A Space Optimal Streaming Algorithm for Sketching Small Moments

Daniel M. Kane
Harvard

Jelani Nelson
MIT

David P. Woodruff
IBM Almaden

December 18, 2009

Streaming moments: problem formulation

Model

- $x = (x_1, x_2, \ldots, x_n)$ starts off as $\vec{0}$
- m updates $(i_1, v_1), (i_2, v_2), \ldots, (i_m, v_m)$
- Update (i, v) causes change $x_i \leftarrow x_i + v$
- $v \in \{-M, \ldots, M\}$

Streaming moments: problem formulation

Model

- $x = (x_1, x_2, \ldots, x_n)$ starts off as $\vec{0}$
- m updates $(i_1, v_1), (i_2, v_2), \ldots, (i_m, v_m)$
- Update (i, v) causes change $x_i \leftarrow x_i + v$
- $v \in \{-M, \ldots, M\}$

Goal: Output $F_p \stackrel{\text{def}}{=} \sum_{i=1}^{n} |x_i|^p = \|x\|_p^p$

Streaming moments: objectives

Objectives

- Minimize space usage
- Minimize update time

Trivial solutions

- Keep x in memory: $O(n \log(mM))$ space / $O(1)$ time
- Keep stream in memory: $O(m \log(nM))$ space / $O(1)$ time

Goal: Get polylogarithmic dependence on n, m

Streaming moments: bad news

Alon, Matias, Szegedy '99: No sublinear space algorithms without

- Approximation (allow output to be $(1 \pm \varepsilon)F_p$)
- Randomization (allow 1% failure probability)

New goal: Output $(1 \pm \varepsilon)F_p$ with probability 99%

Streaming moments: bad news

Alon, Matias, Szegedy '99: No sublinear space algorithms without

- Approximation (allow output to be $(1 \pm \varepsilon)F_p$)
- Randomization (allow 1% failure probability)

New goal: Output $(1 \pm \varepsilon)F_p$ with probability 99%

More bad news: Polynomial space required for $p > 2$
([BJKS '02] and [CKS '03])

Streaming moments: bad news

Alon, Matias, Szegedy '99: No sublinear space algorithms without

- Approximation (allow output to be $(1 \pm \varepsilon)F_p$)
- Randomization (allow 1% failure probability)

New goal: Output $(1 \pm \varepsilon)F_p$ with probability 99%

More bad news: Polynomial space required for $p > 2$
([BJKS '02] and [CKS '03])

Newer goal: Output $(1 \pm \varepsilon)F_p$ with probability 99% for $0 \leq p \leq 2$

Contributions
$(0 < p \leq 2)$

(Notation: $N = \min\{n, m\}$)

Ref	Upper bound	Lower bound	Update time
AMS'99	$O(\varepsilon^{-2}\log(mM))$ (p=2)	$\Omega(\log N)$	$O(1)$ (*)
FKSV'99 (**)	$O(\varepsilon^{-2}\log(mM))$ (p=1)	————	$O\left(\frac{\log(NM)}{\varepsilon^2}\right)$
Indyk'06, Li'08	$O(\varepsilon^{-2}\log(mM)\log N)$	————	$O(\varepsilon^{-2})$
GC'07	$O(\varepsilon^{-(2+p)}\log^2(N)\log(mM))$	————	$\mathrm{polylog}(mM)$
Woodruff'04	————	$\Omega(\varepsilon^{-2})$	————
This work	$O(\varepsilon^{-2}\log(mM))$	$\Omega(\varepsilon^{-2}\log(mM))$	$\tilde{O}(\varepsilon^{-2})$

(*) achieved by CCF'02, TZ'04, (**) L_1-difference only

F_p $(0 < p < 2)$
p-stable distributions

Definition (Zolotarev '86)

For $0 < p \leq 2$, there exists a probability distribution $\mathcal{D}_p$ called the *p-stable distribution* such that if $Q_1, \ldots, Q_n \sim \mathcal{D}_p$ are independent, then $\sum_{i=1}^{n} Q_i x_i \sim \|x\|_p \mathcal{D}_p$.

(In short: $\mathcal{D}_p$ carries information about L_p norms)

F_p $(0 < p < 2)$
p-stable distributions

Definition (Zolotarev '86)

For $0 < p \leq 2$, there exists a probability distribution $\mathcal{D}_p$ called the *p-stable distribution* such that if $Q_1, \ldots, Q_n \sim \mathcal{D}_p$ are independent, then $\sum_{i=1}^{n} Q_i x_i \sim \|x\|_p \mathcal{D}_p$.

