On Deterministic Sketching and Streaming for Sparse Recovery and Norm Estimation

Jelani Nelson
Princeton

joint work with Huy Nguyễn (Princeton) and David Woodruff (IBM Almaden)

The turnstile model of streaming

- Vector $x \in \mathbb{R}^n$ starts as $\vec{0}$
- Receive a sequence of updates $(i_1, v_1), (i_2, v_2), \ldots$ with $(i, v) \in [n] \times \mathbb{R}$
- At end of stream, output some function $f(x)$
- Goal: Use very little memory, e.g. $\log^{O(1)} n$ words

The turnstile model of streaming

- Vector $x \in \mathbb{R}^n$ starts as $\vec{0}$
- Receive a sequence of updates $(i_1, v_1), (i_2, v_2), \ldots$ with $(i, v) \in [n] \times \mathbb{R}$
- At end of stream, output some function $f(x)$
- Goal: Use very little memory, e.g. $\log^{O(1)} n$ words
- This talk: We will focus on *linear sketches*, i.e. algorithms that maintain Ax for some A with $m \ll n$ rows

  (Natural model in some of the compressed sensing applications we will consider. Also, can easily combine linear sketches across datasets to get aggregate statistics or compute distance measures between data.)

Goal of this talk

Revisit classic sketching/streaming problems with the goal of understanding their deterministic complexities

- Point query
- Inner product
- ℓ_1/ℓ_1 sparse recovery
- Norm estimation

Goal of this talk

Revisit classic sketching/streaming problems with the goal of understanding their deterministic complexities

- Point query: Given i, output $x_i \pm \varepsilon \|x\|_1$
- Inner product: Given Ax, Ay, output $\langle x, y \rangle \pm \varepsilon \|x\|_2 \|y\|_2$
- ℓ_1/ℓ_1 sparse recovery: Output $\tilde{x}$ with $\|x - \tilde{x}\|_1 \leq (1+\varepsilon)\|x_{tail(k)}\|_1$
- Norm estimation: Output $(1 \pm \varepsilon)\|x\|_p$

Previous work

- Point query: Given i, output $x_i \pm \varepsilon \|x\|_1$

  Randomized: CountMin [CM05] $m = O(\varepsilon^{-1} \log(1/\delta))$, $m = \Omega(\varepsilon^{-1} \log n)$ for $\delta < 1/n$ [JST11]

  Deterministic: CR-PRECIS [GM07] $m = O(\varepsilon^{-2} \log^2 n / (\log(\varepsilon^{-1} \log n)))$, $m = \Omega(\varepsilon^{-2} + \varepsilon^{-1} \log(\varepsilon n))$ [FPRU10], [Ganguly08], [Gluskin82]

Previous work

- Inner product: Given Ax, Ay, output $\langle x, y\rangle \pm \varepsilon\|x\|_2\|y\|_2$

  Randomized: AMS sketch [AMS96] $m = O(1/\varepsilon^2)$, $m = \Omega(1/\varepsilon^2)$ [KNW10]

  Deterministic: Impossible [AMS96]

Previous work

- ℓ_1/ℓ_1 sparse recovery: Output $\tilde{x}$ with $\|x - \tilde{x}\|_1 \leq (1+\varepsilon)\|x_{tail(k)}\|_1$

  Randomized: $m = O(k \log n \log^3(1/\varepsilon)/\sqrt{\varepsilon})$ [PW11]

  Deterministic: $m = O(k \log(n/k)/\varepsilon^2)$ [IR08]

Previous work

- Norm estimation: Output $(1 \pm \varepsilon)\|x\|_p$

  Randomized: $(0 < p \leq 2)$ $m = O(1/\varepsilon^2)$ [Indyk00], $m = \Omega(1/\varepsilon^2)$ [KNW10]

  $(p > 2)$ $m = \tilde{O}(\varepsilon^{-2} n^{1-2/p})$ [IW05], [BGKS06], [AKO11], [Ganguly11], $m = \tilde{\Omega}(\varepsilon^{-2} n^{1-2/p})$ [BJKS02], [CKS03], [Gronemeier09], [Jayram09], [WZ12], [Ganguly11]

