STA243—HW1

Zhikuan Quan (917800911)   Bohao Zou (917796070)

April 20, 2020

1 Problem

The purpose of this question is to brush-up your linear algebra background. Let $\boldsymbol{A}$ be an $n \times n$ square matrix. Show that the following statements are equivalent:

1. The columns of $\boldsymbol{A}$ are orthonormal vectors.
2. $\boldsymbol{A}\boldsymbol{A}^\top = \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.
3. The rows of $\boldsymbol{A}$ are orthonormal vectors.

Proof. To prove these equivalent expressions, three directive methods are needed: $(\mathbf{1}) \Rightarrow (\mathbf{2})$, $(\mathbf{2}) \Rightarrow (\mathbf{3})$ and $(\mathbf{3}) \Rightarrow (\mathbf{1})$:
Step 1: $(\mathbf{1}) \Rightarrow (\mathbf{2})$:
Set $A = (a_1, a_2, \cdots, a_n)$, where the columns of A $a_k \in \mathbb{R}^{n\times 1}, k = 1, 2, \cdots, n$. Since the columns of A are orthonormal vectors, $a_i^\top a_j = 1$ when $i = j$ and $a_i^\top a_j = 0$ when $i \neq j$:

$$A^\top A = \begin{pmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_n^\top \end{pmatrix} (a_1, a_2, \cdots, a_n) = \begin{pmatrix} a_1^\top a_1 & a_1^\top a_2 & \dots & a_1^\top a_n \\ a_2^\top a_1 & a_2^\top a_2 & \dots & a_2^\top a_n \\ \dots & \dots & \dots & \dots \\ a_n^\top a_1 & a_n^\top a_2 & \dots & a_n^\top a_n \end{pmatrix} = I_n$$

So we have $(A^\top)^{-1} = A$, and then $AA^\top = I_n$.
Step 2: $(\mathbf{2}) \Rightarrow (\mathbf{3})$:
Set $A = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$, where the row of A: $b_k \in \mathbb{R}^{1\times n}, k = 1, 2, \cdots, n$. Since $AA^\top = A^\top A = I_n$, we have:

$$I_n = AA^\top = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} (b_1^\top, b_2^\top, \cdots, b_n^\top) = \begin{pmatrix} b_1 b_1^\top & b_1 b_2^\top & \dots & b_1 b_n^\top \\ b_2 b_1^\top & b_2 b_2^\top & \dots & b_2 b_n^\top \\ \dots & \dots & \dots & \dots \\ b_n b_1^\top & b_n b_2^\top & \dots & b_n b_n^\top \end{pmatrix}$$

In this case, we we have $b_i b_j^\top = 1$ when $i = j$ and $b_i b_j^\top = 0$ when $i \neq j$, so b_is are orthonormal vectors.

Step 3: $(\mathbf{3}) \Rightarrow (\mathbf{1})$:

As in the step 1 and step 2, Set $A = (a_1, a_2, \cdots, a_n) = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$, where the column of A: $a_k \in \mathbb{R}^{n\times 1}$, the row of A: $b_k \in \mathbb{R}^{1\times n}, k = 1, 2, \cdots, n$.
Since the row vector b_i are orthonormal vectors, we have

$$AA^\top = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} (b_1^\top, b_2^\top, \cdots, b_n^\top) = \begin{pmatrix} b_1 b_1^\top & b_1 b_2^\top & \dots & b_1 b_n^\top \\ b_2 b_1^\top & b_2 b_2^\top & \dots & b_2 b_n^\top \\ \dots & \dots & \dots & \dots \\ b_n b_1^\top & b_n b_2^\top & \dots & b_n b_n^\top \end{pmatrix} = I_n$$

So we have $A^\top = A^{-1}$, and $A^\top A = I_n$.

