15-859 Assignment #3

Albert Gao / sixiangg / section 1A

11/03/2022

Task 1

Proof. Note that to prove claim for all $\mathbb{R}^d$ we just need to show for all $x \in \mathbb{R}^d$ with $\|x\|_1 = 1$ by a scaling argument. We first fix x with $\|x\|_1 = 1$ and try to show SAx good. We focus on just the ith row of SAx and observe that

$$\begin{aligned}
(SAx)_i &= \langle (Z_1/m, \cdots, Z_n/m)^T, Ax \rangle && Z_j \text{i.i.d. standard Cauchy} \\
&= \sum_{j=1}^{n} \frac{(Ax)_j}{m} Z_j \\
&= \left\| \frac{Ax}{m} \right\|_1 Z && \text{1 norm invariance} \\
\Rightarrow |(SAx)_i| &= \frac{1}{m} \|Ax\|_1 |Z|.
\end{aligned}$$

Using the cdf of Cauchy we have that

$$\begin{aligned}
\Pr[|Z| < 1 - \epsilon] &= \frac{2}{\pi} \arctan(1 - \epsilon) \\
&\leq \frac{2}{\pi} \left(\frac{\pi}{4} - \frac{\epsilon}{2} \right) \\
&\leq \frac{1}{2} - \frac{\epsilon}{6}.
\end{aligned}$$

The inequality uses observation that arctan is concave on $x \geq 0$ and is thus upper bounded by the tangent line at $x = 1$. Further, for $\epsilon < \sqrt{3} - 1$, $\arctan(1 + \epsilon)$ is lower bounded by line segment from $(1, \pi/4), (\sqrt{3}, \pi/3)$. Hence

$$\begin{aligned}
\Pr[|Z| < 1 + \epsilon] &= \frac{2}{\pi} \arctan(1 + \epsilon) \\
&\geq \frac{2}{\pi} \left(\frac{\pi}{4} + \frac{\epsilon}{\sqrt{3} - 1} \left(\frac{\pi}{3} - \frac{\pi}{4} \right) \right) \\
&= \frac{1}{2} + \frac{2\epsilon}{\sqrt{3} - 1} \frac{1}{12} \\
&\geq \frac{1}{2} + \frac{\epsilon}{6} \\
\Rightarrow \Pr[|Z| > 1 + \epsilon] &\leq \frac{1}{2} - \frac{\epsilon}{6}.
\end{aligned}$$

Then let X_1 denote the sum of m i.i.d. Bernoulli r.v. each of which is 1 iff $|Z| < 1 - \epsilon$. Let X_2

denote sum of m Bernoullis each of which is 1 iff $|Z| > 1+\epsilon$. Then

$$\begin{aligned}
\Pr[X_1 \geq m/2] &= \Pr[X_1 \geq (1+\delta)m(1/2-\epsilon/6)] && \delta = \frac{\epsilon/6}{1/2-\epsilon/6} \\
&\leq \Pr[X_1 \geq (1+\delta)\mathbb{E}X_1] && \text{by analysis above} \\
&\leq \exp\left(-\delta^2 \mathbb{E}X_1/(2+\delta)\right) && \text{Chernoff} \\
&\leq \exp\left(-\delta^2 m/4(2+\delta)\right) && \text{for small enough constant } \epsilon \\
&\leq \exp\left(-(\epsilon^2/9)m(1/12)\right) && \delta \in [\epsilon/3, 1] \text{ for small enough } \epsilon \\
&= \exp(-\epsilon^2 m/108).
\end{aligned}$$

Similarly we have $\Pr[X_2 \geq m/2] \leq \exp(-\epsilon^2 m/108)$. Then

$$\Pr[\|SAx\|_{\text{med}} < (1-\epsilon)\|Ax\|_1/m] \leq \Pr[X_1 \geq m/2] \leq \exp(-\epsilon^2 m/108).$$

Union bound gives

$$\begin{aligned}
\Pr[\|SAx\|_{\text{med}} \notin (1\pm\epsilon)\|Ax\|_1/m] &\leq 2\exp(-\epsilon^2 m/108) \\
&= 2\exp(-\Theta(d\log(d/\epsilon))),
\end{aligned}$$

taking $m = \Theta(d\log(d/\epsilon)/\epsilon^2)$. Also, by hint we may assume there is a γ-net for $\{Ax : \|x\|_1 = 1\}$ of size $(d/\gamma)^{O(d)}$ while taking $\gamma = \frac{\epsilon^3}{d^3\log^2(d/\epsilon)}$. Then by union bound we get that the probability that any vector in the net fails is at most

$$\begin{aligned}
(d/\gamma)^{O(d)} 2\exp(-\Theta(d\log(d/\epsilon))) &= (d^4\log^2(d/\epsilon)/\epsilon^3)^{O(d)} 2\exp(-\Theta(d\log(d/\epsilon))) \\
&= 2\exp(O(d)\log(d^4\log^2(d/\epsilon)/\epsilon^3))\exp(-\Theta(d\log(d/\epsilon))) \\
&\leq 2\exp(O(d)\log(d^6/\epsilon^5))\exp(-\Theta(d\log(d/\epsilon))) \\
&= 2\exp(O(d\log(d/\epsilon)) - \Theta(d\log(d/\epsilon))).
\end{aligned}$$

Note that the constant in $O(d\log(d/\epsilon))$ is at universal and does not dependent on γ, so we can choose constants in $\Theta(d\log(d/\epsilon))$ large enough (by choosing the constants for m large enough) so that the probability of failure is still exponentially small in $\Theta(d\log(d/\epsilon))$.