  Deterministic: Impossible for all p [AMS96]

Our contributions

We improve some upper and lower bounds in the deterministic case

- Point query

  $m = O(\varepsilon^{-2} \cdot \min\{\log n, (\log(n)/\log(1/\varepsilon))^2\})$. Can even get $x_i \pm \varepsilon \|x_{tail(1/\varepsilon^2)}\|_1$ with almost same m.
- Inner product

  Equivalent to point query if we allow $\langle x, y\rangle \pm \varepsilon \|x\|_1 \|y\|_1$
- ℓ_1/ℓ_1 sparse recovery

  We prove $m = \Omega(k/\varepsilon^2 + k\log(n/k)/\varepsilon)$
- Norm estimation

  Settle for $\|x\|_p \pm \varepsilon \|x\|_q$ $(q < p)$. We show the complexity is then characterized by *Gelfand widths* from convex geometry, with the tight bound $m = \Theta(\varepsilon^{-2}\log(\varepsilon^2 n))$ for $p = 2, q = 1$.

Point query/inner product

Reductions between the problems

- Inner product $\Rightarrow$ point query

  Output the estimated inner product of x and e_i. Error is $\varepsilon\|x\|_1\|e_i\|_1 = \varepsilon\|x\|_1$.

- Point query $\Rightarrow$ inner product

  Let $\tilde{x}, \tilde{y}$ be such that $\|x - \tilde{x}\|_\infty \leq \varepsilon\|x\|_1$ and similarly for y. Output $\left\langle \tilde{x}_{head(1/\varepsilon)}, \tilde{y}_{head(1/\varepsilon)} \right\rangle$.

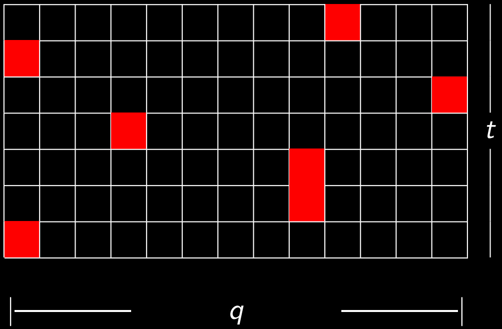
Point query and codes [GM07]

Let $\mathcal{C} = \{C_1, \ldots, C_n\}$ be a code with block length t, alphabet $[q]$, and relative distance $1 - \varepsilon$.

$h_1(i) = (C_1)_i$

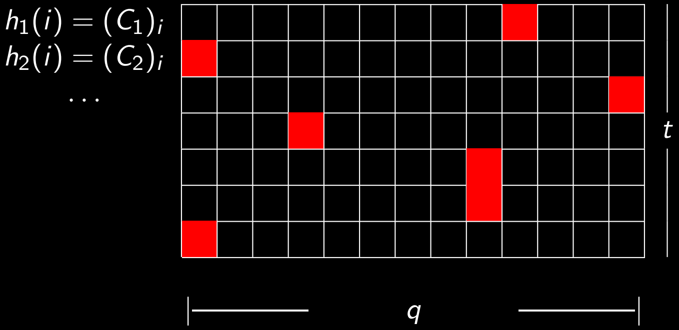
$h_2(i) = (C_2)_i$

$\ldots$



Point query and codes [GM07]

Let $\mathcal{C} = \{C_1, \ldots, C_n\}$ be a code with block length t, alphabet $[q]$, and relative distance $1 - \varepsilon$.



$\tilde{x}_i$ is average of $Y_{j,h_j(i)}$. Equals
$x_i + \sum_{j \neq i}(\#\text{collisions with } i) \cdot x_j/t = x_i \pm \varepsilon \|x_{-i}\|_1$.
[GM07] uses *Chinese remaindering codes*.

Point query and incoherent matrices

A is *incoherent* if each column has unit ℓ_2 norm and $|\langle A_i, A_j \rangle| \leq \varepsilon$ for all $i \neq j$.

