Problem 2.

(a) $\Rightarrow$ (b)
Given $\mathbf{A} \in \mathbb{R}^{n\times n}$, in general, the column vectors of a matrix are orthonormal when all the column vectors are pairwise orthogonal and the inner product of each column with itself is 1 (e.g. have length 1). That is, for column vectors of $\mathbf{A}$ are $a_1, a_2, \cdots, a_n$:

$$a_i^T a_j = \begin{cases} 0, & \text{if } i \neq j \\ 1, & \text{if } i = j \end{cases} \text{ and } \|a_i\| = 1 \text{ for } i = 1, \cdots, n \tag{1}$$

Thus, for a matrix $\mathbf{A}$ whose columns are orthonormal, $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}_n$ where $\mathbf{I}_n$ is the identity matrix. This property is demonstrated below.

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \vdots \\ \vec{a}_n^T \end{bmatrix} \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_n \end{bmatrix} = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \vec{a}_1^T\vec{a}_2 & \dots & \vec{a}_1^T\vec{a}_n \\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \dots & \vec{a}_2^T\vec{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_n^T\vec{a}_1 & \vec{a}_n^T\vec{a}_2 & \dots & \vec{a}_n^T\vec{a}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbf{I}_n \tag{2}$$

Additionally, since $\mathbf{A}$ is an $n \times n$ matrix, then by uniqueness of inverse matrices, $\mathbf{A}\mathbf{A}^T = \mathbf{I}_n$

$$\Longrightarrow \mathbf{A}^T\mathbf{A} = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n \tag{3}$$

(b) $\Rightarrow$ (c)
In addition, the rows of $\mathbf{A}$ are also orthonormal vectors since $\mathbf{A}\mathbf{A}^T = \mathbf{I}_n$. That is, let $\mathbf{B} = \mathbf{A}^T$, as a result the columns of $\mathbf{B}$ are rows of $\mathbf{A}$. It's evident that $\mathbf{B}$ is also an orthogonal matrix because

$$\mathbf{B}^T\mathbf{B} = (\mathbf{A}^T)^T\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \mathbf{I}_n \tag{4}$$

Since $\mathbf{B}^T\mathbf{B} = \mathbf{I}_n$ meaning $\mathbf{B}$ is an orthogonal matrix thus the columns of $\mathbf{B}$ are orthonormal vectors. Thus because the columns of $\mathbf{B}$ are rows of $\mathbf{A}$ we can conclude that the rows of $\mathbf{A}$ are orthonormal vectors.

(c) $\Rightarrow$ (b)
The above line of logic, can be reversed to go from bullet (c) to (b) and then (a). That is, if the rows of $\mathbf{A}$ are orthonormal vectors then all rows are pairwise orthogonal and the inner product of each row with itself is 1. Note that the statement in equation (1) holds but a_i are now row vectors. As a result, equations (2) and (3) also hold. Thus, we can go from bullet (c) to (b).

(b) $\Rightarrow$ (a)
In addition, to go from (b) to (a), the rows of $\mathbf{A}$ are the columns of $\mathbf{A}^T$. Therefore, let $\mathbf{B} = \mathbf{A}^T$. From equation (4), this implies $\mathbf{B}$ is orthogonal which means its columns are orthonormal vectors. Note that the columns of $\mathbf{B}$ are rows of $\mathbf{A}$.

Problem 3.
In order to calculate $\sum_{i,j} Var[\mathbf{M}_{i,j}]$ we used a four step approach. First, M can be decomposed into as sum of outer products $\mathbf{X}_\ell$. Next, we utilized the famous variance calculation trick. That is, for a variable X, $Var(X) = \mathbb{E}(X^2) + [\mathbb{E}(X)]^2$. Next, in the calculation of the expectations $\mathbf{X}_\ell$ and $\mathbf{X}_\ell^2$ we used two separate identities of indicator variables. That is, the expectation for the indicator variable $\mathbb{1}_A$ is $\mathbb{P}(A)$. In addition, the nth power of $\mathbb{1}_A$ equals to $\mathbb{1}_A$. As the last step, all of the pieces of the calculation were combined and simplified.