$$I_n = A^\top A = \begin{pmatrix} a_1^\top \\ a_2^\top \\ \vdots \\ a_n^\top \end{pmatrix} (a_1, a_2, \cdots, a_n) = \begin{pmatrix} a_1^\top a_1 & a_1^\top a_2 & \dots & a_1^\top a_n \\ a_2^\top a_1 & a_2^\top a_2 & \dots & a_2^\top a_n \\ \dots & \dots & \dots & \dots \\ a_n^\top a_1 & a_n^\top a_2 & \dots & a_n^\top a_n \end{pmatrix}$$

Therefore, $a_i^\top a_j = 1$ when $i = j$ and $a_i^\top a_j = 0$ when $i \neq j$, so the column of A: a_is are orthonormal vectors.
In conclusion, since $(\mathbf{1}) \Rightarrow (\mathbf{2})$, $(\mathbf{2}) \Rightarrow (\mathbf{3})$ and $(\mathbf{3}) \Rightarrow (\mathbf{1})$, the three expressions are equivalent. $\square$

2 Problem

Show how do you go from Equation (2) to Equation (3) in the variance calculation in Section (2.1) of the randomized matrix multiplication notes.

Proof. Since $M_{i,j} = \sum\limits_{l=1}^{r}\sum\limits_{k=1}^{n} \frac{1}{rp_k} A_{i,k}B_{k,j} 1\{i_l = k\}$ and $Var(1\{i_l = k\}) = p_k(1 - p_k) = p_k - p_k^2$
For $\frac{1}{rp_k} A_{i,k}B_{k,j} 1\{i_l = k\}$ are independent for different k, We have:

$$\begin{aligned} Var(M_{i,j}) &= \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2 p_k^2} A_{i,k}^2 B_{k,j}^2 Var(1\{i_l = k\}) \\ &= \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2 p_k^2} A_{i,k}^2 B_{k,j}^2 (p_k - p_k^2) \\ &= \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2 p_k^2} A_{i,k}^2 B_{k,j}^2 (p_k) - \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2 p_k^2} A_{i,k}^2 B_{k,j}^2 (p_k^2) \\ &= \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2 p_k} A_{i,k}^2 B_{k,j}^2 - \sum_{l=1}^{r}\sum_{k=1}^{n} \frac{1}{r^2} A_{i,k}^2 B_{k,j}^2 \\ &= \frac{1}{r}\sum_{k=1}^{n} \frac{1}{p_k} A_{i,k}^2 B_{k,j}^2 - \frac{1}{r}\sum_{k=1}^{n} A_{i,k}^2 B_{k,j}^2 \\ &= \frac{1}{r}\sum_{k=1}^{n} \frac{1}{p_k} A_{i,k}^2 B_{k,j}^2 - \frac{1}{r}(A_{i,:}B_{:j})^2 \end{aligned}$$

Therefore, we can prove that: $\sum\limits_{i,j} \text{Var}(M_{i,j}) = \frac{1}{r}\sum\limits_{k}\sum\limits_{i,j} \frac{A_{i,k}^2 B_{k,j}^2}{p_k} - \frac{1}{r}\sum\limits_{i,j}(A_{i,:}B_{:,j})^2$. $\square$

3 Randomized matrix multiplication

1. Implement the algorithm presented in class for randomized matrix multiplication (Algorithm 2 from the class notes on Randomized Linear Algebra)

2. Apply the algorithm to the provided matrices selecting a number of columns $r = 20, 50, 100, 200$. The matrices for this problem can be found in the attached files "STA243_homework_1_matrix_A.csv" and "STA243_homework_1_matrix_B.csv".

3. Calculate the relative approximation error $||M - AB||_F/(||A||_F||B||_F)$ for each of the estimates found in (b). Provide your results in a table.

4. Visualize the estimates from (b) using the image() function in R. Combine the plots for $r = 20, 50, 100, 200$ into a single image using par(mfrow=c(2,2)).

Solution. *For (1) and (2), the R codes are shown in the R attachment. The table of the relative approximation error for each of the estimates is shown as follows:*

r	20	50	100	200
Error	0.2315	0.1284	0.0856	0.0596

To visualize the estimates, heat-maps are used to show the result.

