Numerical Linear Algebra in the Streaming Model

David Woodruff
IBM Almaden
Data Streams

- A data stream is a sequence of data, that is too large to be stored in available memory

- Examples
  - Internet search logs
  - Network Traffic
  - Sensor networks
  - Scientific data streams (astronomical, genomics, physical simulations)
Data Stream Models

• Underlying object an \( n \times d \) matrix \( A \)

• Row-Insertion Model
  – See rows (or columns) of \( A \) one at a time in an arbitrary order
  – E.g., document/term entries

• Turnstile Model
  – See entries of \( A \) one at a time in an arbitrary order
  – E.g., customer/item entries
  – Stream may be a long interleaved sequence of arbitrary additive updates \( A_{i,j} \leftarrow A_{i,j} + \Delta \) to entries

• Goals:
  – 1 pass (or small number of passes) over the data
  – Low space complexity
  – Fast processing time per update
Linear Algebra Problems

• **Approximate Matrix Product**
  – Given matrices A and B, approximate A*B

• **Regression**
  – Given a matrix A and a vector b, find an x which approximately minimizes |Ax-b|
  – Least squares, least absolute deviation, M-estimators

• **Low Rank Approximation**
  – Given a matrix A, find a rank-k matrix A’ for which |A’-A| is as small as possible
  – Frobenius, spectral, robust

• **Leverage Score Approximation**
  – Given a matrix A, if A = Q*R where Q has orthonormal columns, estimate |Q_i,*|^2 for all rows i
  – Sampling based algorithms
Linear Algebra Problems Con’d

• **Sketching norms**
  – Given a matrix $A$, approximate its trace, Frobenius, and operator norms
  – Lower bounds imply lower bounds for harder problems, such as low rank approximation in spectral norm

• **Graph sparsification**
  – Given the Laplacian $L$ of a graph $G$, approximate the quadratic form $x^T L x$ for all vectors $x$
  – Approximately preserve all cut values
Talk Outline

• Overview of techniques
  – Oblivious Subspace Embeddings
  – Leverage Score Sampling

• Sample of known results for linear algebra problems

• Open problems
Example Sketching Technique: Least squares regression [S]

• Suppose $A$ is an $n \times d$ matrix with $n \leq d$.

• How to find an approximate solution $x$ to $\min_x |Ax-b|_2$?

• **Goal:** output $x'$ for which $|Ax'-b|_2 \cdot (1+\varepsilon) \min_x |Ax-b|_2$ w.h.p.

• Draw $S$ from a $k \times n$ random family of matrices, for $k \leq n$

• Compute $S^*A$ and $S^*b$. Output solution $x'$ to $\min_{x'} |(S A)x-(S b)|_2$

• Streaming implementation: maintain $S^*A$ and $S^*b$
How to choose the right sketching matrix $S$?

- Recall: output the solution $x'$ to $\min_{x'} ||(SA)x-(Sb)||_2$
- Lots of matrices work
- $S$ is $d/\epsilon^2 \times n$ matrix of i.i.d. Normal random variables
- Computing $S*A$ may be slow…
Fast JL [AC, S]

• S is a Fast Johnson Lindenstrauss Transform

  – $S = P\times H\times D$

  – D is a diagonal matrix with +1, -1 on diagonals

  – H is the Hadamard transform

  – P just chooses a random (small) subset of rows of H*D

  – S*A can be computed much faster

• In a stream, useful if you see one column of A at a time
Even faster sketching matrices $S$ 
[CW, MM, NN]

- CountSketch matrix

- Define $k \times n$ matrix $S$, for $k = \frac{d^2}{\varepsilon^2}$

- $S$ is really sparse: single randomly chosen non-zero entry per column

\[
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 1 & 0 & -1 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
\]

- Easy to maintain in a stream

Surprisingly, this works!
Leverage Score Sampling [DMM]

- Main reason sketching works is
  - $|S(Ax-b)|_2 = (1\pm\epsilon) |Ax-b|_2$ for all $x$ in $\mathbb{R}^d$
  - $S$ is a **subspace embedding** for column span of $[A, b]$

- Leverage score sampling also provides a subspace embedding
  - If $[A, b] = Q*R$ where $Q$ has orthonormal columns, sample row $i$ of $[A, b]$ w.pr. $\gg |Q_{i,*}|_2^2$ for all rows $i$

  - Let $S$ implement sampling of $d \log d / \epsilon^2$ rows of $A$. $|S(Ax-b)|_2 = (1\pm\epsilon) |Ax-b|_2$ for all $x$ in $\mathbb{R}^d$

  - Gives a coreset, not directly implementable in a stream, but possible
Talk Outline

• Overview of techniques
  – Oblivious Subspace Embeddings
  – Leverage Score Sampling

• Sample of known results for linear algebra problems

• Open problems
Regression

- Least Squares Regression [CW, MM, NN]
  - $\mathcal{O}(d^2 / \epsilon)$ space in a stream, $O(1)$ update time

- Least Absolute Deviation Regression [SW]
  - $\text{poly}(d/\epsilon)$ space in a stream, $O(\sim 1)$ update time
Low Rank Approximation [S,CW]

• A is an n x n matrix

• Want to output a rank k matrix A’, so that w.h.p.,
  $\|A-A’\|_F \cdot (1+\epsilon) \|A-A_k\|_F$
  where $A_k$ is the best rank-k approximation to A

• $O(\sim (n/poly(\epsilon)))$ space in a stream, $O(1)$ update time
Matrix Norms in A Stream [LNW]

- A is an n x n matrix

- $p$-th Schatten norm is $\sum_{i=1}^{\text{rank}(A)} \sigma_{i}^{p}(A)$

- $p = 2$ is the Frobenius norm
  - $O(\sim(1))$ space in a stream, $O(1)$ update time

- $p = 1$ is trace norm
  - $\Omega(n^{1/2})$ space in a stream, no nontrivial upper bound!

- $p = 1$ is the operator norm $\max_{\text{unit } x, y} x^{T}Ay$
  - $\Omega(n^2)$ space in a stream
  - Same lower bound for operator norm low rank approximation
Graph Sparsification [KLMMS]

- Given graph $G$, let $H$ be a subgraph with reweighted edges.
- Let $L_G$ be the Laplacian of $G$ and $L_H$ be the Laplacian of $H$.
- Want $x^T L_H x = (1 \pm \epsilon) x^T L_G x$ for all $x$.
- $O(\frac{n}{\epsilon^2})$ space in a stream of edges possible.
- Clever recursive leverage score sampling in a stream [MP].
Open Problems

• Optimal bounds in terms of $\varepsilon$ in streaming model
  – Tradeoff with number of passes

• Spectral low rank approximation not possible in a stream, but maybe can get $O(\text{nnz}(A))$ time offline?
  – Current best $\text{nnz}(A) \text{ poly}(k/\varepsilon)$

• Robust low rank approximation:
  Output a rank k matrix $A'$, so that
  $|A-A'|_1 \cdot (1+\varepsilon) |A-A_k|_1$