(In short: $\mathcal{D}_p$ carries information about L_p norms)

- $p = 2$: Gaussian
- $p = 1$: Cauchy
- $p = 1/2$: Lévy

Algorithms based on p-stable sketch matrices

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{r,1} & \cdots & A_{r,n} \end{bmatrix},$$ the $A_{i,j}$ are i.i.d. from $\mathcal{D}_p$,

Maintain $Ax = y$

Algorithms based on p-stable sketch matrices

$$A = \begin{bmatrix} A_{1,1} & \cdots & A_{1,n} \\ \vdots & \ddots & \vdots \\ A_{r,1} & \cdots & A_{r,n} \end{bmatrix}, \text{ the } A_{i,j} \text{ are i.i.d. from } \mathcal{D}_p,$$

Maintain $Ax = y$

- Idea introduced by Indyk '06
- Indyk '06: Estimate F_p as $\mathrm{median}\{|y_j|^p\}_{j=1}^r$
- Li '08: Estimate F_p as $\frac{\prod_{j=1}^r |y_j|^{p/r}}{\left[\frac{2}{\pi}\Gamma\left(\frac{p}{r}\right)\Gamma\left(1-\frac{1}{r}\right)\sin\left(\frac{\pi}{2}\cdot\frac{p}{r}\right)\right]^r}$
- Both cases: $r = \Theta(1/\varepsilon^2)$

Too much randomness

- In Indyk'06 and Li'08, $\Omega(n/\varepsilon^2)$ bits needed to store matrix A

Too much randomness

- In Indyk'06 and Li'08, $\Omega(n/\varepsilon^2)$ bits needed to store matrix A
- Indyk derandomized using Nisan's pseudorandom generator (but blowed up space)

Too much randomness

- In Indyk'06 and Li'08, $\Omega(n/\varepsilon^2)$ bits needed to store matrix A
- Indyk derandomized using Nisan's pseudorandom generator (but blowed up space)

Is there a more efficient derandomization?

Our Contributions

Yes, via k-wise independence!

- For fixed i, make the $A_{i,j}$ k-wise independent
- Make the seeds used to generate rows of A pairwise independent

Our Contributions

Yes, via k-wise independence!

- For fixed i, make the $A_{i,j}$ k-wise independent
- Make the seeds used to generate rows of A pairwise independent

- $k = \tilde{\Theta}(1/\varepsilon^p)$ fools Indyk's estimator
- A different estimator works with $k = \Theta(\log(1/\varepsilon)/\log\log(1/\varepsilon))$.

Our Contributions

A different estimator
(works with $k = O(\log(1/\varepsilon)/\log\log(1/\varepsilon))$)

1. Maintain $Ax = y$ and $A'x = y'$.
2. A has $k = \Theta(\log(1/\varepsilon)/\log\log(1/\varepsilon))$, $r = \Theta(1/\varepsilon^2)$.
3. A' has $k', r' = \Theta(1)$.
4. $y'_{\mathrm{med}} \leftarrow \mathrm{median}\{|y'_j|\}_{j=1}^{r'}$.
5. Output $-y'^p_{\mathrm{med}} \cdot \ln\left(\frac{1}{r}\sum_{j=1}^{r}\cos\left(\frac{y_j}{y'_{\mathrm{med}}}\right)\right)$.

Analyzing median F_p algorithm (full independence)

An argument for the median:

Define

$$I_{[a,b]}(x) = \begin{cases} 1, & \text{if } x \in [a, b], \\ 0, & \text{otherwise} \end{cases}$$

- $Q = \sum_i Q_i x_i$.

Analyzing median F_p algorithm (full independence)

An argument for the median:

Define

$$I_{[a,b]}(x) = \begin{cases} 1, & \text{if } x \in [a,b], \\ 0, & \text{otherwise} \end{cases}$$

- $Q = \sum_i Q_i x_i$.
- "$\mathrm{median}(|Q|/\|x\|_p) = 1$" means $\mathbf{E}[I_{[-1,1]}(Q/\|x\|_p)] = 1/2$.

Analyzing median F_p algorithm (full independence)

An argument for the median:

Define

$$I_{[a,b]}(x) = \begin{cases} 1, & \text{if } x \in [a,b], \\ 0, & \text{otherwise} \end{cases}$$

- $Q = \sum_i Q_i x_i$.
- "$\mathrm{median}(|Q|/\|x\|_p) = 1$" means $\mathbf{E}[I_{[-1,1]}(Q/\|x\|_p)] = 1/2$.

- $\mathbf{E}[I_{[-1+\varepsilon,1-\varepsilon]}(Q/\|x\|_p)] = 1/2 - \Theta(\varepsilon)$
- $\mathbf{E}[I_{[-1-\varepsilon,1+\varepsilon]}(Q/\|x\|_p)] = 1/2 + \Theta(\varepsilon)$
- Take $r = \Theta(1/\varepsilon^2)$ trials $Q_1, \ldots, Q_r$. Number of counters inside interval is concentrated by Chebyshev.

$\Rightarrow$ median of the $|Q_j|$ is $(1 \pm \Theta(\varepsilon))\|x\|_p$ with probability 2/3

Analyzing median F_p algorithm (k-wise independence)

One possible path

- Replace $I_{[a,b]}$ with a well-approximating low-degree polynomial.
- k-wise independence fools polynomials.

Analyzing median F_p algorithm (k-wise independence)

One possible path

- Replace $I_{[a,b]}$ with a well-approximating low-degree polynomial.
- k-wise independence fools polynomials.

What we actually do (for good reason)

- Replace $I_{[a,b]}$ with a well-approximating smooth function $\tilde{I}_{[a,b]}$.
- Show $\tilde{I}_{[a,b]}$ is fooled by k-wise independence via Taylor's theorem.

