

15-859 Assignment #2

Albert Gao / sixiangg / section 1A
10/16/2022

Task 1

1. *Proof.* We may let $A = U\Sigma V^T$ be its singular value decomp. And if u_i is the i th row of U , written as a column vector, then we have

$$\begin{aligned} a_i^T &= u_i^T \Sigma V^T \\ \Rightarrow a_i^T (A^T A)^{-1} a_i &= a_i^T (V \Sigma U^T U \Sigma V^T)^{-1} a_i \\ &= u_i^T \Sigma V^T (V \Sigma^2 V^T)^{-1} V \Sigma u_i \\ &= u_i^T \Sigma V^T V (\Sigma^2)^{-1} V^T V \Sigma u_i \\ &= u_i^T u_i \\ &= l_i. \end{aligned}$$

Note that the invertibility of $A^T A$ comes from the fact that A has rank d . □

2. (a) *Proof.* Let's first show that $A_{i-1}^T A_{i-1} + \lambda I$ is invertible for any $i \geq 1$. Note that $x^T A_{i-1}^T A_{i-1} x \geq 0$ so $A_{i-1}^T A_{i-1}$ is positive semidefinite. It is also then clear that $S_i = A_{i-1}^T A_{i-1} + \lambda I$ is positive definite. This implies that S_i has strictly positive eigenvalues and thus is invertible. By Matrix Determinant Lemma we then have that

$$\begin{aligned} \det(S_{i+1}) &= \det(S_i + a_i a_i^T) \\ &= (\det S_i) (1 + a_i^T (A_{i-1}^T A_{i-1} + \lambda I)^{-1} a_i) \\ &\geq \det(S_i) (1 + l_i) \\ &\geq \det(S_i) e^{l_i/2}. \end{aligned}$$

By calculus and that $0 < l_i \leq 1$

Note that the first equality comes from the observation that for any matrices A, B we have $AB = \sum_k A_{*k} B_{k*}$. Then we may conclude that

$$\begin{aligned} \det(A^T A + \lambda I) &= \det(S_{n+1}) \\ &\geq \det(S_n) e^{l_n/2} \\ &\geq \det(S_1) e^{\sum_1^n l_i/2} \\ &= \det(\lambda I) \exp\left(\sum l_i/2\right). \end{aligned} \quad \square$$

- (b) *Proof.* We first note that determinant of a square matrix is the product of its eigenvalues, that singular values and eigen values coincide for real symmetric positive definite matrices, and that the top singular value is just the operator norm of the matrix. Combining these

facts gives

$$\begin{aligned}
 \det(A^T A + \lambda I) &= \prod_i \lambda_i && \text{the eigenvalues} \\
 &= \prod_i \sigma_i && \text{the singular values} \\
 &\leq \sigma_1^d && \text{positive definite} \\
 &= \|A^T A + \lambda I\|_2^d \\
 &= (\|A^T A\|_2 + \lambda)^d \\
 &= \left(\sup_x x^T A^T A x + \lambda \right)^d && \text{3.13 from 15-884 linalg resource} \\
 &= (\|A\|_2^2 + \lambda)^d. && \square
 \end{aligned}$$

Combining the above two lemmas gives us that

$$\begin{aligned}
 \det(\lambda I) \exp\left(\sum l_i/2\right) &\leq (\|A\|_2^2 + \lambda)^d \\
 \Rightarrow \exp\left(\sum l_i/2\right) &\leq \left(\frac{\|A\|_2^2 + \lambda}{\lambda}\right)^d \\
 \Rightarrow \sum l_i/2 &\leq d \log(1 + \|A\|_2^2/\lambda) \\
 \Rightarrow \sum l_i &= O(d \log(1 + \|A\|_2^2/\lambda)).
 \end{aligned}$$

Task 2

1. (a) *Proof.* Expanding with SVD gives us

$$\begin{aligned}
 y &= (AA^T)^r Ag \\
 &= (U\Sigma V^T V\Sigma U^T)^r Ag \\
 &= (U\Sigma^2 U^T)^r Ag \\
 &= U\Sigma^{2r} U^T U\Sigma V^T g \\
 &= U\Sigma^{2r+1} V^T g \\
 &= U\Sigma^{2r+1} g'
 \end{aligned}$$

where $g' = V^T g$ is a random vec with its coordinates independent normal, because V^T has orthonormal rows and we can apply rotational invariance from lecture 1's slides ($\langle u, v \rangle = 0 \Rightarrow \langle g, u \rangle, \langle g, v \rangle$ independent; sum of norm is norm gives each coordinate is norm of mean 0 and variance still 1). \square

