Problem 1.
[10 points] 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.

Solution:
First show **1** implies **2**.
Let the columns of $\mathbf{A}$ be denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. The assumption that the columns are orthonormal means:

$$\mathbf{a}_i^\top \mathbf{a}_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \delta_{ij}$$

δ_{ij} denotes the Kronecker delta.
Now consider the product $\mathbf{A}^\top\mathbf{A}$. The (i,j)-th entry of this matrix is:

$$(\mathbf{A}^\top\mathbf{A})_{ij} = \mathbf{a}_i^\top \mathbf{a}_j = \delta_{ij}$$

This is a matrix composed of Kronecker deltas which is just the identity.
Hence, we have:

$$\mathbf{A}^\top\mathbf{A} = \mathbf{I}$$

Since $\mathbf{A}$ is square and has full rank (its columns are linearly independent), it is invertible. From the identity $\mathbf{A}^\top\mathbf{A} = \mathbf{I}$, it follows that:

$$\mathbf{A}^{-1} = \mathbf{A}^\top$$

Multiplying both sides of this identity by $\mathbf{A}$ gives:

$$\mathbf{A}\mathbf{A}^\top = \mathbf{I}$$

Thus, both

$$\mathbf{A}^\top\mathbf{A} = \mathbf{I} \quad \text{and} \quad \mathbf{A}\mathbf{A}^\top = \mathbf{I}$$

.

Now, we will show **2** implies **1**.
Assume

$$\mathbf{A}\mathbf{A}^\top \;=\; \mathbf{A}^\top\mathbf{A} \;=\; \mathbf{I}.$$

Let the columns of $\mathbf{A}$ be denoted by $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{R}^n$. Then the (i,j)-entry of $\mathbf{A}^\top\mathbf{A}$ is:

$$(\mathbf{A}^\top\mathbf{A})_{ij} = \mathbf{a}_i^\top \mathbf{a}_j.$$

Since $\mathbf{A}^\top\mathbf{A} = \mathbf{I}$, it follows that

$$\mathbf{a}_i^\top \mathbf{a}_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

which shows that the columns of $\mathbf{A}$ are orthonormal.

Now, we will show **2** implies **3**.
Let the rows of $\mathbf{A}$ be denoted as $\mathbf{r}_1^\top, \ldots, \mathbf{r}_n^\top$, where each $\mathbf{r}_i \in \mathbb{R}^n$.
The (i,j)-th entry of the product $\mathbf{A}\mathbf{A}^\top$ is given by the inner product:

$$(\mathbf{A}\mathbf{A}^\top)_{ij} = \mathbf{r}_i^\top \mathbf{r}_j$$

But from the assumption that $\mathbf{A}\mathbf{A}^\top = \mathbf{I}$, we know:

$$(\mathbf{A}\mathbf{A}^\top)_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} = \delta_{ij}$$

This means:

$$\mathbf{r}_i^\top \mathbf{r}_j = \delta_{ij}$$

That is, the rows of $\mathbf{A}$ are orthonormal: each has unit length and distinct rows are orthogonal.

Now, to show that they are all equal, we will show that **3** implies **2**.
Let the rows of $\mathbf{A}$ be denoted as $\mathbf{r}_1^\top, \ldots, \mathbf{r}_n^\top$, where each $\mathbf{r}_i \in \mathbb{R}^n$. Orthonormality means

$$\mathbf{r}_i^\top \mathbf{r}_j = \delta_{ij}$$

Since $\left(\mathbf{A}\mathbf{A}^\top\right)_{ij} = \mathbf{r}_i^\top \mathbf{r}_j$, we get

$$\mathbf{A}\mathbf{A}^\top = I.$$

Because $AA^\top = I$ is invertible, $A^{-1} = A^\top$. Multiplying $AA^\top = I$ by $A^\top$ yields

$$A^\top A A^\top = A^\top \implies A^\top A = I.$$

Hence

$$AA^\top = I_n \quad \text{and} \quad A^\top A = I.$$

Problem 2.
[10 points] Show how do you go from Equation (2) to Equation (3) in the variance calculation in slide 20 of the RandLA.pdf slides.
Solution:

$$\begin{aligned}
\mathbb{E}\left[\|M - AB\|_F^2\right] &= \mathbb{E}\left[\sum_{i,j} (M_{ij} - A_{i,:}B_{:,j})^2\right] \\
&= \sum_{i,j} \mathbb{E}\left[(M_{ij} - A_{i,:}B_{:,j})^2\right] \\
&= \sum_{i,j} \mathbb{E}\left[(M_{ij} - \mathbb{E}[M_{ij}])^2\right] \\
&= \sum_{i,j} \operatorname{Var}(M_{ij})
\end{aligned}$$

To compute $\mathbb{E}[M_{ij}]$, we use:

$$\begin{aligned}
M_{ij} &= \sum_{l=1}^{r} \sum_{k=1}^{n} \frac{1}{rp_k} A_{ik} B_{kj} \cdot 1_{\{i_l = k\}} \\
\mathbb{E}[M_{ij}] &= \sum_{l=1}^{r} \sum_{k=1}^{n} \frac{1}{rp_k} A_{ik} B_{kj} \cdot \mathbb{E}[1_{\{i_l = k\}}] \\
&= \sum_{l=1}^{r} \sum_{k=1}^{n} \frac{1}{rp_k} A_{ik} B_{kj} \cdot p_k \\
&= \sum_{l=1}^{r} \sum_{k=1}^{n} \frac{1}{r} A_{ik} B_{kj} \\
&= \sum_{k=1}^{n} A_{ik} B_{kj} = A_{i,:}B_{:,j}
\end{aligned}$$