Defining $\tilde{I}_{[a,b]}$ FT-mollification

Define

$$b(x) = \begin{cases} e^{-\frac{x^2}{1-x^2}} & \text{for } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{I}^c_{[a,b]}(x) = \frac{1}{2\pi}(c \cdot \hat{b}(ct) * I_{[a,b]}(t))(x)$$

Defining $\tilde{I}_{[a,b]}$ FT-mollification

Define

$$b(x) = \begin{cases} e^{-\frac{x^2}{1-x^2}} & \text{for } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{I}^c_{[a,b]}(x) = \frac{1}{2\pi}(c \cdot \hat{b}(ct) * I_{[a,b]}(t))(x)$$

Then, for $c > 1$,

i. $\|(\tilde{I}^c_{[a,b]})^{(\ell)}\|_\infty = O(c^\ell)$ for $\ell \geq 0$.

ii. For $c = \tilde{O}(1/\varepsilon)$, $|\tilde{I}^c_{[a,b]} - I_{[a,b]}| < \varepsilon$ except potentially at $a \pm \varepsilon$ and $b \pm \varepsilon$.

Defining $\tilde{I}_{[a,b]}$
FT-mollification

Define

$$b(x) = \begin{cases} e^{-\frac{x^2}{1-x^2}} & \text{for } |x| < 1 \\ 0 & \text{otherwise} \end{cases}$$

and

$$\tilde{I}^c_{[a,b]}(x) = \frac{1}{2\pi}(c \cdot \hat{b}(ct) * I_{[a,b]}(t))(x)$$

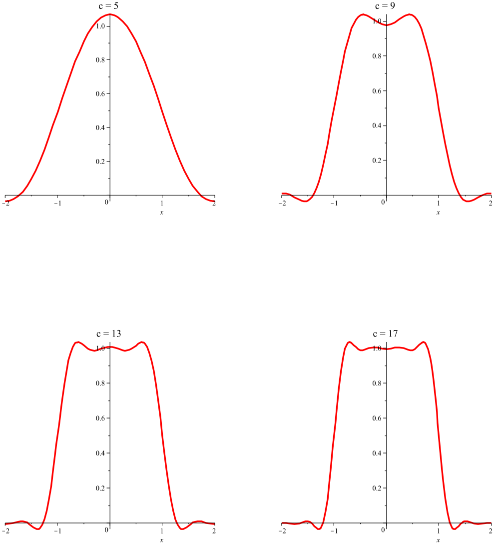
Then, for $c > 1$,

i. $\|(\tilde{I}^c_{[a,b]})^{(\ell)}\|_\infty = O(c^\ell)$ for $\ell \geq 0$.

ii. For $c = \tilde{O}(1/\varepsilon)$, $|\tilde{I}^c_{[a,b]} - I_{[a,b]}| < \varepsilon$ except potentially at $a \pm \varepsilon$ and $b \pm \varepsilon$.

For c large, $\tilde{I}^c_{[a,b]}$ looks like $I_{[a,b]}$.

$\tilde{I}^{c}_{[-1,1]}$ plots



Proof Outline

- Let R_i be k-wise independent from $\mathcal{D}_p$, and Q_i be i.i.d.
- Let $R = \sum_i R_i x_i$ and $Q = \sum_i Q_i x_i$.
- Suppose $\|x\|_p = 1$.

Proof Outline

- Let R_i be k-wise independent from $\mathcal{D}_p$, and Q_i be i.i.d.
- Let $R = \sum_i R_i x_i$ and $Q = \sum_i Q_i x_i$.
- Suppose $\|x\|_p = 1$.

Want: $\mathbf{E}[I_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[I_{[a,b]}(R)]$

Proof Outline

- Let R_i be k-wise independent from $\mathcal{D}_p$, and Q_i be i.i.d.
- Let $R = \sum_i R_i x_i$ and $Q = \sum_i Q_i x_i$.
- Suppose $\|x\|_p = 1$.

Want: $\mathbf{E}[I_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[I_{[a,b]}(R)]$

Proof: $\mathbf{E}[I_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[a,b]}(R)] \approx_\varepsilon \mathbf{E}[I_{[a,b]}(R)]$

(1)→(2) $\tilde{I}^c$ well-approximates I except for two length-$O(\varepsilon)$ strips. Use anticoncentration.

(2)→(3) Main technical lemma.

(3)→(4) Same as (1)→(2), but must prove anticoncentration.

Main technical lemma

Lemma

- $\|f^{(\ell)}(x)\|_\infty = O(\alpha^\ell)$ *for all* $\ell \geq 0$
- $k = \max\{\log(1/\varepsilon), \alpha^p\}$
- R_i *are* $\Theta(k)$*-wise indep.,* Q_i *are fully indep., from* $\mathcal{D}_p$
- $R = \sum_i R_i x_i, Q = \sum_i Q_i x_i$
- $\|x\|_p = O(1)$

$$\Rightarrow |\mathbf{E}[f(R)] - \mathbf{E}[f(Q)]| < \varepsilon$$

Proof strategy

- Approximate f by a polynomial (Taylor-expand), and bound expected difference using Taylor's theorem, by analyzing moments $\mathbf{E}[X_i^k]$ and high-order derivatives of f

Proof strategy

- Approximate f by a polynomial (Taylor-expand), and bound expected difference using Taylor's theorem, by analyzing moments $\mathbf{E}[X_i^k]$ and high-order derivatives of f
- Problem: $\mathcal{D}_p$ has infinite moments for $p < 2$

Proof strategy (modified)

Linearity of expectation:

$$\mathbf{E}[f(R)] = \mathbf{E}\left[\sum_{A \in \mathcal{A}} \mathbf{1}_A \cdot f(R)\right] = \sum_{A \in \mathcal{A}} \mathbf{E}[\mathbf{1}_A \cdot f(R)]$$

where events in $\mathcal{A}$ partition probability space

Proof strategy (modified)

Linearity of expectation:

$$\mathbf{E}[f(R)] = \mathbf{E}\left[\sum_{A \in \mathcal{A}} \mathbf{1}_A \cdot f(R)\right] = \sum_{A \in \mathcal{A}} \mathbf{E}[\mathbf{1}_A \cdot f(R)]$$

where events in $\mathcal{A}$ partition probability space

What events should we consider?