- (b) *Proof.* Since we know $g'_i \sim N(0, 1)$, by hint we have

$$\begin{aligned}
 \Pr[|g'_i| \geq 2\sqrt{\log n}] &\leq e^{-2\log n} / (2\sqrt{\log n}) = \frac{1}{2n^2\sqrt{\log n}} \\
 \Rightarrow \Pr[\max_i |g'_i| \geq 2\sqrt{\log n}] &\leq \frac{1}{2n\sqrt{\log n}} = 1/\text{poly}(n).
 \end{aligned}$$

Further, we have for any α ,

$$\begin{aligned}
 \Pr[|g'_1| < \alpha] &= \frac{2}{\sqrt{2\pi}} \int_0^\alpha e^{-x^2/2} dx \\
 &\leq \frac{\sqrt{2}}{\pi} \int_0^\alpha 1 dx \\
 &= \frac{\sqrt{2}\alpha}{\pi} \\
 &= 1/\text{poly}(n). \quad \text{if we take } \alpha = 1/\text{poly}(n)
 \end{aligned}$$

Then combining the above gives us that with high probability $1 - 1/\text{poly}(n)$, both desired events occur. That is, $\max |g'_i| \leq 2\sqrt{\log n}$ and $|g'_1| \geq 1/\text{poly}(n)$. \square

- (c) *Proof.* Let u_i be the i th column of U . We then have that

$$\begin{aligned}
 \|y\|_2^2 &= \|U\Sigma^{2r+1}g'\|_2^2 \\
 &= \left\| U \begin{pmatrix} \sigma_1^{2r+1} g'_1 \\ \vdots \\ \sigma_n^{2r+1} g'_n \end{pmatrix} \right\|_2^2 \\
 &= \left\| \sum_{i=1}^n \sigma_i^{2r+1} g'_i u_i \right\|_2^2 \\
 &= \sum_{i=1}^n \|\sigma_i^{2r+1} g'_i u_i\|_2^2 \quad \text{Pythagorean since } U \text{ has orthonormal columns} \\
 &= \sum_{i=1}^n \sigma_i^{4r+2} (g'_i)^2.
 \end{aligned}$$

Similarly,

$$\begin{aligned}
\|y^T A\|_2^2 &= \|(g')^T \Sigma^{2r+1} U^T U \Sigma V^T\|_2^2 \\
&= \|(g')^T \Sigma^{2r+2} V^T\|_2^2 \\
&= \|V \Sigma^{2r+2} g'\|_2^2 \\
&= \sum_{i=1}^n \sigma_i^{4r+4} (g'_i)^2.
\end{aligned}$$

same reason as above

We can then conclude that

$$\begin{aligned}
\|u^T A\|_2^2 &= \left\| \frac{y^T}{\|y\|_2} A \right\|_2^2 \\
&= \frac{\|y^T A\|_2^2}{\|y\|_2^2} \\
&= \frac{\sum \sigma_i^{4r+4} (g'_i)^2}{\sum \sigma_i^{4r+2} (g'_i)^2}
\end{aligned}$$

as desired. \square

- (d) *Proof.* With probability $1 - 1/\text{poly}(n)$ we know $|g'_1| \geq 1/\text{poly}(n)$ and $\max |g'_i| \leq 2\sqrt{\log n}$ and so