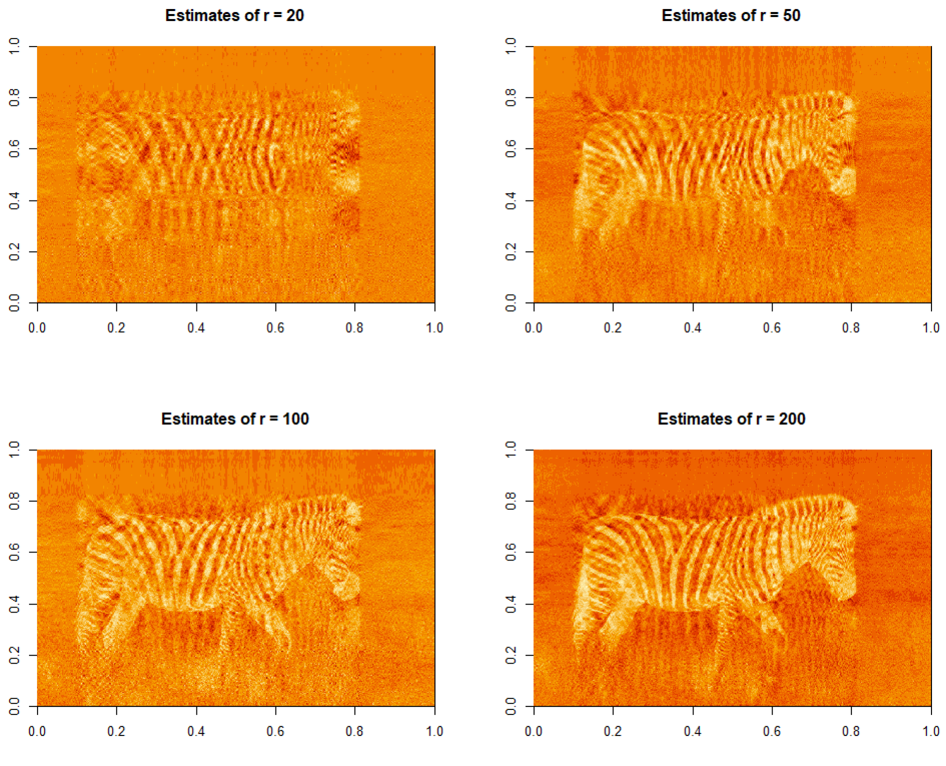

Figure 1: Visualize the estimates

4 Power method

Let $\mathbf{X}$ be a 10×10 matrix such that

$$\mathbf{X} = \lambda(\boldsymbol{v}\boldsymbol{v}^\top) + \mathbf{E},$$

where $\mathbf{E}$ is a *random matrix* with each entry being an i.i.d. standard Gaussian variable. Note that $\mathbf{X}$ is a rank-1 matrix perturbed with a random noise matrix ($\mathbf{E}$). Your goal in this problem is to *estimate* the eigenvector $\boldsymbol{v}$. The file `power_sim.R` consists of a test routine that fixes the true eigenvector, initial vector used for power method and the noise matrix. The end goal is produce a plot of λ versus how well the estimated eigenvector (using the power method) is correlated (measured via inner-product) with the true eigenvector. For this you have to implement your own power method (function `power_iteration`). Complete the code and run the test routine to produce the plot.

Solution. *The R codes for Power Method are shown in the R attachment. The result is shown in the following plot:*