Proof strategy (modified)

Linearity of expectation:

$$\mathbf{E}[f(R)] = \mathbf{E}\left[\sum_{A\in\mathcal{A}} \mathbf{1}_A \cdot f(R)\right] = \sum_{A\in\mathcal{A}} \mathbf{E}[\mathbf{1}_A \cdot f(R)]$$

where events in $\mathcal{A}$ partition probability space

What events should we consider? Truncation

Proof strategy (modified)

Linearity of expectation:

$$\mathbf{E}[f(R)] = \mathbf{E}\left[\sum_{A \in \mathcal{A}} \mathbf{1}_A \cdot f(R)\right] = \sum_{A \in \mathcal{A}} \mathbf{E}[\mathbf{1}_A \cdot f(R)]$$

where events in $\mathcal{A}$ partition probability space

What events should we consider? Truncation

Define random variables:

$$R'_i = \begin{cases} R_i, & |R_i x_i| \leq \lambda \\ 0, & \text{otherwise} \end{cases}$$

Proof strategy (modified)

$$R_i' = \begin{cases} R_i & |R_i x_i| \leq \lambda \\ 0 & \text{otherwise} \end{cases}$$

For $S \subseteq [n]$, event $\mathbf{1}_S$ indicates that S is *exactly* the set of truncated R_i'

Proof strategy (modified)

$$R_i' = \begin{cases} R_i & |R_i x_i| \leq \lambda \\ 0 & \text{otherwise} \end{cases}$$

For $S \subseteq [n]$, event $\mathbf{1}_S$ indicates that S is *exactly* the set of truncated R_i'

$$\begin{aligned} \mathbf{E}[f(R)] &= \sum_{S \subseteq [n]} \mathbf{E}\left[\mathbf{1}_S \cdot f(R)\right] \\ &= \sum_{S \subseteq [n]} \mathbf{E}\left[\mathbf{1}_S \cdot f\left(\sum_{i \in S} R_i x_i + \sum_{i \notin S} R_i' x_i\right)\right] \end{aligned}$$

Proof strategy (modified)

$$R'_i = \begin{cases} R_i & |R_i x_i| \le \lambda \\ 0 & \text{otherwise} \end{cases}$$

For $S \subseteq [n]$, event $\mathbf{1}_S$ indicates that S is *exactly* the set of truncated R'_i

$$\begin{aligned} \mathbf{E}[f(R)] &= \sum_{S \subseteq [n]} \mathbf{E}\left[\mathbf{1}_S \cdot f(R)\right] \\ &= \sum_{S \subseteq [n]} \mathbf{E}\left[\mathbf{1}_S \cdot f\left(\sum_{i \in S} R_i x_i + \sum_{i \notin S} R'_i x_i\right)\right] \end{aligned}$$

Problem: How to reason about $\mathbf{1}_S$ using k-wise indep.?

Dealing with $\mathbf{1}_S$

For $S \subseteq [n]$, $\mathbf{1}'_S$ indicates that S is a *subset* of the truncated R'_i

Dealing with $\mathbf{1}_S$

For $S \subseteq [n]$, $\mathbf{1}'_S$ indicates that S is a *subset* of the truncated R'_i
Use inclusion-exclusion!

Dealing with $\mathbf{1}_S$

For $S \subseteq [n]$, $\mathbf{1}'_S$ indicates that S is a *subset* of the truncated R'_i
Use inclusion-exclusion!

$$\begin{aligned}\mathbf{1}_S &= \mathbf{1}'_S \cdot \left(\prod_{i \notin S} \left(1 - \mathbf{1}'_{\{i\}} \right) \right) \\ &= \sum_{T \subseteq [n] \setminus S} (-1)^{|T|} \mathbf{1}'_{S \cup T}\end{aligned}$$

Dealing with $\mathbf{1}_S$

For $S \subseteq [n]$, $\mathbf{1}'_S$ indicates that S is a *subset* of the truncated R'_i
Use inclusion-exclusion!

$$\begin{aligned}\mathbf{1}_S &= \mathbf{1}'_S \cdot \left(\prod_{i \notin S} \left(1 - \mathbf{1}'_{\{i\}}\right) \right) \\ &= \sum_{T \subseteq [n] \setminus S} (-1)^{|T|} \mathbf{1}'_{S \cup T}\end{aligned}$$

Now

$$\mathbf{E}[f(R)] = \sum_{S \subseteq [n]} \sum_{T \subseteq [n] \setminus S} (-1)^{|T|} \mathbf{E}\left[\mathbf{1}'_{S \cup T} \cdot f\left(\sum_{i \in S} R_i x_i + \sum_{i \notin S} R'_i x_i \right) \right]$$