$$\begin{aligned} \mathbf{M} &= \sum_{\ell=1}^{r} \mathbf{X}_\ell \\ &= \sum_{\ell=1}^{r}\sum_{k=1}^{n} \frac{1}{rp_k} \mathbf{A}_{:,k}\mathbf{B}_{k,:}\mathbb{1}\{i_\ell = k\} \end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_\ell] &= \mathbb{E}\Big[\sum_{k=1}^{n}\frac{1}{rp_k}\mathbf{A}_{:,k}\mathbf{B}_{k,:}\mathbb{1}\{i_\ell = k\}\Big] \\ &= \sum_{k=1}^{n}\frac{p_k}{rp_k}\mathbf{A}_{:,k}\mathbf{B}_{k,:} \\ &= \sum_{k=1}^{n}\frac{1}{r}\mathbf{A}_{:,k}\mathbf{B}_{k,:} \\ &= \frac{1}{r}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_\ell^2] &= \mathbb{E}\Big[\Big(\sum_{k=1}^{n}\frac{1}{rp_k}\mathbf{A}_{:,k}\mathbf{B}_{k,:}\mathbb{1}\{i_\ell = k\}\Big)^2\Big] \\ &= \sum_{k=1}^{n}\frac{1}{r^2p_k^2}\mathbf{A}_{:,k}^2\mathbf{B}_{k,:}^2\mathbb{E}\big[(\mathbb{1}\{i_\ell = k\})^2\big] \\ &= \sum_{k=1}^{n}\frac{p_k}{r^2p_k^2}\mathbf{A}_{:,k}^2\mathbf{B}_{k,:}^2 \\ &= \sum_{k=1}^{n}\frac{1}{r^2p_k}\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2\end{aligned}$$

$$\begin{aligned}\sum_{i,j}Var[\mathbf{M}_{i,j}] &= \sum_{i,j}\sum_{\ell=1}^{r}Var[X_\ell] \\ &= \sum_{i,j}\sum_{\ell=1}^{r}\left[\mathbb{E}[X_\ell^2] - (\mathbb{E}[X_\ell])^2\right] \\ &= \sum_{i,j}\sum_{\ell=1}^{r}\left[\sum_{k=1}^{n}\frac{1}{r^2p_k}\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2 - \Big(\frac{1}{r}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})\Big)^2\right] \\ &= \sum_{i,j}\sum_{\ell=1}^{r}\left[\sum_{k=1}^{n}\frac{1}{r^2p_k}\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2 - \Big(\frac{1}{r}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})\Big)^2\right] \\ &= \sum_{i,j}\sum_{\ell=1}^{r}\left[\sum_{k=1}^{n}\frac{1}{r^2p_k}\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2 - \frac{1}{r^2}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})^2\right] \\ &= \sum_{i,j}\left[\sum_{k=1}^{n}\frac{1}{rp_k}\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2 - \frac{1}{r}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})^2\right] \\ &= \frac{1}{r}\sum_{k=1}^{n}\sum_{i,j}\frac{\mathbf{A}_{i,k}^2\mathbf{B}_{k,j}^2}{p_k} - \frac{1}{r}\sum_{i,j}(\mathbf{A}_{i,:}\mathbf{B}_{:,j})^2\end{aligned}$$

Problem 4.

