}$ lower bound [AKL17]
- Nothing better known for other values of p for dense matrices
- When the streaming algorithm is a linear sketch:
  - Not clear if these lower bounds imply lower bounds for streams (though would be surprising if not)
  - $n^{2-4/p}$ bound for every $p \geq 2$, tight for even integers [LNW14,LW16]
  - For p not an even integer, conjecture an $n^{2-g(\epsilon)}$ lower bound