Dealing with $\mathbf{1}_S$

For $S \subseteq [n]$, $\mathbf{1}'_S$ indicates that S is a *subset* of the truncated R'_i
Use inclusion-exclusion!

$$\begin{aligned}\mathbf{1}_S &= \mathbf{1}'_S \cdot \left(\prod_{i \notin S} \left(1 - \mathbf{1}'_{\{i\}}\right)\right) \\ &= \sum_{T \subseteq [n] \setminus S} (-1)^{|T|} \mathbf{1}'_{S \cup T}\end{aligned}$$

Now

$$\mathbf{E}[f(R)] = \sum_{S \subseteq [n]} \sum_{T \subseteq [n] \setminus S} (-1)^{|T|} \mathbf{E}\left[\mathbf{1}'_{S \cup T} \cdot f\left(\sum_{i \in S} R_i x_i + \sum_{i \notin S} R'_i x_i\right)\right]$$

Still a problem: How to deal with large S, T?

Approximate Inclusion-Exclusion

Introduced to streaming by Bar-Yossef et al. '02
(analyzed balls and bins with limited independence)

Approximate Inclusion-Exclusion

Introduced to streaming by Bar-Yossef et al. '02
(analyzed balls and bins with limited independence)

$$\mathbf{E}[f(R)] = \sum_{S\subseteq[n]} \sum_{T\subseteq[n]\setminus S} (-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S} R_i x_i + \sum_{i\notin S} R'_i x_i\right)\right]$$

Approximate Inclusion-Exclusion

Introduced to streaming by Bar-Yossef et al. '02
(analyzed balls and bins with limited independence)

$$\begin{aligned}
\mathbf{E}[f(R)] &= \sum_{S\subseteq[n]}\ \sum_{T\subseteq[n]\setminus S}(-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S}R_ix_i+\sum_{i\notin S}R'_ix_i\right)\right]\\
&\overset{?}{\approx} \sum_{\substack{S\subseteq[n]\\|S|\le Ck}}\ \sum_{\substack{T\subseteq[n]\setminus S\\|T|\le Ck}}(-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S}R_ix_i+\sum_{i\notin S}R'_ix_i\right)\right]
\end{aligned}$$

Approximate Inclusion-Exclusion

Introduced to streaming by Bar-Yossef et al. '02
(analyzed balls and bins with limited independence)

$$\begin{aligned}\mathbf{E}[f(R)] &= \sum_{S\subseteq[n]}\sum_{T\subseteq[n]\setminus S}(-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S}R_ix_i+\sum_{i\notin S}R'_ix_i\right)\right]\\ &\approx \sum_{\substack{S\subseteq[n]\\ |S|\leq Ck}}\sum_{\substack{T\subseteq[n]\setminus S\\ |T|\leq Ck}}(-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S}R_ix_i+\sum_{i\notin S}R'_ix_i\right)\right]\end{aligned}$$

Approximate Inclusion-Exclusion

Introduced to streaming by Bar-Yossef et al. '02
(analyzed balls and bins with limited independence)

$$\mathbf{E}[f(R)] = \sum_{S\subseteq[n]} \sum_{T\subseteq[n]\setminus S} (-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S} R_i x_i + \sum_{i\notin S} R'_i x_i\right)\right]$$

$$\approx \sum_{\substack{S\subseteq[n]\\ |S|\le Ck}} \sum_{\substack{T\subseteq[n]\setminus S\\ |T|\le Ck}} (-1)^{|T|}\mathbf{E}\left[\mathbf{1}'_{S\cup T}\cdot f\left(\sum_{i\in S} R_i x_i + \sum_{i\notin S} R'_i x_i\right)\right]$$

$$\mathbf{E}[f(R)] \approx_\varepsilon \mathbf{E}[F(\vec{R})] \approx_\varepsilon \sum_{\substack{S,T\subseteq[n]\\ |S|,|T|\le Ck\\ S\cap T=\emptyset}} (-1)^{|T|}\mathbf{E}_{\substack{R_i\\ i\in S\cup T}}\left[\mathbf{1}'_{S\cup T}\cdot\mathbf{E}\left[p_{k,R_i}\left(\sum_{i\notin S\cup T} R'_i x_i\right)\right]\right]$$

$$= \sum_{\substack{S,T\subseteq[n]\\ |S|,|T|\le Ck\\ S\cap T=\emptyset}} (-1)^{|T|}\mathbf{E}_{\substack{Q_i\\ i\in S\cup T}}\left[\mathbf{1}'_{S\cup T}\cdot\mathbf{E}\left[p_{k,Q_i}\left(\sum_{i\notin S\cup T} Q'_i x_i\right)\right]\right] \approx_\varepsilon \mathbf{E}[F(\vec{Q})] \approx_\varepsilon \mathbf{E}[f(Q)]$$

Proof Outline

- Let R_i be k-wise independent from $\mathcal{D}_p$, and Q_i be i.i.d.
- Let $R = \sum_i R_i x_i$ and $Q = \sum_i Q_i x_i$.
- Suppose $\|x\|_p = 1$.

Want: $\mathbf{E}[I_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[I_{[a,b]}(R)]$

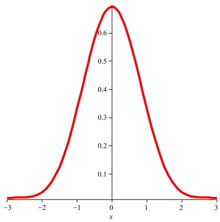
Proof: $\mathbf{E}[I_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[a,b]}(Q)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[a,b]}(R)] \approx_\varepsilon \mathbf{E}[I_{[a,b]}(R)]$

(1)→(2) $\tilde{I}^c$ well-approximates I except for two length-$O(\varepsilon)$ strips. Use anticoncentration.

