CS 15-851: Algorithms for Big Data	Spring 2025
Lecture 5 Part 1 — 2/13/2025	
Prof. David Woodruff	*Scribe: Oswaldo Ramirez*

1 Leverage Score Sampling

We're looking at another subspace embedding: Sampling based. If A is sparse then SA is sparse. $A = U\Sigma V^\top$ be an $n \times d$ matrix d rank. The i^{th} leverage score: $\ell(i)$ of A to be $|U_{i,*}|_2^2$.

Leverage score distribution depends on the matrix.

Let $(q_1, \ldots, q_n)$ be a distribution with $q_i \geq \frac{\beta \ell(i)}{d}$, where β is a parameter (think of $\beta = 1/2$).

Define sampling matrix $S = D \cdot \Omega^\top$, where D is a $k \times k$ rescaling matrix and Ω is a sampling matrix. Note: the rescaling matrix is for the norm when sampling.

Leverage score doesn't depend on the choice of orthonormal basis U for A. Let U, U' be orthonormal bases.

Claim 1. $|e_i U|_2^2 = |e_i U'|_2^2 \quad \forall i \in [n]$

Proof. Since U, U' have the same column space as A, there exists a change of basis matrix Z such that $U = U'Z$. Because U, U' have orthonormal columns, Z must be rotation matrix (orthonormal rows and columns).

$$|Ux|_2^2 = |U'Zx|_2^2 \qquad \text{(Substitution)}$$

The left-hand side is equal to $|x|_2^2$ while the right-hand size is $|Zx|_2^2$ because U, U' have orthonormal columns. Therefore, Z is a $d \times d$ orthonormal matrix. We can conclude that

$$|e_i U|_2^2 = |e_i U'Z|_2^2 = |e_i U'|_2^2$$

■

1.1 Leverage Score Sampling gives a Subspace Embedding

We want to show for $S = D \cdot \Omega^\top$, that

$$|SAx|_2^2 = (1 \pm \epsilon)|Ax|_2^2 \quad \forall x.$$

Well, let's write A in terms of SVD. Then, it's equivalent to showing

$$\begin{aligned} |Sy|_2^2 &= (1 \pm \epsilon)|Uy|_2^2 && (y = \Sigma V^\top x) \\ &= (1 \pm \epsilon)|y|_2^2 && (U \text{ has orthonormal columns}) \end{aligned}$$

for all y. It suffices to show $|U^\top S^\top SU - I|_2 \leq \epsilon$ with high probability. However, to analyze $U^\top S^\top SU$, we'll use Matrix Chernoff Bound (Lecture 4 slide 79).

First, we'll define a few items to be used for Matrix Chernoff.

- Let $i(j)$ denote the index of U sampled in the j^{th} trial.
- Let $X_j = I_d - \frac{U_{i(j)}^\top U_{i(j)}}{q_{i(j)}}$, where $U_{i(j)}$ is the j^{th} sampled row of U.
- The X_j are independent copies of a symmetric random variable.

Next, we'll need to evaluate a few items and satisfy conditions.

$$\begin{aligned}
\mathbf{E}[X_j] &= I_d - \sum_i q_i \frac{U_i^\top U}{q_i} \\
&= I_d - I_d && (U \text{ has orthonormal columns}) \\
&= 0
\end{aligned}$$

$$\begin{aligned}
|X_j|_2 &\leq |I_d| + \frac{|U_{i(j)}^\top U_{i(j)}|}{q_{i(j)}} \\
&\leq 1 + \max_i \frac{|U_i|_2^2}{q_i} && (\text{Definition of Leverage Score}) \\
&\leq 1 + \frac{d}{\beta} && (q_i \geq \tfrac{\beta \cdot \ell(i)}{d})
\end{aligned}$$

Finally,

$$\begin{aligned}
\mathbf{E}[X^\top X] &= I_d - 2\mathbf{E}\left[\frac{U_{i(j)}^\top U_{i(j)}}{q_{i(j)}}\right] + \mathbf{E}\left[\frac{U_{i(j)}^\top U_{i(j)} U_{i(j)}^\top U_{i(j)}}{q_{i(j)}^2}\right] \\
&= I_d - 2I_d + \mathbf{E}\left[\frac{U_{i(j)}^\top U_{i(j)} U_{i(j)}^\top U_{i(j)}}{q_{i(j)}^2}\right] \\
&= \mathbf{E}\left[\frac{U_{i(j)}^\top U_{i(j)} U_{i(j)}^\top U_{i(j)}}{q_{i(j)}^2}\right] - I_d \\
&= \sum_i \frac{U_i^\top U_i U_i^\top U_i}{q_i} - I_d \\
&\leq (\frac{d}{\beta} \sum_i U_i^\top U_i - I_d) \\
&\leq (\frac{d}{\beta} - 1) I_d.
\end{aligned}$$