- Measurement: Ax
- Recovery: $\tilde{x} = A^T Ax$

Point query and incoherent matrices

A is *incoherent* if each column has unit ℓ_2 norm and $|\langle A_i, A_j \rangle| \leq \varepsilon$ for all $i \neq j$.

- Measurement: Ax
- Recovery: $\tilde{x} = A^T Ax$

  Proof: $(A^T Ax)_i = A_i^T Ax = \sum_{j=1}^{n} \langle A_i, A_j \rangle x_j = x_i \pm \varepsilon \|x_{-i}\|_1$.

How to get an incoherent matrix

- Johnson-Lindenstrauss Given any $v_1, \ldots, v_N$ can find A with $O(\varepsilon^{-2} \log N)$ rows so that $\|Av_i - Av_j\|_2 = (1 \pm \varepsilon)\|v_i - v_j\|_2$ for all i, j. If vectors are $0, e_1, \ldots, e_n$, this implies A is incoherent. Using Fast JL [Ailon-Chazelle'06], ... can have A, A^T with fast multiplication times.

How to get an incoherent matrix

- Johnson-Lindenstrauss Given any $v_1, \ldots, v_N$ can find A with $O(\varepsilon^{-2} \log N)$ rows so that $\|Av_i - Av_j\|_2 = (1 \pm \varepsilon)\|v_i - v_j\|_2$ for all i, j. If vectors are $0, e_1, \ldots, e_n$, this implies A is incoherent. Using Fast JL [Ailon-Chazelle'06], ... can have A, A^T with fast multiplication times.
- Codes Given q-ary code $C_1, \ldots, C_n$ with block length t and relative distance $1 - \varepsilon$. Form matrix with n columns and $m = qt$ rows. A_i broken up into t blocks of size q, with 1s indicating what symbol C_i has in each position and 0s elsewhere. [GM07] used *chinese-remainder codes*, but both Reed-Solomon codes and random codes do better.

How to get an incoherent matrix

- Johnson-Lindenstrauss Given any $v_1, \ldots, v_N$ can find A with $O(\varepsilon^{-2} \log N)$ rows so that $\|Av_i - Av_j\|_2 = (1 \pm \varepsilon)\|v_i - v_j\|_2$ for all i, j. If vectors are $0, e_1, \ldots, e_n$, this implies A is incoherent. Using Fast JL [Ailon-Chazelle'06], ... can have A, A^T with fast multiplication times.
- Codes Given q-ary code $C_1, \ldots, C_n$ with block length t and relative distance $1 - \varepsilon$. Form matrix with n columns and $m = qt$ rows. A_i broken up into t blocks of size q, with 1s indicating what symbol C_i has in each position and 0s elsewhere. [GM07] used *chinese-remainder codes*, but both Reed-Solomon codes and random codes do better.
- Almost pairwise independence A set $S \subseteq \{-1, 1\}^n$ such that $|\mathbb{E}_{x \in S} \prod_{i \in T} x_i| \leq \varepsilon$ for all $|T| = 1, 2$. Form a $|S| \times n$ matrix where rows are elements of S scaled down by $\sqrt{|S|}$.

ℓ_1/ℓ_1 sparse recovery

ℓ_1/ℓ_1 sparse recovery lower bound

- Show $m = \Omega(k/\varepsilon^2)$ via the probabilistic method (based on an argument of Gluskin, rediscovered by Ganguly)
- Show $m = \Omega(k \log(n/k)/\varepsilon)$ via communication complexity

$m = \Omega(k/\varepsilon^2)$

- Recall we must produce $\tilde{x}$ with $\|x - \tilde{x}\|_1 \leq (1+\varepsilon)\|x_{tail(k)}\|_1$
- Thus if $x = 0$, we must output $\tilde{x} = 0$.

  The plan: Show that there exists $x \in \ker(A)$ such that $\|x_{head(k)}\|_1 > \varepsilon\|x_{tail(k)}\|_1$, i.e. 0 is not an acceptable output.