$$\begin{aligned}
\|u^T A\|_2^2 &= \frac{\sum \sigma_i^{4r+4} (g'_i)^2}{\sum \sigma_i^{4r+2} (g'_i)^2} \\
&\geq \frac{\sigma_1^{4r+4} (g'_1)^2}{\sigma_1^{4r+2} (g'_1)^2 + \sum_{i=2}^n \sigma_i^{4r+2} 4 \log n} \\
&\geq \frac{\sigma_1^{4r+4} (g'_1)^2}{\sigma_1^{4r+2} (g'_1)^2 + 4(n-1) \sigma_2^{4r+2} \log n} \\
&= \sigma_1^2 - \frac{4(n-1) \sigma_1^2 \sigma_2^{4r+2} \log n}{\sigma_1^{4r+2} (g'_1)^2 + 4(n-1) \sigma_2^{4r+2} \log n} \\
&\geq \sigma_1^2 - \frac{4(n-1) \sigma_1^2 \sigma_2^{4r+2} \log n}{\sigma_1^{4r+2} (g'_1)^2} \\
&\geq \sigma_1^2 - \frac{4(n-1) \sigma_1^{4r+2} \sigma_2^2 \log n}{2^{4r} \sigma_1^{4r+2} (g'_1)^2} \\
&= \sigma_1^2 - \frac{4(n-1) \log n}{2^{4r} (g'_1)^2} \sigma_2^2 \\
&\geq \sigma_1^2 - \frac{4(n-1) \log n \text{poly}(n)^2}{16^{k \log(n/\gamma)}} \sigma_2^2 \\
&\quad r = k \log(n/\gamma) = \Omega(\log(n/\gamma)) \\
&\geq \sigma_1^2 - \frac{4(n-1) \log n \text{poly}(n)^2}{(n/\gamma)^k} \sigma_2^2 \\
&\geq \sigma_1^2 - \gamma^k \sigma_2^2 \\
&\quad \text{pick } k \text{ large enough so } n^k \geq 4(n-1) \log n \text{poly}(n)^2 \\
&\geq \sigma_1^2 - \gamma \sigma_2^2,
\end{aligned}$$

where last step follows from $\gamma \leq 1$. \square

(e) *Proof.* First we may observe that with $1 - 1/\text{poly}(n)$,

$$\begin{aligned}
\sum_{i=m+1}^n \sigma_i^{4r+2} (g'_i)^2 &\leq \sum_{i=m+1}^n \sigma_i^{4r+2} (4 \log n) \\
&\leq \sum_{i=m+1}^n (1 - \gamma)^{4r+2} \sigma_1^{4r+2} (4 \log n) \\
&\leq (n - m) e^{-(4r+2)\gamma} \sigma_1^{4r+2} (4 \log n) \\
&\leq (n - m) e^{-(4k \log(n)/\gamma + 2)\gamma} \sigma_1^{4r+2} (4 \log n) \\
&\leq (n - m) n^{-4k} \sigma_1^{4r+2} (4 \log n).
\end{aligned}$$

So so long as $(g'_1)^2 \geq 1/\text{poly}(n)$ for any f in $\text{poly}(n)$ we can find a large enough constant k such that

$$f/(g'_1)^2 \sum_{i=m+1}^n \sigma_i^{4r+2} (g'_i)^2 \leq \sigma_1^{4r+2}$$

It then follows that

$$\begin{aligned}
\sigma_1^2 \frac{1 + (\sigma_m/\sigma_1)^2 \Theta}{1 + \Theta + 1/\text{poly}(n)} &= \sigma_1^2 \frac{(g'_1)^2 \sigma_1^{4r+2} + (\sigma_m/\sigma_1)^2 \sum_{i=2}^m (g'_i)^2 \sigma_i^{4r+2}}{(g'_1)^2 \sigma_1^{4r+2} + \sum_{i=2}^m (g'_i)^2 \sigma_i^{4r+2} + (g'_1)^2 \sigma_1^{4r+2}/\text{poly}(n)} \\
&= \frac{(g'_1)^2 \sigma_1^{4r+4} + (\sigma_m)^2 \sum_{i=2}^m (g'_i)^2 \sigma_i^{4r+2}}{(g'_1)^2 \sigma_1^{4r+2} + \sum_{i=2}^m (g'_i)^2 \sigma_i^{4r+2} + (g'_1)^2 \sigma_1^{4r+2}/\text{poly}(n)} \\
&\leq \frac{\sum_{i=1}^n (g'_i)^2 \sigma_i^{4r+4}}{\sum_{i=1}^m (g'_i)^2 \sigma_i^{4r+2} + (g'_1)^2 \sigma_1^{4r+2}/\text{poly}(n)} \\
&\leq \frac{\sum_{i=1}^n (g'_i)^2 \sigma_i^{4r+4}}{\sum_{i=1}^m (g'_i)^2 \sigma_i^{4r+2} + \sum_{i=m+1}^n (g'_i)^2 \sigma_i^{4r+2}} \\
&= \|u^T A\|_2^2. \quad \square
\end{aligned}$$