Now, take an arbitrary vector x such that $\|x\|_1 = 1$. Take a vector y in the γ-net such that $\|Ax - y\|_1 \leq \gamma$. Note we have that $\|SAx\|_\infty \leq \|SAx\|_1 \leq O(d\log d)\|Ax\|_1$ for all x for constant $9/10$ probability, by hints 1 and 3. Also we may assume that A is an Auerbach basis since proof for any basis extends to all x for original A since the column spans are the same. Then so

$$\begin{aligned}
\|SAx\|_{\text{med}} &= \|Sy + S(Ax-y)\|_{\text{med}} \\
&\in \|Sy\|_{\text{med}} \pm \|S(Ax-y)\|_\infty \\
&\subseteq \|Sy\|_{\text{med}} \pm O(d\log d)\|Ax-y\|_1 \\
&\subseteq \|Sy\|_{\text{med}} \pm O(d\log d)\gamma \\
&= \|Sy\|_{\text{med}} \pm O\left(d\log(d/\epsilon)\frac{\epsilon^3}{d^3\log^2(d/\epsilon)}\right) \\
&= \|Sy\|_{\text{med}} \pm O\left(\frac{\epsilon^3}{d^2\log(d/\epsilon)}\right) \\
&= \|Sy\|_{\text{med}} \pm O\left(\frac{\epsilon}{dm}\right) \\
&= \|Sy\|_{\text{med}} \pm O(\|x\|_\infty \epsilon/m) && \|x\|_1 = 1 \\
&= \|Sy\|_{\text{med}} \pm O\left(\epsilon\frac{\|Ax\|_1}{m}\right). && \text{Auerbach basis}
\end{aligned}$$

Further, since $\gamma = O\left(\epsilon\frac{\|Ax\|_1}{m}\right)$ for same reasoning above we have

$$\begin{aligned}\|SAx\|_{\text{med}} &\in \|Sy\|_{\text{med}} \pm O\left(\epsilon\frac{\|Ax\|_1}{m}\right) \\ &\subseteq (1\pm\epsilon)(\|y\|_1/m) \pm O(\epsilon\|Ax\|_1/m) \\ &\subseteq (1\pm\epsilon)(\|Ax\|_1/m + \|Ax-y\|_1/m) \pm O(\epsilon\|Ax\|_1/m) \\ &\subseteq (1\pm\epsilon)(\|Ax\|_1/m + \gamma/m) \pm O(\epsilon\|Ax\|_1/m) \\ &= (1\pm O(\epsilon))(\|Ax\|_1/m).\end{aligned}$$

Finally, we can scale $O(\epsilon)$ to have constant 1 by scaling constants chosen for m appropriately and achieve our desired result. $\square$

Task 2

Proof. Suppose we just use a countsketch matrix S of dimension $k \times n$ with $k = O(\gamma\epsilon^{-2})$, kept track of using seeds for the hashes. Then total space usage is just $O(\gamma\epsilon^{-2}\log(n))$ due to bit sizes for each entry. To be more precise, we have a dimension k vector y initialized to 0. For each update $x_i \leftarrow x_i + a$ we perform $y \leftarrow y + S(ae_i) = y + aS_{*i}$. Note we don't store the column S_{*i} anywhere and we simply find the location of vector y to update. We eventually obtain Su, Sv, and a $(1 \pm \epsilon/2)$ approximation a to $\|u\|_2^2$ and we want to approximate $\|u+v\|_2^2$. To do this, we may output

$$a + \|Sv\|_2^2 + 2\langle Su, Sv\rangle.$$

We shall argue that given corret constants for k we shall obtain a correct approximation. Given our S, the JL property from class still holds. That is, for all $\|x\|_2 = 1$, $\mathbb{E}_S \left|\|Sx\|_2^2 - 1\right|^2 \le \epsilon^2\delta$, so long as $k = \frac{2}{\epsilon^2\delta}$. In the proof for approximate matrix product from class we have that for arbitrary vectors x, y,

$$\begin{aligned}
\frac{\mathbb{E}|\langle Sx, Sy\rangle - \langle x, y\rangle|}{\|x\|_2\|y\|_2} &\le 3\epsilon\delta^{1/2} \\
\Pr[|\langle Sx, Sy\rangle - \langle x, y\rangle| \ge 60\epsilon\delta^{1/2}\|x\|_2\|y\|_2] &\le \frac{1}{20}.
\end{aligned}$$

Take $\delta = c/\gamma$ where c is a constant gives us correct k. We determine this c later. Hence with at least $9/10$ probability (by union bound) we have that

$$\begin{aligned}
&\begin{cases} |\langle Su, Sv\rangle - \langle u, v\rangle| \le 60\epsilon\delta^{1/2}\|u\|_2\|v\|_2 \le 60\epsilon\delta^{1/2}\sqrt{\gamma}\|u\|_2^2 \\ |\langle Sv, Sv\rangle - \langle v, v\rangle| \le 60\epsilon\delta^{1/2}\sqrt{\gamma}\|u\|_2^2 \end{cases} \\
\Rightarrow a + \|Sv\|_2^2 + 2\langle Su, Sv\rangle &= (1 \pm \epsilon/2)\|u\|_2^2 + \|v\|_2^2 + 2\langle u, v\rangle \pm 180\epsilon\delta^{1/2}\sqrt{\gamma}\|u\|_2^2 \\
&= \|u\|_2^2 + \|v\|_2^2 + 2\langle u, v\rangle \pm (\epsilon/2 + 180\epsilon\sqrt{c})\|u\|_2^2 \\
&= \|u+v\|_2^2 \pm \epsilon\|u\|_2^2 && c = 1/360^2 \\
&= (1 \pm \epsilon)\|u+v\|_2^2.
\end{aligned}$$

Hence our algorithm behaves as desired. □