Let's show why $\sum_i \frac{U_i^\top U_i U_i^\top U_i}{q_i} \leq \frac{d}{\beta} \sum_i U_i^\top U_i x$. First, we must recall $A \leq B$ for square matrices A, B.

$$A \leq B \iff \forall x, \quad x^\top A x \leq x^\top B x.$$

We have

$$x^\top \sum_i \frac{U_i^\top U_i U_i^\top U_i}{q_i} x \leq \frac{d}{\beta} x^\top \sum_i U_i^\top U_i x.$$

The right-hand size is equal to $\sum_i \frac{d}{\beta} \langle U_i, x \rangle^2$ while the left-hand size is equal to $\sum_i \frac{|U_i|_2^2}{q_i} \langle U_i, x \rangle^2$. Then, it boils down to showing $\frac{|U_i|_2^2}{q_i} \leq \frac{d}{\beta}$ but this comes from the fact that $\frac{|U_i|_2^2}{d} \leq q_i$.

Therefore, we see $|\mathbf{E}[X^\top X]| \leq \frac{d}{\beta} - 1$.

After collecting everything we can get back to Matrix Chernoff. Let's find W.

$$\begin{aligned} |W|_2 &= \left| \frac{1}{k} \sum_{j=1}^{k} \left(I_d - \frac{U_{i(j)}^\top U_{i(j)}}{q_{i(j)}} \right) \right|_2 \\ &= \left| I_d - \sum_{j=1}^{k} \frac{U_{i(j)}^\top U_{i(j)}}{k \cdot q_{i(j)}} \right|_2 \\ &= |I_d - U^\top S^\top S U|_2 \end{aligned}$$

By Matrix Chernoff,

$$\mathbf{P}\left[|I_d - U^\top S^\top SU| > \epsilon \right] \leq 2d \cdot e^{-k\epsilon^2 \Theta(\frac{\beta}{d})}.$$

We set $k = \Theta(\frac{d \cdot log(d)}{\beta \epsilon^2})$.

Note: we need $\Omega(dlg(d))$ samples. Consider the matrix

$$A = \begin{pmatrix} I_d \\ 0 \end{pmatrix},$$

where A is a $(n+d) \times d$ matrix. The Coupon collector says $\Omega(d \cdot lg(d))$ samples.

1.2 Fast Computation of Leverage Scores

We need to compute SVD but as we know that is slow. Suppose we compute SA for subspace embedding S. Let $SA = QR^{-1}$, where Q has orthonormal columns. Similar to the previous lecture, we'll define a sketch.

Set $\ell_i' = |e_i AR|_2^2$. Since AR has column span of A, $AR = UT^{-1}$. We end up wanting to show T^{-1} is a rotation matrix to preserve the norm.

Claim 2. $(1 \pm O(\epsilon))|x|_2 = |ARx|_2 = |UT^{-1}|_2 = |T^{-1}x|_2$

Proof.

$$\begin{aligned} (1-\epsilon)|ARx|_2 &\leq |SARx|_2 \\ &= |x|_2 && (Q = SAR) \end{aligned}$$

$$\begin{aligned} (1+\epsilon)|ARx|_2 &\geq |SARx|_2 \\ &= |x|_2 && (Q = SAR) \end{aligned}$$

■

Therefore, $(1 \pm O(\epsilon))|x|_2 = |ARx|_2$. Because $AR = UT^{-1}$ and U has orthonormal columns, we get $|ARx|_2 = |UT^{-1}|_2 = |T^{-1}x|_2$.

Let's get back to the normal leverage score:

$$\ell_i = |e_i ART|_2^2 = (1 \pm O(\epsilon))|e_i AR|_2^2 = (1 \pm O(\epsilon))\ell_i'.$$

However, we still need to compute AR and we don't know anything about R. As a solution, we'll sketch R on the right-hand side.

- $\ell_i = (1 \pm O(\epsilon))\ell_i'$. It suffices to set this ϵ to be constant. $\ell_i' = |e_i AR|_2^2$ takes too long.
- Let G be a $d \times O(log(n))$ matrix of i.i.d. normal random variables.
- For any vector z, $\mathbf{P}[|zG|_2^2 = (1 \pm 1/2)|z|_2^2] \geq 1 - \frac{1}{n^2}$. Instead, set $\ell_i' = |e_i ARG|_2^2$.

Note:

- Can compute in (nnz(A) $+d^2)log(n)$ time.
- Can solve regression in nnz$(A)log(n)$+poly$(d(log(n))/\epsilon)$ time.