(f) *Proof.* Since $\gamma \leq 1$ we have $\Omega(\log(n/\gamma)) \geq \Omega(\log(n)) \geq \Omega(\log(n)/\gamma)$. So by (d) we know in the case of $\sigma_2 \leq \sigma_1/2$ the desired inequality is satisfied. So let us only focus on the case when $\sigma_2 \geq \sigma_1/2$. Then

$$\begin{aligned}
\|u^T A\|_2^2 &\geq \sigma_1^2 \frac{1 + (\sigma_m/\sigma_1)^2 \Theta}{1 + \Theta + 1/\text{poly}(n)} \\
&= \sigma_1^2 - \sigma_1^2 \frac{\Theta + 1/\text{poly}(n) - (\sigma_m/\sigma_1)^2 \Theta}{1 + \Theta + 1/\text{poly}(n)} \\
&\geq \sigma_1^2 - 4\sigma_2^2 \frac{\Theta + 1/\text{poly}(n) - (\sigma_m/\sigma_1)^2 \Theta}{1 + \Theta + 1/\text{poly}(n)} \\
&\geq \sigma_1^2 - 4\sigma_2^2 \frac{\Theta + 1/\text{poly}(n) - (1 - \gamma)^2 \Theta}{1 + \Theta + 1/\text{poly}(n)} \\
&= \sigma_1^2 - 4\sigma_2^2 \frac{(2\gamma - \gamma^2)\Theta + 1/\text{poly}(n)}{\Theta + 1/\text{poly}(n) + 1} \\
&\geq \sigma_1^2 - 4\sigma_2^2 \frac{2\gamma\Theta + 1/\text{poly}(n)}{\Theta + 1/\text{poly}(n) + 1} \\
&\geq \sigma_1^2 - \sigma_2^2 (8\gamma + 1/\text{poly}(n)).
\end{aligned}$$

Instead of $(1 - \gamma)$ threshold we can have the threshold be such that m is the largest index with $\sigma_m \geq (1 - (1/16)\gamma)\sigma_1$. Also take $1/\text{poly}(n)$ to be $1/(8n^2)$. Then the proof of part (e) proceeds the same way and we get

$$\begin{aligned}
\|u^T A\|_2^2 &\geq \sigma_1^2 - 4\sigma_2^2 \frac{\Theta + 1/(8n^2) - (1 - (1/16)\gamma)^2 \Theta}{1 + \Theta + 1/(8n^2)} \\
&\geq \sigma_1^2 - 4\sigma_2^2 \frac{(1/8)\gamma\Theta + 1/(8n^2)}{1 + \Theta + 1/(8n^2)} \\
&\geq \sigma_1^2 - \sigma_2^2((1/2)\gamma + 1/(2n^2)) \\
&\geq \sigma_1^2 - \sigma_2^2(\gamma/2 + \gamma/2) \\
&= \sigma_1^2 - \gamma\sigma_2^2.
\end{aligned}$$

We're done. □

2. *Proof.* We may proceed with the power method on $(A^T, r = O(\log(n)/\gamma))$ (where γ to be specified later) to get unit vector w such that $\|w^T A^T\|_2^2 \geq \sigma_1^2 - \gamma\sigma_2^2$ with probability $1 - 1/\text{poly}(n)$. Setting final vectors $(u, v) = (Aw, w)$ then gives us that

$$\begin{aligned}
\|uv^T - A\|_F^2 &= \text{Tr}((uv^T - A)(vu^T - A^T)) \\
&= \text{Tr}(uu^T - uu^T - uu^T + AA^T) \\
&= \text{Tr}(-uu^T) + \sum_{i=1}^n \sigma_i^2 \\
&= -\|u\|_2^2 + \sum_{i=1}^n \sigma_i^2 \\
&= -\|Aw\|_2^2 + \sum_{i=1}^n \sigma_i^2 \\
&\leq -\sigma_1^2 + \gamma\sigma_2^2 + \sum_{i=1}^n \sigma_i^2 \\
&= \gamma\sigma_2^2 + \|A - A_1\|_F^2 \\
&\leq (\gamma\alpha + 1)\|A - A_1\|_F^2.
\end{aligned}$$

Then if we take $\gamma = \epsilon/\alpha$ we get the desired inequality. The runtime is $2r + 1$ matrix-vector multiplications to get v and another to get u , where the matrix used is always A or A^T so each multiplication is done in $\text{nnz}(A)$ time. Hence total time is $O(\text{nnz}(A)r) = O(\text{nnz}(A)\log(n)/\gamma) = O(\text{nnz}(A)\log(n)(1 + \alpha/\epsilon))$ (where we have plus 1 to ensure we can take r as an integer). □

