Lecture 2-1 — 14th September , 2017

Prof. David Woodruff *Scribe: Nimrah Shakeel*

Choosing the right sketching matrix S

Our goal is to choose a suitable sketching matrix S so that S*A can be computed in O(ndlog(n)) time. To this end, we choose a matrix S = P*H*D, such that P is a diagonal matrix with ± 1 on the diagonals, H is a Hadamard matrix and P is a matrix that chooses a (small) subset of rows of P*H. We can view P as having one 1 per row, where the column of each 1 indicates a row of P*H which is to be selected.

We are assuming that the columns of A are orthonormal. We have to show that such a sketching matrix satisfies the subspace embedding property i.e. for any unit vector x, the distortion due to transformation by S is bounded by $\pm\epsilon$.

Claim 1. $\mid SAX \mid_2{}^2 = \mid PHDAx \mid_2{}^2 = 1 \pm \epsilon$
Proof: Since HD is a rotation matrix, we can say that multiplication by it does not change the norm

$$\mid HDA \mid_2{}^2 = \mid Ax \mid_2{}^2 = 1, \forall x$$

Let $y = Ax$
We will make use of the flattening lemma.

Lemma 1. *For any particular y,*

$$Pr[\mid HDY \mid_\infty \geq C\sqrt{\frac{log(\frac{nd}{\delta})}{n}}] < \frac{\delta}{2d}$$

Proof: Let $C > 0$ be a constant. We will first show that for any $i \in [n]$, we can say that $[Pr[\mid HDY \mid_\infty > C\sqrt{\frac{log(\frac{nd}{\delta})}{n}}] < \frac{\delta}{2nd}$
Afterwards, we can apply a union bound over all i's.
Notice that any particular ith row of HDY can be written as

$$\mid (HDY)_i \mid = \sum_j H_{ij} D_{jj} y_j$$

Recall the Azuma-Hoeffding lemma:

Lemma 2. *: For zero mean bounded random variable Z_j such that $Z_j \leq \beta_j$ with probability 1 $[Pr[\mid \sum_j Z_j \mid > t] \leq 2e^{-(\frac{t^2}{2\sum_j(\beta_j)^2})}$*
Here , let $Z_j = H_{ij} D_{jj} y_j$.
Then Z_j's are bounded random variables with 0 mean and $\mid Z_j \mid \leq \frac{|y_j|}{\sqrt{n}} = \beta_j$ with probability 1, for all j.

Consequently, we have $\sum_j(\beta_j)^2 = \frac{1}{n}$ *, since y has unit norm.*
This gives us

$$Pr[|\sum_j Z_j |> C\sqrt{\frac{\log\frac{nd}{\delta}}{n}}] \leq 2e^{-\frac{C^2\log\frac{nd}{\delta}}{2}} \leq \frac{\delta}{2nd}$$

for a suitable C.

Consequences of the flattening lemma:

Remember that A has orthonormal columns ,and so has HDA.
The flattening lemma implies that for any particular $i \in d$

$$| HDAe_i |_\infty \leq C\sqrt{\frac{\log\frac{nd}{\delta}}{n}}$$

with probability $1-\frac{\delta}{2d}$
This gives us that $| e_jHDAe_i |_\infty \geq C\sqrt{\frac{\log\frac{nd}{\delta}}{n}}$ *with probability* $1-\frac{\delta}{2}$*, for all i,j.*
Further, we can say that $| e_jHDA |_2 \leq C\sqrt{\frac{d\log\frac{nd}{\delta}}{n}}$ *, for all j.*