$m = \Omega(k/\varepsilon^2)$

- Recall we must produce $\tilde{x}$ with $\|x - \tilde{x}\|_1 \le (1+\varepsilon)\|x_{tail(k)}\|_1$
- Thus if $x = 0$, we must output $\tilde{x} = 0$.

  The plan: Show that there exists $x \in \ker(A)$ such that $\|x_{head(k)}\|_1 > \varepsilon\|x_{tail(k)}\|_1$, i.e. 0 is not an acceptable output.
- Without loss of generality A has orthonormal rows.

$m = \Omega(k/\varepsilon^2)$

- Recall we must produce $\tilde{x}$ with $\|x - \tilde{x}\|_1 \le (1+\varepsilon)\|x_{tail(k)}\|_1$
- Thus if $x = 0$, we must output $\tilde{x} = 0$.

  The plan: Show that there exists $x \in \ker(A)$ such that $\|x_{head(k)}\|_1 > \varepsilon\|x_{tail(k)}\|_1$, i.e. 0 is not an acceptable output.
- Without loss of generality A has orthonormal rows.
- The bad x: Note $x = (I - A^T A)y$ is in $\ker(A)$ for any y. Choose y randomly as $\sum_{i=1}^k \sigma_i e_{\pi(i)}$ where π is a random permutation and σ_i are independent random signs. We show $\|x_{head(k)}\|_1 > \varepsilon\|x_{tail(k)}\|_1$ with positive probability.

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\mathbb{E}\|x_{tail}(k)\|_1 \quad \leq \quad \mathbb{E}\|x\|_1$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\begin{aligned} \mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\ &\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \end{aligned}$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\begin{aligned} \mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\ &\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\ &\leq k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \end{aligned}$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\begin{aligned}
\mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\
&\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\
&\leq k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A A^T A y\right)^{1/2}
\end{aligned}$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\begin{aligned}
\mathbb{E}\|x_{tail}(k)\|_1 &\le \mathbb{E}\|x\|_1 \\
&\le \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\
&\le k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A y\right)^{1/2}
\end{aligned}$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$
\begin{aligned}
\mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\
&\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\
&\leq k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} \left\langle \sum_{j=1}^{k} \sigma_j A_{\pi(j)}, \sum_{j=1}^{k} \sigma_j A_{\pi(j)} \right\rangle\right)^{1/2}
\end{aligned}
$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$\begin{aligned}
\mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\
&\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\
&\leq k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E}\left\langle \sum_{j=1}^{k} \sigma_j A_{\pi(j)}, \sum_{j=1}^{k} \sigma_j A_{\pi(j)} \right\rangle\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\sum_{j=1}^{k} \mathbb{E}\|A_{\pi(j)}\|_2^2\right)^{1/2}
\end{aligned}$$

The tail

$x = (I - A^T A)y, y = \sum_{i=1}^{k} \sigma_i e_{\pi(i)}$

$$
\begin{aligned}
\mathbb{E}\|x_{tail}(k)\|_1 &\leq \mathbb{E}\|x\|_1 \\
&\leq \mathbb{E}\|y\|_1 + \mathbb{E}\|A^T A y\|_1 \\
&\leq k + \sqrt{n} \cdot \left(\mathbb{E}\|A^T A y\|_2^2\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E} y^T A^T A y\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\mathbb{E}\left\langle \sum_{j=1}^{k} \sigma_j A_{\pi(j)}, \sum_{j=1}^{k} \sigma_j A_{\pi(j)} \right\rangle\right)^{1/2} \\
&= k + \sqrt{n} \cdot \left(\sum_{j=1}^{k} \mathbb{E}\|A_{\pi(j)}\|_2^2\right)^{1/2} \\
&= k + \sqrt{km}
\end{aligned}
$$

Wrapping up

- Calculating expectations and using Markov also shows that with constant probability an $\Omega(1)$ fraction of the coordinates $\pi([k])$ are at least a constant in x, i.e. $\|x_{head(k)}\|_1 = \Omega(k)$ with $2/3$ probability.
- Thus with positive probability we simultaneously have both $\|x_{head(k)}\|_1 = \Omega(k)$ and $\|x_{tail(k)}\|_1 \leq 3(k + \sqrt{km})$. This implies the existence of a bad vector if $m < ck/\varepsilon^2$ for some $c > 0$.