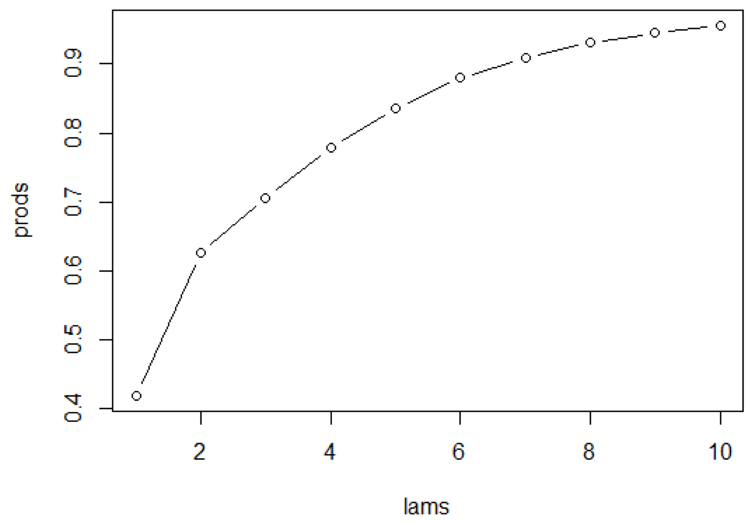

Figure 2: Test Result

As we can see from the plot, the larger the λ is, the better the performance of estimators, the closer to the true eign vectors.

5 Sketching for Least-squares

1. Implement the Sketched-OLS algorithm presented in class (Algorithm 1 from the class notes on Randomized Algorithms for Least Squares).

2. Generate a 1048576×20 design matrix $\mathbf{X}$ and a 1048576×1 response $\mathbf{y}$ with elements drawn iid from a Uniform$(0, 1)$ distribution.

3. Compare the calculation time for the full least squares problem and the sketched OLS. For the matrix $\mathbf{\Phi} = \mathbf{S}^T\mathbf{HD}$, first calculate $\mathbf{X}_* = \mathbf{\Phi X}$ and $\mathbf{y}_* = \mathbf{\Phi y}$. Once finished, use the system.time() function in R to time the calculation of $(\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{y}_*$ and compare to the calculation time of $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. Repeat these steps for $\epsilon = .1, .05, .01, .001$ and present your results in a table.

Solution. *The R codes are shown in the attachment. We input X, Y and error and then output the estimators β and matrix $\mathbf{X}_* = \mathbf{\Phi X}$ and $\mathbf{y}_* = \mathbf{\Phi y}$, where $\mathbf{\Phi} = \mathbf{S}^T\mathbf{HD}$. The results of time for each ϵ are shown in the table, we calculate the time for user, system and elasped. As we can see from the following table, the time for sketched OLS is obviously less than the classical OLS.*

ϵ	*OLS*	0.1	0.05	0.01	0.001
Time.user	*0.73*	*0*	*0*	*0*	*0.13*
Time.system	*0.83*	*0*	*0*	*0*	*0*
Time.elapsed	*1.73*	*0*	*0*	*0*	*0.12*

Only using the function $system.time()$ *in R is very hard to see the differences between the different error because the speed of computing is fast and the consuming of time is less than the precision which this function can be recorded. For this reason, we have used another function* $microbenchmark()$ *which can record the time more precisely than* $system.time()$*. (Note : The time unit of this function is milliseconds.)*

ϵ	*OLS*	0.1	0.05	0.01	0.001
Min	*569.95*	*1.05*	*2.09*	*12.09*	*127.68*
Mean	*612.27*	*1.27*	*2.35*	*13.15*	*139.91*
Medium	*592.65*	*1.17*	*2.28*	*12.50*	*135.27*
Max	*724.03*	*2.06*	*3.34*	*36.10*	*167.60*

Pledge:

Please sign below (print full name) after checking (✓) the following. If you can not honestly check each of these responses, please email me at kbala@ucdavis.edu to explain your situation.

- ✓ We pledge that we are honest students with academic integrity and we have not cheated on this homework.
- ✓ These answers are our own work.
- ✓ We did not give any other students assistance on this homework.
- ✓ We understand that to submit work that is not our own and pretend that it is our is a violation of the UC Davis code of conduct and will be reported to Student Judicial Affairs.
- ✓ We understand that suspected misconduct on this homework will be reported to the Office of Student Support and Judicial Affairs and, if established, will result in disciplinary sanctions up through Dismissal from the University and a grade penalty up to a grade of "F" for the course.

Team Member 1 Zhikuan Quan

Zhikuan Quan

Team Member 2 Bohao Zou

Bohao Zou