CSE 522: Sublinear (and Streaming) Algorithms **Spring 2014**

Lecture 9: Approximating F_p (or $||f||_p$) for $p > 2$

April 28, 2014

Lecturer: Paul Beame *Scribe: Paul Beame*

We continue the analysis of Andoni's algorithm, which uses the exponential distribution $Exp(1)$ given by $\mathrm{P}[x > t] = e^{-t}$ for $t \geq 0$ and rescales each f_j using the an independent exponential distribution to get z_j so that with good probability $||z||_\infty$ is a constant factor approximation to $||f||_p$. The algorithm runs a variant of the Count sketch on z in order to approximate $||z||_\infty$. Though the Count sketch does not yield constant factor approximations in general, z is sufficiently skewed that this variant does work well.

More precisely, for each j, define $z_j = f_j/u_j^{1/p}$ where $u_j \sim Exp(1)$ are chosen independently.

The vector z will be part of our analysis but does not directly appear in the description of the algorithm as a streaming algorithm.

Max-stable algorithm for $||f||_p$ approximation:

<u>Initialize:</u>
$k \leftarrow \lceil M^{1-2/p} \log_2 M \rceil$
$y \leftarrow$ length k vector of real numbers
Use Nisan generator to approximate the following random choices:
Choose $u_1, \ldots, u_M \sim Exp(1)$ independently.
Choose $h : [M] \to [k]$ uniformly at random.
Choose $g : [M] \to \{-1, 1\}$ uniformly at random.
<u>Process:</u>
for each i **do**
    $y_{h(x_i)} \leftarrow y_{h(x_i)} + c_i \cdot g(x_i)/u_{x_i}^{1/p}$
end for
<u>Output:</u> $||y||_\infty = \max\{y_x \ : \ a \in [k]\}$.

We will show that the above algorithm produces a factor 4 approximation, say, for $||f||_p$ with probability bounded above 1/2 and hence using the usual median, running $O(\log(1/\delta))$ copies in parallel yields a factor 4 approximation with probability at least $1 - \delta$.

This algorithm corresponds to a sketch matrix of the following form:

$$\begin{bmatrix} 0 & 0 & -1/u_3^{1/p} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & +1/u_2^{1/p} & 0 & 0 & 0 & \cdots & -1/u_{M-2}^{1/p} & 0 & 0 \\ 0 & 0 & 0 & -1/u_4^{1/p} & 0 & \cdots & 0 & 0 & 0 \\ +1/u_1^{1/p} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & -1/u_M^{1/p} \\ 0 & 0 & 0 & 0 & +1/u_5^{1/p} & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & -1/u_{M-1}^{1/p} & 0 \end{bmatrix}$$

which is the product of

$$P_{g,h} = \begin{bmatrix} 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & \cdots & \cdots & -1 & 0 & 0 & 0 \\ 0 & +1 & 0 & 0 & 0 & 0 & +1 & 0 & \cdots & \cdots & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ +1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & +1 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots & 0 & 0 & -1 & 0 \end{bmatrix}$$

and

$$D_u = \begin{bmatrix} 1/u_1^{1/p} & & & & & & & & \\ & 1/u_2^{1/p} & & & & & & & \\ & & 1/u_3^{1/p} & & & & & & \\ & & & 1/u_4^{1/p} & & & & & \\ & & & & 1/u_5^{1/p} & & & & \\ & & & & & \ddots & & & \\ & & & & & & 1/u_{M-2}^{1/p} & & \\ & & & & & & & 1/u_{M-1}^{1/p} & \\ & & & & & & & & 1/u_M^{1/p} \end{bmatrix}.$$

The vector $z = P_{g,h} \cdot f$ and $y = D_u P_u f$.

Last time we proved

Claim 1: $\mathbb{P}[\frac{||f||_p}{2} \leq ||z||_\infty \leq 2||f||_p] > 3/4$.

In order to prove that $||y||_\infty$ is a good estimate for $||z||_\infty$ we need to show that z is sufficiently skewed.

Claim 2: For any H, $\mathbb{E}[\#\{j \ : \ |z_j| \geq \frac{|f||_p}{H}\}] \leq H^p$.

Proof. Let $Y_j = \begin{cases} 1 & \text{if } |z_j| \geq ||f||_p/H \\ 0 & \text{otherwise.} \end{cases}$ Then

$$\begin{aligned}
\mathbb{E}(\sum_j Y_j) &= \sum_j \mathbb{P}[|z_j| \geq \frac{||f||_p}{H}\} \\
&= \sum_j \mathbb{P}[\frac{|f_j|^p}{u_j} \geq \frac{||f||_p^p}{H^p}\} \\
&= \sum_j \mathbb{P}[u_j \geq \frac{H^p \cdot |f_j|^p}{||f||_p^p}\} \\
&= \sum_j (1 - e^{-\frac{H^p \cdot |f_j|^p}{||f||_p^p}\}}) \qquad \text{since } u_j \sim Exp(1) \\
&\leq \sum_j \frac{H^p \cdot |f_j|^p}{||f||_p^p}\} \qquad \text{since } e^{-x} \geq 1 - x \\
&= H^p
\end{aligned}$$

$\square$

Therefore, by Markov's inequality

$$\mathbb{P}[\#\{j \; : \; |z_j| \geq \frac{|f||_p}{H}\} \geq 100H^p] \leq 1/100.$$

We choose $H = c\log_2 M$ for some constant $c > 0$ and let $K = 100H^p$.