Matrix Chernoff Bound

Let $X_1, ..., X_s$ *be independent copies of a symmetric random matrix* $X \in R^{dxd}$ *with*
$E[X] = 0$, $| X |_2 \leq \gamma$ *and* $| E[X^TX] |_2 \leq \Sigma^2$
Let $W = \frac{1}{s}\sum_{i\in[S]} X_i$
For any $\epsilon > 0$

$$Pr[|W|_2 > \epsilon] \leq 2d.e^{-s\epsilon^2 \ \Sigma^2+\frac{\gamma\epsilon}{3}}$$

where $|W|_2$ *denotes the spectral norm such that*

$$|W|_2 = sup\frac{|Wx|_2}{|x|_2}$$

Let $V = HDA$ *, then V is a matrix with orthonormal columns.*
Suppose P in $S = PHD$ *samples s rows uniformly with replacement. If row i is sampled in the jth sample,* $P_{ji} = \sqrt{\frac{n}{s}}$*, and is 0 otherwise.*
Let Y_i *be the ith sampled row of* $V = HDA$
Let $X_i = I_d - n.Y_i{}^TY_i$
Then

$$E[X_i] = I_d - n.\sum_j(\frac{1}{n})V_j{}^TV_j = I_d - V^TV = 0^{dxd}$$

$$|X_i|_2 \leq |I_d|_2 + n.max|e_jHDA|_2^2 = 1 + n.C^2\log\frac{nd}{\delta}.\frac{d}{n} = \Theta(d\log\frac{nd}{\delta})$$

$$E[X^TX+I_d] = I_d + I_d - 2nE[Y_i^T] + n^2E[Y_i^TY_iY_i^TY_i]$$
$$= 2I_d - 2I_d + n^2\sum_i(\frac{1}{n}).v_i^Tv_iv_i^Tv_i = n\sum_i v_i^Tv_i.|v_i|_2^2$$

Let us define

$$Z = n\sum_i v_i^Tv_iC^2\log(\frac{nd}{\delta}).\frac{d}{n} = C^2d\log(\frac{nd}{\delta})I_d$$

Note that $E[X^TX+I_d]$ and Z are real symmetric , with non-negative eigenvalues.

Claim 2. : For all vectors y, we have

$$y^TE[X^TX+I_d]y \leq y^TZy$$

Proof:

$$y_TE[X^TX+I_d]y = n\sum_i y^Tv_i^Tv_iy|v_i|_2^2 = n\sum_i v_i,y|v_i|_22$$

Also,

$$y_TZy = n\sum_i y^Tv_i^Tv_iyC^2\log(\frac{nd}{\delta}).\frac{d}{n} = d\sum_i v_iyC^2\log(\frac{nd}{\delta})$$

Hence

$$|E[X^TX]|_2 \leq |E[X^TX]+I_d|_2 + \mid I_d\mid_2 = \mid \exp X^Tx + I_d\mid_2 + 1$$
$$\leq |Z|_2 + 1 \leq C^2d\log(\frac{nd}{\delta}) + 1$$
$$|E[X^TX]|_2 = \omega(d\log(\frac{nd}{\delta}))$$

Recall the Matrix Chernoff bound which says that
For $X_1,...,X_s$ which are independent copies of a symmetric random matrix $X \in R^{dxd}$ with $E[X]=0$, $\mid X\mid_2 \leq \gamma$ and $|E[X^TX]|_2 \leq \sigma^2$
Let $W = \frac{1}{s}\sum_{i\in[S]}X_i$.
For any $\epsilon > 0$,

$$Pr[|W|_2 > \epsilon] \leq 2d.e^{-s\epsilon^2/(\sigma^2+\frac{\gamma\epsilon}{3})}$$

Then $Pr[|I_d - (PHDA)^T(PHDA)|_2 > \epsilon] \leq 2d.e^{-s\epsilon^2/(\Theta(d\log(\frac{nd}{\delta})))}$
If we set $s = d\log(\frac{nd}{\delta})\frac{\log(\frac{nd}{\delta})}{\epsilon^2}$ to make this probability less than $\frac{\delta}{2}$
This implies that for every unit vector x , $|1 - |PHDAx|_2^2| = |x^Tx - x^T(PHDA)^T(PHDA)x| \leq \epsilon$, so $|PHDAx|_2^2 \in 1\pm\epsilon$ for all unit vectors x.
Thus if we consider the column span of A adjoined with b, we can solve the regression problem in $\omega(nd\log n) + poly(\frac{d\log n}{\epsilon})$, which is nearly optimal in matrix dimensions i.e. when $n \gg d$.