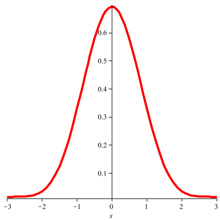
(2)→(3) Main technical lemma.

(3)→(4) Same as (1)→(2), but must prove anticoncentration.

Anticoncentration of R



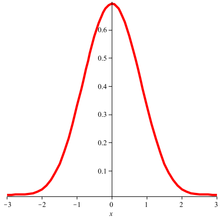
Anticoncentration of R



- $\|f^{(\ell)}\|_\infty = O(1/\varepsilon)^\ell$
  $\Rightarrow$ fooled with $k = O(1/\varepsilon^p)$
- Easy to show $\mathbf{E}[f(Q)] = O(\varepsilon)$
- $\Rightarrow \mathbf{E}[f(R)] = O(\varepsilon)$ by main technical lemma
- $\Rightarrow$ anticoncentration in interval $[-\varepsilon, \varepsilon]$

Shift f to show anticoncentration in any width-$O(\varepsilon)$ interval.

Anticoncentration of R



$$f(x) = -\int_{-\infty}^{x/\varepsilon} \frac{\sin^4(y)}{y^3} dy$$

- $\|f^{(\ell)}\|_\infty = O(1/\varepsilon)^\ell$
  $\Rightarrow$ fooled with $k = O(1/\varepsilon^p)$
- Easy to show $\mathbf{E}[f(Q)] = O(\varepsilon)$
- $\Rightarrow$ $\mathbf{E}[f(R)] = O(\varepsilon)$ by main technical lemma
- $\Rightarrow$ anticoncentration in interval $[-\varepsilon, \varepsilon]$

Shift f to show anticoncentration in any width-$O(\varepsilon)$ interval.

Intuition for the new estimator

Our new estimator's final step:

"Let $y'_{\mathrm{median}} = \mathrm{median}\{|y'_j|\}_{j=1}^{r'}$.
Output $-y'^{p}_{\mathrm{median}} \cdot \ln\left(\frac{1}{r}\sum_{j=1}^{r} \cos\left(\frac{y_j}{y'_{\mathrm{median}}}\right)\right)$."

Intuition for the new estimator

Our new estimator's final step:

"Let $y'_{\mathrm{median}} = \mathrm{median}\{|y'_j|\}_{j=1}^{r'}$.
Output $-y'^{p}_{\mathrm{median}} \cdot \ln\left(\frac{1}{r}\sum_{j=1}^{r}\cos\left(\frac{y_j}{y'_{\mathrm{median}}}\right)\right)$."

- We know $y'_{\mathrm{median}} = \Theta(\|x\|_p)$.
- Apply main technical lemma with $f(x) = \cos(x)$ to refine y'_{median} to a $(1 \pm \varepsilon)$-approximation.

Correcting to $(1 \pm \varepsilon)$-approximation

$$Z \sim \mathcal{D}_p$$

$$\mathbf{E}[\cos(BZ)] = \mathbf{E}\left[\frac{e^{iBZ} + e^{-iBZ}}{2}\right]$$

Can look at Fourier transform of pdf of $\mathcal{D}_p$ to show $\mathbf{E}[\cos(BZ)] = e^{-|B|^p}$

Correcting to $(1 \pm \varepsilon)$-approximation

$$Z \sim \mathcal{D}_p$$

$$\mathbf{E}[\cos(BZ)] = \mathbf{E}\left[\frac{e^{iBZ} + e^{-iBZ}}{2}\right]$$

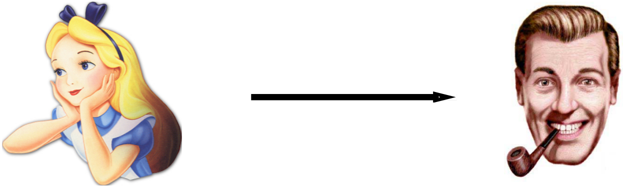
Can look at Fourier transform of pdf of $\mathcal{D}_p$ to show $\mathbf{E}[\cos(BZ)] = e^{-|B|^p}$

- Apply technical lemma to $f\left(\frac{y_j}{y'_{\text{median}}}\right)$ with $f(x) = \cos(x)$
- Use Chebyshev's inequality

Lower bounds

Streaming lower bounds via communication complexity

Alice   Bob



$x \in \mathcal{X}$ $y \in \mathcal{Y}$

- Alice, Bob know $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$
- Bob needs to compute $f(x, y)$
- Communication lower bounds $\Rightarrow$ streaming space lower bounds (Alon, Matias, Szegedy '99)

Previous F_p lower bound

Woodruff '04 and Jayram, Kumar, Sivakumar '08

Indexing

- $\mathcal{X} = \{0,1\}^t$, $\mathcal{Y} = \{1,\ldots,t\}$
- $f(x,y) = x_y$

Previous F_p lower bound

Woodruff '04 and Jayram, Kumar, Sivakumar '08

Indexing

- $\mathcal{X} = \{0,1\}^t$, $\mathcal{Y} = \{1,\ldots,t\}$
- $f(x,y) = x_y$

Gap-Hamming

- $\mathcal{X} = \{0,1\}^{t'}$, $\mathcal{Y} = \{0,1\}^{t'}$
- $$f(x,y) = \begin{cases} 1 & \Delta(x,y) \geq \frac{t'}{2} + \sqrt{t'} \\ 0 & \Delta(x,y) \leq \frac{t'}{2} - \sqrt{t'} \end{cases}$$