$m = \Omega(k \log(n/k)/\varepsilon)$

- Communication complexity, reduction from the EQUALITY problem on strings of length $r = \Theta((k/\varepsilon)\log n \log(n/k))$.
- In EQUALITY, Alice and Bob receive strings $x, y \in \{0,1\}^r$, respectively, and they must decide whether $x = y$. Known that $\Omega(r)$ communication is required deterministically.

Reduction from EQUALITY details

- Inspired by an approach of [DIPW10].
- Let S be the set of all strings in $\{0, c\varepsilon/k\}^n$ with ℓ_1 norm 1. Note $\log|S| = \Theta((k/\varepsilon)\log(\varepsilon n/k))$.
- Alice, Bob each get strings in $\{0,1\}^r$ with $r = \log n \cdot \log|S|$.
- Alice treats her input as $\log n$ indices into elements of S: $x^1, \ldots, x^{\log n}$. Similarly Bob does this with his input to get $y^1, \ldots, y^{\log n}$.
- Alice computes $u = \sum_{i=1}^{\log n} 2^i x^i$ and Bob computes $v = \sum_{i=1}^{\log n} 2^i y^i$.

Reduction from EQUALITY details

- Inspired by an approach of [DIPW10].
- Let S be the set of all strings in $\{0, c\varepsilon/k\}^n$ with ℓ_1 norm 1. Note $\log|S| = \Theta((k/\varepsilon)\log(\varepsilon n/k))$.
- Alice, Bob each get strings in $\{0,1\}^r$ with $r = \log n \cdot \log|S|$.
- Alice treats her input as $\log n$ indices into elements of S: $x^1, \ldots, x^{\log n}$. Similarly Bob does this with his input to get $y^1, \ldots, y^{\log n}$.
- Alice computes $u = \sum_{i=1}^{\log n} 2^i x^i$ and Bob computes $v = \sum_{i=1}^{\log n} 2^i y^i$.
- Alice sends $A'u$ where A' is A rounded to $O(\log n)$ precision.

Reduction from EQUALITY details

- Inspired by an approach of [DIPW10].
- Let S be the set of all strings in $\{0, c\varepsilon/k\}^n$ with ℓ_1 norm 1. Note $\log|S| = \Theta((k/\varepsilon)\log(\varepsilon n/k))$.
- Alice, Bob each get strings in $\{0,1\}^r$ with $r = \log n \cdot \log|S|$.
- Alice treats her input as $\log n$ indices into elements of S: $x^1, \ldots, x^{\log n}$. Similarly Bob does this with his input to get $y^1, \ldots, y^{\log n}$.
- Alice computes $u = \sum_{i=1}^{\log n} 2^i x^i$ and Bob computes $v = \sum_{i=1}^{\log n} 2^i y^i$.
- Alice sends $A'u$ where A' is A rounded to $O(\log n)$ precision.
- Bob computes $A'(u-v)$ and says "Equal" if the result is 0, else says "Not equal".
- Communication is $\#\mathrm{rows}(A) \cdot O(\log n)$, so lower bound on number of rows is $\Omega(r/\log n)$.

Norm estimation

Connection to Gelfand widths

We show:

Theorem: For $1 \le q < p \le \infty$, let m be the minimum number such that there is an $n - m$ dimensional subspace S of $\mathbb{R}^n$ satisfying $\sup_{v \in S} \frac{\|v\|_p}{\|v\|_q} \le \varepsilon$. Then there is an $m \times n$ matrix A and associated output procedure Out which for any $x \in \mathbb{R}^n$, given Ax, outputs an estimate of $\|v\|_p$ with additive error at most $\varepsilon \|v\|_q$. Moreover, any matrix A with fewer rows will fail to perform the same task.

