Fast JL Transform
(and everything else)

Piotr Indyk
MIT
What else is there in streaming/sketching?

- Fast JL transform
- AMS sampling (and entropy algorithm)
- Lp sampling
- Parsing (Dyck languages)
- Random order streams
- Model-based compressed sensing
Fast FL Transform
Fast JL Transform

• **JL Lemma:**
  – Let $A$ be a $k \times n$ matrix, with i.i.d. entries chosen from $N(0,1)$ (or other distributions)
  – Then, for any $x \in \mathbb{R}^n$, we have
    \[
    Q = \Pr\left[ | \|Ax\|_2^2 - k\|x\|^2 | > \varepsilon k\|x\|^2 \right] \leq \exp(-C \varepsilon^2 k)
    \]

• Alternatively: $k = O(\log(1/Q)/\varepsilon^2)$
• Takes $O(nk)$ time to compute
• Can we do it faster?
[Ailon-Chazelle’06]

• Compute

\[ Ax = P H D x \]

where:

– **D**: diagonal matrix, with \( D_{ii} \in \mathbb{R}\{-1,1\} \) i.i.d.
– **H**: normalized Hadamard matrix
  • Entries in \( \{-1/n^{1/2}, 1/n^{1/2}\} \)
  • Columns orthogonal
– **P**: a random projection matrix
  • In the simplest case, \( P=P' P'' \), where
    – \( P'' \) projects the vector \( H D x \) on \( l \) random coordinates, chosen with replacement
    – To get optimal reduced dimension, apply standard JL matrix \( P' \) to the \( l \)-dimensional vector \( P'' H D x \)

• Total time: \( O(n \log n + lk) \) time
  – As long as \( l < n \) and \( k > \log n \), we get an improvement in the running time
  – Can set \( l = O(\log(1/Q)^2/\epsilon^2) \)
Improvements

• [Ailon-Liberty, June‘10]: keep only $P=P''$, i.e., the matrix $A=PHD$ is $l \times n$
  – $l=\log(1/Q) \cdot (\log n + 1/\varepsilon)^{O(1)}$ suffices to get probability of failure $<Q$

• [Krahmer-Ward, September‘10]
  – $l=\log(1/Q)/\varepsilon^2 \cdot (\log n)^{O(1)}$
  – Any matrix $H$ satisfying RIP of order $k=O(\log(1/Q)/\varepsilon^2)$ suffices
Improvements II

• Can make the original JL sparse:
  – [Dasgupta-Kumar-Sarlos’10]:
    about $O(\log(1/Q)^3/\varepsilon)$ non-zeros per column
  – [Kane-Nelson’10], [Braverman-Ostrovski-Rabani’10]:
    about $O(\log(1/Q)^3/\varepsilon)$ non-zeros per column
  – [Kane-Nelson’10]:
    $O(\log(1/Q)/\varepsilon)$ non-zeros per column

• Note the last entry is always not greater than $k=O(\log(1/Q)/\varepsilon^2)$
AMS Sampling and extensions
Old algorithm for $L_k$ estimation under insertions only

- Let $F_k = \sum_{i=1}^{m} x_i^k = ||x||^k_k$ where $x$ defined by the stream $i_1...i_n$

- Algorithm A: two passes
  - Pass 1: Pick a stream element $i=i_j$ uniformly at random
    (i.e., sample $i$ with prob. $|x_i|/||x||_1 = |x_i|/n$)
  - Pass 2: Compute $x_i$
  - Return $Y = n x_i^{k-1}$

- Alternative view:
  - Little birdy that samples $i$ and returns $x_i$
    (Sublinear-Time Algorithms lectures)
Analysis

• Estimator \( Y = n x_i^{k-1} \)
• Expectation
  \[
  E[Y] = \frac{\sum_i x_i}{n} * n x_i^{k-1} = \sum_i x_i^k = F_k
  \]
• Second moment (\( \geq \) variance)
  \[
  E[Y^2] = \frac{\sum_i x_i}{n} * n^2 x_i^{2k-2} = n \sum_i x_i^{2k-1} = n F_{2k-1}
  \]
• Claim:
  \[
  n F_{2k-1} \leq m^{1-1/k} (F_k)^2
  \]
• Therefore, averaging over \( O(m^{1-1/k}/\epsilon^2) \) samples + Chebyshev does the job
Extensions

• One pass sampling
  – Pick \( i = i_j \) uniformly at random from the stream
  – Compute \( r = \# \text{occurrences of } i \) in the \textbf{reminder} of the stream
  – Estimator: \( Y' = n \ (r^k - (r-1)^k) \)

• Can solve other problems
  – Entropy \( \sum_i x_i / n \ \log(x_i / n) \) in \( O(\log n/\varepsilon^2) \) words of space [Chakrabarti-Cormode-McGregor’08]

• Insertions and deletions
  – Can sample from distribution close to \( |x_i| / ||x||_1 \) using \( (\log n)^{O(1)} \) space [Thatchar-Woodruff’09]
Parsing expressions
Parsing expressions

• Problem:
  – A stream of $n$ parenthesis ($T$ types)
  – Is it well-formed

• Cases:
  – $T=1$:
    • Can check using one counter
  – $T=2$ [Magniez-Mathieu-Nayak’10]
    • $O(n^{1/2})$ space sufficient and necessary
    • But $(\log n)^{O(1)}$ space if allowed additional pass in reverse
  – $T>2$:
    • Reduce to $T=2$ with $O(\log T)$ overhead
Random order streams
Random order streams

- We have seen how to find \((1/2 \pm \varepsilon)\) - quantile in \(O(1/\varepsilon \log n)\) space
- Gives exact median in \(\log n\) passes using \((\log n)^{O(1)}\) space

- What if the stream elements come in random order?
  - Can find \((1/2 \pm 1/n^a)\)-quantile in one pass
  - Therefore, can find exact median in \(O(\log \log n)\) passes [Guha-MacGregor’08]
Model-based compressed sensing

- We have seen sketches with
  - Guarantee: (1) \( \|x-x^*\|_1 \leq C \min_{k\text{-sparse } x''} \|x-x''\|_1 \)
  - Sketch length \( m = \mathcal{O}(k \log (n/k)) \)
- Can we get shorter sketches?
  - Not in general [DoBa-Indyk-Price-Woodruff’10]
- An alternative
  \( \|x-x^*\|_1 \leq C \min_{\text{supp}(x'')} \text{ in } M \) \( \|x-x''\|_1 \)
- Example model: \( M = I \) blocks, each of length \( b \)
  - \( m = \mathcal{O}(lb + I \log n) \) with superconstant \( C \) (deterministic) [Baraniuk-Cevher-Duarte-Hedge’10]
  - \( m = \mathcal{O}(lb + I \log N) \), constant \( C \) (randomized) [Price’11]