Previous F_p lower bound

Woodruff '04 and Jayram, Kumar, Sivakumar '08

Indexing

- $\mathcal{X} = \{0,1\}^t$, $\mathcal{Y} = \{1,\ldots,t\}$
- $f(x,y) = x_y$

Gap-Hamming

- $\mathcal{X} = \{0,1\}^{t'}$, $\mathcal{Y} = \{0,1\}^{t'}$
- $$f(x,y) = \begin{cases} 1 & \Delta(x,y) \geq \frac{t'}{2} + \sqrt{t'} \\ 0 & \Delta(x,y) \leq \frac{t'}{2} - \sqrt{t'} \end{cases}$$

Indexing $\xrightarrow{\text{JKS}'08}$ Gap-Hamming $\xrightarrow{\text{Woodruff}'04}$ F_p

Led to $\Omega(\min\{N, \varepsilon^{-2}\})$ lower bound for F_p

The new F_p lower bound

Augmented-Indexing

- $\mathcal{X} = \{0,1\}^t$, $\mathcal{Y} = \{1, \ldots, t\}$
- Bob also gets x_i for $i > y$
- $f(x, y) = x_y$

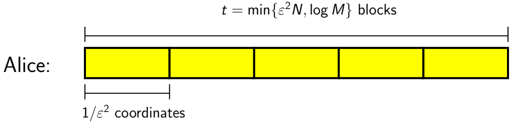
Requires $\Omega(t)$ communication (MNSW '98)

An F_1 lower bound

Theorem

$(1 \pm \varepsilon)$-approximation of F_1 requires $\Omega(\min\{N, \varepsilon^{-2} \log M\})$ space

Proof.



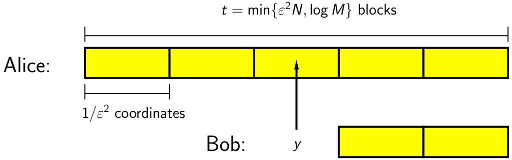
□

An F_1 lower bound

Theorem

$(1 \pm \varepsilon)$-approximation of F_1 requires $\Omega(\min\{N, \varepsilon^{-2} \log M\})$ space

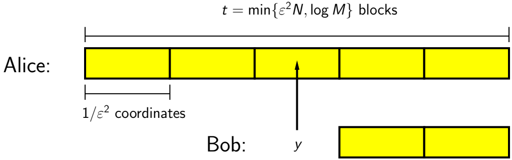
Proof.



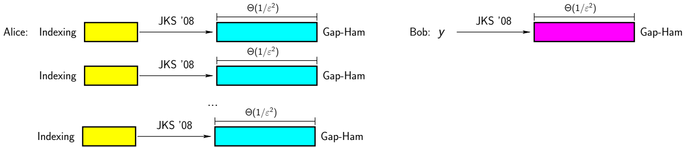
□

An F_1 lower bound

Step 1:

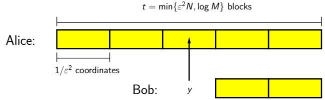


Step 2:

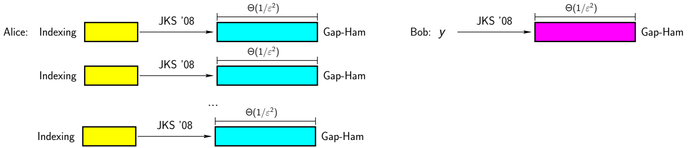


An F_1 lower bound

Step 1:



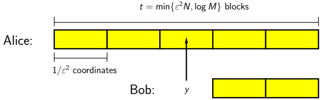
Step 2:



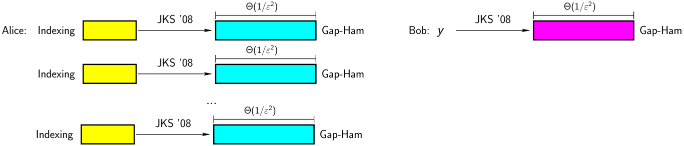
Step 3: For ith Gap-Ham vector z_i, if $z_{i,j} = 1$ Alice puts $((i, j), 2^i)$ in stream

An F_1 lower bound

Step 1:



Step 2:

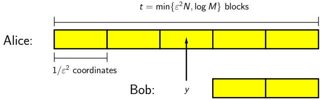


Step 3: For ith Gap-Ham vector z_i, if $z_{i,j} = 1$ Alice puts $((i,j), 2^i)$ in stream

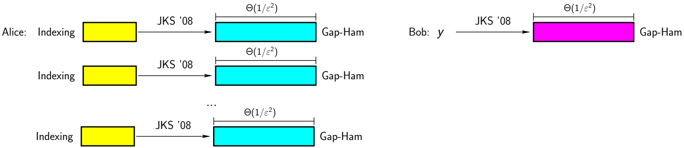
Step 4: Alice sends algorithm state + weight of each block

An F_1 lower bound

Step 1:



Step 2:



Step 3: For ith Gap-Ham vector z_i, if $z_{i,j} = 1$ Alice puts $((i, j), 2^i)$ in stream

Step 4: Alice sends algorithm state + weight of each block

Step 5: Bob deletes contribution of blocks larger than his own

Open problems

- F_p in optimal space with $O(1)$ update time?
- Find other applications for FT-mollification.