We know from discussion (and above) that

$$\mathbb{E}\left[\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}}\right]=\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}p_k=A_{i:}B_{:j},$$

and

$$\begin{aligned}\mathbb{E}\left[(\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}})^2\right]&=\mathbb{E}[\sum_{k=1}^{n}\frac{1}{p_k^2}A_{ik}^2B_{kj}^21_{\{i_1=k\}}^2+\sum_{k\neq t}\frac{1}{p_kp_t}A_{ik}B_{kj}A_{it}B_{tj}1_{\{i_1=k\}}1_{\{i_1=t\}}]\\&=\mathbb{E}[\sum_{k=1}^{n}\frac{1}{p_k^2}A_{ik}^2B_{kj}^21_{\{i_1=k\}}^2]=\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}^2B_{kj}^2\end{aligned}$$

We know from the definition of M_{ij},

$$\sum_{i,j}Var(M_{ij})=\sum_{i,j}\frac{1}{r}Var(\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}})$$

And by applying the definition of variance, $\text{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2$, to our results above ;

$$\begin{aligned}\sum_{i,j}\frac{1}{r}Var(\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}})&=\frac{1}{r}\sum_{i,j}[\mathbb{E}\left[(\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}})^2\right]-(\mathbb{E}\left[\sum_{k=1}^{n}\frac{1}{p_k}A_{ik}B_{kj}1_{\{i_1=k\}}\right])^2]\\&=\frac{1}{r}\sum_{k=1}^{n}\sum_{i,j}\frac{1}{p_k}A_{ik}^2B_{kj}^2-\frac{1}{r}\sum_{i,j}(A_{i,:}B_{:,j})^2\end{aligned}$$

which shows how to go from Equation (2) to Equation (3).
Problem 3.

r	Relative Error
20	0.203035
50	0.155827
100	0.094211
200	0.074369

Table 1: Relative error of the solution as a function of r.

We see that the matrix products form a zebra and that increasing the number of columns increases the resolution of the zebra.

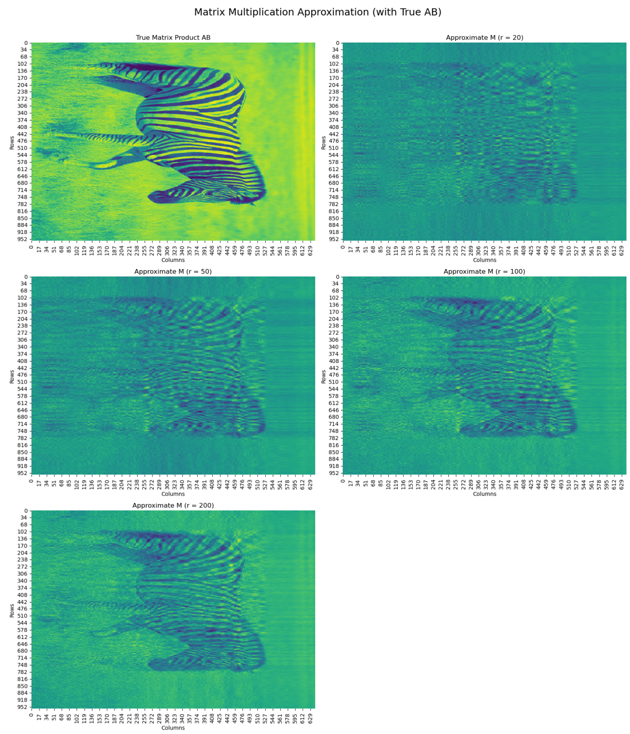

Figure 1: Matrix Multiplication Approximation (with True AB)

Problem 4. As λ increases the rank-1 signal overtakes the Gaussian noise, so the power-method estimate's correlation with the true eigenvector rises from its random baseline up toward 1.

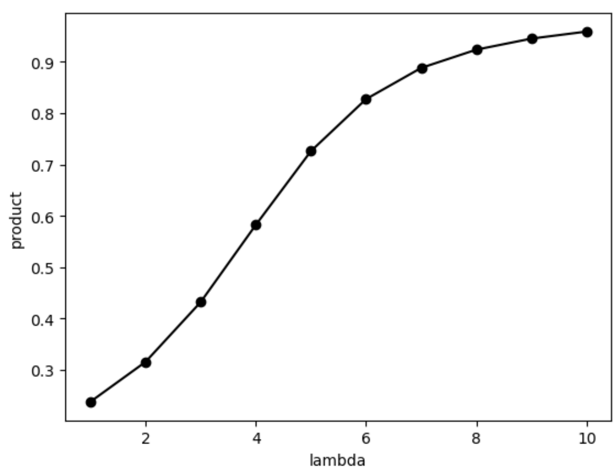

Figure 2: λ vs correlation

Problem 5.
It seems the sketched solution is signifantly faster.

Method	Running Time (s)
$\epsilon = 0.1$	0.002573
$\epsilon = 0.05$	0.000375
$\epsilon = 0.01$	0.007836
$\epsilon = 0.001$	0.029770
OLS	0.102390

Table 2: Running time comparison of sketch-based methods and ordinary least squares (OLS) for various ϵ values.

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.

Name: Shawheen Naderi