~~Faster~~
~~Simpler~~ ~~SVD & Approximation~~

From Nick Harvey's notes

Fast Low-rank Approximation in the Spectral Norm

Schmidt-Eckert-Young-Mirsky says: Given A, the matrix A_k ensures $\|A - A_k\|_2 \le \sigma_{k+1}(A)$.

In particular, if $A = UDV^T$ and $V_k = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{pmatrix}$ is the top-k-right singular value vector, then $P_k = V_k V_k^T$ is the projection matrix. And hence

$$\|A - AP_k\|_2 \le \sigma_{k+1}(A).$$

Takes SVD time.

Faster? Here's one way. $A = n \times d$ matrix

Let $\tilde{A}$ be T samples from distribution that adds in row i (i.e. $\frac{a_i^T}{\sqrt{T}\sqrt{p_i}}$ w.p. p_i) (scaled)

Here $p_i = \dfrac{\|a_i\|^2}{\|A_F\|_2^2}$.

Compute the SVD $\tilde{A} = \tilde{U}\tilde{D}\tilde{V}^T$. Let $\tilde{V}_k$ be the top k right singular vectors of $\tilde{A}$.

$\tilde{P} := \tilde{V}_k \tilde{V}_k^T$.

Theorem: whp, $\|A - A\tilde{P}\|_2 \le \sigma_{k+1}(A) + \epsilon \|A\|_2$

loss over the bare minimum.

Main Lemma: for any ~~matrix~~ $(T \times d)$ matrix B, let $P_{B,k}$ be the projection onto the top k right singular vectors of B.

then $\|A - AP_{B,k}\|_2^2 \le \sigma_{k+1}(A)^2 + 2\|A^TA - B^TB\|_2$.

Pf: Later

So it suffices to show that whp, $\|A^TA - \tilde{A}^T\tilde{A}\|_2 \le (\epsilon^2/2)\|A\|_2$.

Pf: By scaling down, can assume that $\|A\|_2 = 1$.

$$A^TA = \sum_{k=1}^{n} a_k a_k^T \qquad \text{where } a_k^T \text{ was the } k^{th} \text{ row of } A.$$

Note that each sample adds in $\dfrac{a_k a_k^T}{T p_i}$ to $\tilde{A}^T\tilde{A}$ w.p. p_i

Let Y be equal to $\frac{a_k a_k^T}{p_k}$ w/p p_k. $(k \in [1..n])$

then $E[Y] = \sum_{k=1}^{n} a_k a_k^T = A^T A$.

Since $\|A\|_2 \le 1$ this means $A^T A$ (which is psd) has eigenvalues ≤ 1.

$\|E[Y]\|_2 \le 1$. Also $\|Y\|_2 \le \max_k \frac{\|a_k a_k^T\|_2}{p_k} = \|A\|_F^2$ by choice of p_k.

Now Let $Y_1, Y_2 \ldots Y_T$ be samples distributed according to Y, independently.

Each Y_i satisfies $\left.\begin{array}{l} \|Y_i\|_2 \le \|A\|_F^2 \\ \text{and } \|E(Y_i)\|_2 \le 1 \end{array}\right\}$

then by AHlswede Winter (AWS) we get

$$\Pr\Big[\Big\| \frac{1}{T}\sum_{i=1}^{T} Y_i - A^T A \Big\|_2 \ge \underbrace{\epsilon^2/2}_{= \epsilon^2/2 \|A\|_1 \text{ by our choice of scaling}} \Big] \le 2d \cdot \exp\left(- \frac{(\epsilon^2/2)^2 T}{4\|A\|_F^2}\right)$$

$$\le \delta$$

if we set $T = \frac{16 \|A\|_F^2 \ln \delta^{-1}}{\epsilon^4}$.

What is this T? $\Theta\left(\frac{\|A\|_F^2 \ln \delta^{-1}}{\epsilon^4}\right)$ but $\|A\|_2 = 1 = \sigma_1$

$\Rightarrow \|A\|_F^2 = \sum_{i=1}^{r} \sigma_i^2 \le r$ for sure

In fact $\frac{\|A\|_F^2}{\|A\|_2^2}$ is called the "stable" rank of A. $\le$ rank of A.

$\Rightarrow$ in fact we can set $T = \Theta\left(\frac{r \ln n}{\epsilon^4}\right)$ and get the approximation w/p $1 - 1/\text{poly}(n)$.

Finally it remains to prove the Main Lemma.

Recall: **Lemma**: Let $A \in \mathbb{R}^{n\times d}$. Let $B \in \mathbb{R}^{T\times d}$, and Let $P_{B,k}$ be the projection onto the top k right singular vectors of A.

then $\|A - AP_{B,k}\|_2^2 \le \sigma_{k+1}(A)^2 + 2\|A^TA - B^TB\|$.

Pf: Let Q be the projection onto the kernel of $P_{B,k}$. ($= P$ for short)

(ie onto the span of the bottom $d-k$ right singular vectors of B).

$$\|A - AP\|_2 = \|A(I-P)\|_2 = \|AQ\|_2$$

$$= \sup_{x : \|x\|=1} \|AQx\|$$

$$= \sup_{\substack{x \in \mathrm{span}(Q): \\ \|x\|=1}} \|A\cancel{Q}x\|.$$

so
$$\|A - AP\|^2 = \sup_{\substack{x \in \mathrm{Ker}(P): \\ \|x\|=1}} \|Ax\|^2$$

$$= \sup_{x \in \mathrm{Ker}(P) : \|x\|=1} \langle A^TAx, x\rangle$$

$$= \underline{\qquad} \left(\langle (A^TA - B^TB)x, x\rangle + \langle B^TBx, x\rangle\right)$$

$$\le \sup_{\substack{x \in \mathrm{ker}(P) \\ \|x\|=1}} \langle (A^TA - B^TB)x, x\rangle + \sup_{\substack{x \in \mathrm{Ker}(P): \\ \|x\|=1}} \langle B^TBx, x\rangle$$

$$\le \underbrace{\|A^TA - B^TB\|} + \sigma_{k+1}(B^TB)$$

eigenvalues are bounded by max singular value

this is where we use that P is projection on top k singular values of B, and hence on top k eigenspace of B^TB.

$$\le 2\|A^TA - B^TB\| + \sigma_{k+1}(A^TA) = \sigma_{k+1}^2(A) \,..$$

☺

Using Fact: X, Y symmetric with $\lambda_1(X) \ge \lambda_2(X) \ge \cdots \ge \lambda_n(X)$
$\lambda_1(Y) \ge \lambda_2(Y) \ge \cdots \ge \lambda_n(Y)$. then

$$\max_i |\lambda_i(X) - \lambda_i(Y)| \le \|X - Y\|_2$$