Open problems (some progress)

- F_p in optimal space with $O(1)$ update time?

  [N., Woodruff] $p = 1$ with $\varepsilon^{-2} \log^{O(1)}(nmM)$ space, $\log^{O(1)}(nmM)$ update time

- Find other applications for FT-mollification.

  [Kane, N., Woodruff] FT-mollification actually gives an alternative proof that bounded independence fools regular halfspaces ([DGJ$^+$09]).

  [Diakonikolas, Kane, N.] Showed bounded independence fools degree-2 threshold functions, via FT-mollification.

Other news announcements

[Kane, N., Woodruff]: Optimal distinct elements algorithm.

- $O(\varepsilon^{-2} + \log(n))$ bits of space
- $O(1)$ worst-case update and reporting times

Fooling regular halfspaces

- $H_{a,\theta} = \{x : \langle a, x\rangle \geq \theta\}$ (a halfspace).
- Theorem [DGJ+09]: $\Pr[x \in H_{a,\theta}] \approx_\varepsilon \Pr[y \in H_{a,\theta}]$ for $k = \tilde{O}(1/\varepsilon^2)$. x_i are i.i.d., y_i are k-wise independent.
- The [DGJ+09] proof outline:
  1. Reduce to case when $|a_i| \leq \varepsilon$ for all i
  2. Show the theorem in the case when every $|a_i| \leq \varepsilon$ (the "regular" case)

Fooling regular halfspaces

- $H_{a,\theta} = \{x : \langle a, x\rangle \geq \theta\}$ (a halfspace).
- Theorem [DGJ+09]: $\Pr[x \in H_{a,\theta}] \approx_\varepsilon \Pr[y \in H_{a,\theta}]$ for $k = \tilde{O}(1/\varepsilon^2)$. x_i are i.i.d., y_i are k-wise independent.
- The [DGJ+09] proof outline:
  1. Reduce to case when $|a_i| \leq \varepsilon$ for all i
  2. Show the theorem in the case when every $|a_i| \leq \varepsilon$ (the "regular" case)
- Proof of 2 via FT-mollification: $\mathbf{E}[I_{[\theta,\infty)}(\langle a, x\rangle)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(\langle a, x\rangle)] \approx_\varepsilon \mathbf{E}[\tilde{I}^c_{[\theta,\infty)}(\langle a, y\rangle)] \approx_\varepsilon \mathbf{E}[I_{[\theta,\infty)}(\langle a, y\rangle)]$.

Fooling degree-2 threshold functions

Statement: $\mathbf{E}[\mathrm{sign}(p(x))] \approx_\varepsilon \mathbf{E}[\mathrm{sign}(p(y))]$ for $k = \mathrm{poly}(1/\varepsilon)$, p a degree-2 polynomial.

Fooling degree-2 threshold functions

Statement: $\mathbf{E}[\mathrm{sign}(p(x))] \approx_\varepsilon \mathbf{E}[\mathrm{sign}(p(y))]$ for $k = \mathrm{poly}(1/\varepsilon)$, p a degree-2 polynomial.

- Some savings in the known applications: (1) $\Omega(1/\varepsilon^p)$-wise independence fools Indyk's estimator, (2) $\Omega(1/\varepsilon^2)$-wise independence ε-fools regular halfspaces (no more logs).
- A new statement: Bounded independence fools Goemans-Williamson hyperplane rounding.

Fooling degree-2 threshold functions

Statement: $\mathbf{E}[\mathrm{sign}(p(x))] \approx_\varepsilon \mathbf{E}[\mathrm{sign}(p(y))]$ for $k = \mathrm{poly}(1/\varepsilon)$, p a degree-2 polynomial.

- Some savings in the known applications: (1) $\Omega(1/\varepsilon^p)$-wise independence fools Indyk's estimator, (2) $\Omega(1/\varepsilon^2)$-wise independence ε-fools regular halfspaces (no more logs).
- A new statement: Bounded independence fools Goemans-Williamson hyperplane rounding.
- Idea of proof:
  1. $p = p_1 - p_2 + p_3 + p_4 + C$, p_1, p_2 pos. semidef. with no small non-zero eigenvalues, p_3 indefinite with only small eigenvalues, p_4 a linear form, C a constant.
  2. Let Δ be the trace of the symmetric matrix associated with p_3.
  3. Define $R \subseteq \mathbb{R}^4$ by $R = \{z : z_1^2 - z_2^2 + z_3 + z_4 + \Delta + C > 0\}$.
  4. $\mathbf{E}[I_R(M(x))] \approx_\varepsilon \mathbf{E}[\tilde{I}_R^c(M(x))] \approx_\varepsilon \mathbf{E}[\tilde{I}_R^c(M(y))] \approx_\varepsilon \mathbf{E}[I_R(M(y))]$ for $M(z) = (\sqrt{p_1(z)}, \sqrt{p_2(z)}, p_3(z) - \Delta, p_4(z))$.