Fix $u = (u_1, \ldots, u_M)$ which fixes z. Call j heavy if $|z_j| > ||f||_p/H$ and let $L \subseteq [M]$ be the set of light (non-heavy elements in $[M]$.

Now for $p > 2$, $k = cM^{1-2/p}\log_2 M$ is $M^{\Omega(1)}$ and so is larger than the $50K^2$ for sufficiently large M, so the probability that any two of the heavy elements collide under h is at most $1/100$.

In order to show that $||y||_\infty$ approximates $||z||_\infty$ well, all we need to show is that the contribution of the light elements won't affect the contribution of any heavy j element of z_j by too much.

Let $a \in [k]$. Then $\mathbb{E}_g(\sum_{j \in L,\, h(j)=a} g(j)z_j) = \sum_{j \in L,\, h(j)=a} \mathbb{E}_g(g(j))z_j = 0$ for each fixed h.

Therefore

$$\begin{aligned}
Var_g, h(\sum_{j\in L,\ h(j)=a} g(j)z_j) &= \mathbb{E}_g, h((\sum_{j\in L,\ h(j)=a} g(j)z_j)^2) \\
&= \mathbb{E}_g, h(\sum_{i\in L,\ h(j)=a} \sum_{j\in L,\ h(j)=a} g(i)g(j)z_i z_j) \\
&= \mathbb{E}_h(\sum_{i\in L,\ h(j)=a} \sum_{j\in L,\ h(j)=a} \mathbb{E}_g(g(i)g(j))z_i z_j) \\
&= \mathbb{E}_h(\sum_{j\in L,\ h(j)=a} z_j^2) \qquad \text{by pairwise independence} \\
&\leq \mathbb{E}_h(\sum_{h(j)=a} z_j^2) \\
&= \frac{\sum_j z_j^2}{k} \\
&= ||z||_2^2/k.
\end{aligned}$$

Now in order to understand the variance of the contribution of the light elements overall, we let u vary.

$$\begin{aligned}
\mathbb{E}_u(||z||_2^2) &= \sum_j \mathbb{E}(\frac{f_j^2}{u_j^{2/p}}) \\
&= \sum_j f_j^2 \mathbb{E}(\frac{1}{u_j^{2/p}}) \\
&\leq c'||f||_2^2
\end{aligned}$$

for some constant $c' = \int_0^\infty e^{-\lambda}/\lambda^{2/p}\mathrm{d}\lambda$ since $u_j \sim Exp(1)$.

Now $||z||_\infty$ is roughly $||f||_p$ so we need the variance to be small relative to $||f||_p^2$ rather than $||f||_2^2$. Therefore k needs to be small enough to reduce $||f||_2^2$ sufficiently to achieve this. To relate these two we use Hölder's Inequality.

Proposition 0.1 (Hölder's Inequality)**.** *For arbitrary vectors* u *and* v*,* $\langle u, v\rangle \leq ||u||_p \cdot ||v||_q$ *for* $\frac{1}{p} + \frac{1}{q} = 1$.

We apply Hölder's inequality to the vectors $(f_1^2, \ldots, f_M^2)$ and $(1, \ldots, 1)$ and $p' = p/2$, $q' = 1/(1-$

$1/p') = 1/(1-2/p)$: Then

$$\begin{aligned} ||f||_2^2 &= \sum_j f_j^2 \cdot 1 \\ &= (\sum_j (f_j^2)^{p/2})^{2/p} \, (\sum_j 1^{1/(1-2/p)})^{1-2/p} \\ &= (\sum_j f_j^p)^{2/p} \, M^{1-2/p} \\ &= ||f_j||_p^2 \cdot M^{1-2/p}. \end{aligned}$$

Therefore the variance of the contribution of the light elements is $\leq \frac{c'||f||_p^2 M^{1-2/p}}{k}$. With our choice of k, we get variance for each single bucket $a \in [k]$ at most $\frac{\varepsilon ||f||_p^2}{\log_2 M}$.

Now, because the expectation for a bucket is 0, and it is given by a sum of independent random variables with total variance is at most $\frac{\varepsilon ||f||_p^2}{\log_2 M}$, we can apply a variant of Chernoff bounds which says that the probability that such a random variable is at least K standard deviations aboves its mean decays exponentially in K^2 to show that the probability that a single bucket has a contribution at least $||f||_p/10$ from light elements is at most $1/(100M)$ for ε sufficiently small. By a union bound, except with probability $1/100$, every bucket has a contribution at most $||f||_p/10$ from light elements. Together with the fact that the heavy elements are hashed to distinct bins except with probability $1/100$ we get that $||y||_\infty$ is between $||f||_p/3$ and $3||f||_p$ except with probability $1/3$.

Finally, we run $O(\log(1/\delta))$ independent copies of the protocol and take the median of the answers to derive a constant factor approximation with probability at least $1 - \delta$.