Connection to Gelfand widths

We show:

Theorem: For $1 \leq q < p \leq \infty$, let m be the minimum number such that there is an $n-m$ dimensional subspace S of $\mathbb{R}^n$ satisfying $\sup_{v \in S} \frac{\|v\|_p}{\|v\|_q} \leq \varepsilon$. Then there is an $m \times n$ matrix A and associated output procedure *Out* which for any $x \in \mathbb{R}^n$, given Ax, outputs an estimate of $\|v\|_p$ with additive error at most $\varepsilon\|v\|_q$. Moreover, any matrix A with fewer rows will fail to perform the same task.

Definition: The *Gelfand width of order m* is the infimum over all $(n-m)$-dimensional subspaces S of $\mathbb{R}^n$ of $\sup_{v \in S} \frac{\|v\|_p}{\|v\|_q}$.

Thus, the above theorem tells us that the optimal number of rows is just the smallest m such that the Gelfand width of order m is at most ε (the subspace is then just the kernel of A). For example, [Foucart et al, 2010] and [Garnaev-Gluskin, 1984] for $p = 2, q = 1$ give that the width of order m is $\Theta(\sqrt{(1 + \log(n/m))/m})$, so $m = \Theta(\varepsilon^{-2} \log(\varepsilon^2 n))$ is optimal for these norms.

Proof of theorem

Theorem: For $1 \leq q < p \leq \infty$, let m be the minimum number such that there is an $n - m$ dimensional subspace S of $\mathbb{R}^n$ satisfying $\sup_{v \in S} \frac{\|v\|_p}{\|v\|_q} \leq \varepsilon$. Then there is an $m \times n$ matrix A and associated output procedure *Out* which for any $x \in \mathbb{R}^n$, given Ax, outputs an estimate of $\|v\|_p$ with additive error at most $\varepsilon \|v\|_q$. Moreover, any matrix A with fewer rows will fail to perform the same task.

Proof: Suppose A has fewer than m rows. Then some $v \in \ker(A)$ has $\|v\|_p > \varepsilon \|v\|_q$, so 0 is not a valid approximation of $\|v\|_p$ but we must output 0 whenever $Av = 0$ to be correct on the 0 vector.

Proof of theorem

Theorem: For $1 \le q < p \le \infty$, let m be the minimum number such that there is an $n - m$ dimensional subspace S of $\mathbb{R}^n$ satisfying $\sup_{v \in S} \frac{\|v\|_p}{\|v\|_q} \le \varepsilon$. Then there is an $m \times n$ matrix A and associated output procedure *Out* which for any $x \in \mathbb{R}^n$, given Ax, outputs an estimate of $\|v\|_p$ with additive error at most $\varepsilon \|v\|_q$. Moreover, any matrix A with fewer rows will fail to perform the same task.

Proof: For the other direction, let A be such that the above $n - m$ dimensional subspace is its kernel. For any sketch z, we must output a number in the range $[\|x\|_p - \varepsilon\|x\|_q, \|x\|_p + \varepsilon\|x\|_q]$ for any x with $Ax = z$. Assume not possible, so $\exists x, y$ with $Ax = Ay$ but $\|x\|_p - \varepsilon\|x\|_q > \|y\|_p + \varepsilon\|y\|_q$. But $x - y \in \ker(A)$, so

$$\|x\|_p - \varepsilon\|y\|_p \le \|x - y\|_p \le \varepsilon\|x - y\|_q \le \varepsilon(\|x\|_q + \|y\|_q)$$

This is a contradiction. Thus in fact, our *Out* procedure can just be to output $\min_{x : Ax = z} \|x\|_p + \varepsilon\|x\|_q$. This can be solved in poly time using the ellipsoid method; details in paper.

Future Directions (open)

- Find correct complexity for point query/inner product/heavy hitters.
- Obtain better time complexity for deterministic point query.
- Nail exactly complexity for ℓ_1/ℓ_1 sparse recovery.
- Norm estimation with a faster *Out* procedure (avoid ellipsoid method!).
- Understand deterministic complexities for other streaming problems.
- (known?) What's a simple "very explicit" construction of an incoherent matrix with $m = O(\varepsilon^{-2} \log n)$ rows?