Task 3

1. *Proof.* Let X^* be the optimal solution to

$$\min_X \|AX - B\|_F^2.$$

Let $B^* = AX^* - B$. Let $A' = SA, B' = SB, \widehat{B}' = A'\widehat{X} - B'$. Then we have

$$\begin{aligned} \|A\widehat{X} - B\|_F^2 &= \|A(\widehat{X} - X^*)\|_F^2 + \|B^*\|_F^2 && \text{Pythagorean} \\ &\leq 4\|A'(\widehat{X} - X^*)\|_F^2 + \|B^*\|_F^2 && \text{subspace embedding w/ prob 19/20} \\ &= 4\|(A'X^* - B') - (A'\widehat{X} - B')\|_F^2 + \|B^*\|_F^2 \\ &= 4\|(A'X^* - B') - \widehat{B}'\|_F^2 + \|B^*\|_F^2 \\ &= 4\|A'X^* - B'\|_F^2 - 4\|\widehat{B}'\|_F^2 + \|B^*\|_F^2 && \text{Pythagorean} \\ &\leq 4\|SB^*\|_F^2 + \|B^*\|_F^2 \\ &\leq 80\|B^*\|_F^2 + \|B^*\|_F^2 && \Pr[\|SB^*\|_F^2 \geq 20\|B^*\|_F^2] \leq \frac{1}{20} \\ &= O(1)\|AX^* - B\|_F^2 \end{aligned}$$

with probability at least $1 - \frac{2}{20} = \frac{9}{10}$ by union bound as desired. Further, we can write

$$\widehat{X} = (SA)^-(SB)$$

where $(SA)^-$ denotes the pseudoinverse of SA . Since S is a $O(d \log d) \times n$ leverage score sampling matrix, SB consists of $O(d \log d)$ (scaled and) sampled rows of B and each row of \widehat{X} is a linear combination of these rows by definition of row vector and matrix multiplication. \square

2. *Proof.* Let S be a $O(k \log k) \times n$ leverage score sampling matrix for \widehat{U} which is of dimension $n \times k$. As in hint, treat $A = \widehat{U}, B = A$ and let $\widehat{X} = (S\widehat{U})^-(SA)$. Applying the previous part gives us that

$$\begin{aligned} \|\widehat{U}(S\widehat{U})^-(SA) - A\|_F^2 &= \|\widehat{U}\widehat{X} - A\|_F^2 \\ &= O(1)\|\widehat{U}X^* - A\|_F^2 && \text{Part 1} \\ &\leq O(1)\|\widehat{U}\widehat{V} - A\|_F^2 && X^* \text{ optimal} \\ &= O(1)\|A - A_k\|_F^2. && \text{low rank approximation} \end{aligned}$$

This means that SA just works since $\widehat{U}(S\widehat{U})^-$ is one such solution for U and the minimum must be at most $O(1)\|A - A_k\|_F^2$. Note that to get R we need to express S as $D\Omega^T$ where D is for scaling and Ω^T is for sampling, and take $R = \Omega^T A$. \square

3. *Proof.* Consider the minimizing problem $\min_X \|R^T X - A^T\|_F^2$, we know from above that the minimum is $O(1)$ from $\|A - A_k\|_F^2$. We may use leverage score sampling for R^T to solve the minimizing problem with $S \in \mathbb{R}^{O(k \log k \log(k \log k)) \times n} = \mathbb{R}^{O(k \log^2 k) \times n}$. Let \widehat{X} be the optimal solution for

$$\min_X \|SR^T X - SA^T\|_F^2.$$

Then we can express $\hat{X} = (SR^T)^-(SA^T)$. It follows that

$$\begin{aligned}
 \|(AS^T)((SR^T)^-)^T R - A\|_F^2 &= \|R^T(SR^T)^-(SA^T) - A^T\|_F^2 \\
 &= O(1)\|R^T X^* - A^T\|_F^2 && \text{part 1} \\
 &= O(1)\|(X^*)^T R - A\|_F^2 \\
 &= O(1)\|A - A_k\|_F^2. && \text{part 2}
 \end{aligned}$$

Notice that AS^T are sampled $O(k \log^2 k)$ columns of A if we leave the scaling, so we've indeed found our valid C and R . Yay! \square