(a) Consider two matrices $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$. To implement the randomized matrix multiplication algorithm, we began by defining n and an empty matrix $\mathbf{M} \in \mathbb{R}^{m\times p}$. Next, we calculated p_k by doing non-uniform random sampling, which is defined as

$$p_k = \frac{\|\mathbf{A}_{:,k}\|_2\|\mathbf{B}_{k,:}\|_2}{\sum_l \|\mathbf{A}_{:,l}\|_2\|\mathbf{B}_{l,:}\|_2}$$

After calculating the probabilities, we generated an array of random numbers of size r from 1 to n where each element had probability p_k of being selected. Lastly, to construct $\mathbf{M}$, we iterated through the array of random numbers and for each number in the array, the estimated matrix $\mathbf{M}$ was calculated by

$$\mathbf{M} = \sum_{i=1}^{r}\frac{1}{rp_i}\mathbf{A}_{:,i}\mathbf{B}_{i,:}$$

(b) The algorithm above was implemented and applied to the two provided matrices. The estimated matrices and the Relative Approximation Error ("RAE") were recorded for r values 20, 50, 100, and 200.

(c) Summary table of relative approximation errors and number of columns r.

Number of Columns r	Relative Approximation Error
20	0.2238
50	0.1623
100	0.1045
200	0.0819

Table 1: Number of Columns r and Relative Approximation Error.

(d) Refer to figure 1 for the plots of the matrix estimates from (b).

Problem 5.
The algorithm to implement the Power Method is split into two parts. In the first part, the variables and initial values are initialized (e.g. create a variable t to track the number of iterations). In the second stage, the steps of the algorithm are implemented to run until one of the termination conditions is reached. There are two conditions: 1) the maximum number of iterations are reached or 2) the ℓ_2 norm of the difference between estimated eigenvectors from the current and past iterations is less than some user-defined ϵ. Once one of the conditions is reached, the eigenvector from the current iteration is returned. The steps of the algorithm are as follows:

1. The matrix A is multiplied on the right with the estimated eigenvector from the previous iteration. In this case, matrix A is the matrix whose top eigenvector is estimated.
2. The resulting vector is normalized (e.g. divided by it's ℓ_2 norm).
3. The difference between estimated eigenvectors from the current and the past iterations are computed (e.g. ℓ_2 norm of the difference).
4. The iterator t is incremented by one.

The plot of how well the estimated eigenvector is correlated (measured via inner-product) with the true eigenvector versus λ can be found in figure 2. For this plot, the top eigenvector for matrix $\mathbf{X} = \lambda(vv^T) + \mathbf{E}$ is estimated where $\mathbf{E}$ is defined to be a random noise matrix.

Problem 6.
The Sketched-OLS algorithm implementation includes three essential steps. First, the algorithm is initialized. That is the summary information about the inputted data is collected and stored (e.g. dimensions of the design matrix). In the next step, two random samples are drawn. First, n numbers are sampled with replacement from the set {-1, 1} with equal probabilities. Second, a sample of r indices from 1 to the total number of rows in the design matrix is generated. These random samples are then passed to a subroutine that calculates the transformation matrix Φ on $\mathbf{X}$ and $\mathbf{Y}$. In order to speed up calculation speed at this stage, instead of using matrix multiplication, each transformation that is part of Φ is iteratively applied. For example, the signs on rows are flipped according to the sample of {-1, 1} described above. In addition, instead of multiplying by the $\mathbf{S}^T$ matrix, the second sample of r indices is used to subset the rows of the target matrix. Note that the effects of these two operations are the same as when doing matrix multiplication, however, it takes up much less computational resources. At the last stage, the Sketched-OLS $\hat{\beta}$ is estimated with $(\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{y}_*$ where $\mathbf{X}_* = \Phi\mathbf{X}$ and $\mathbf{Y}_* = \Phi\mathbf{Y}$.

Based on our implementation, Sketched OLS is capable of significantly shortening computation time. When applied on a 1048576 $\times$ 20 design matrix $\mathbf{X}$ and a 1048576 $\times$ 1 response vector $\mathbf{y}$ with elements drawn i.i.d from a Uniform(0, 1) distribution, we obtained the following calculation times for $(\mathbf{X}_*^T\mathbf{X}_*)^{-1}\mathbf{X}_*^T\mathbf{y}_*$ in seconds given ϵ values.

ϵ	**Calculation Time (sec)**
0.1	0.0003
0.05	0.0003
0.01	0.0013
0.001	0.0014

Table 2: Comparison between ϵ and the corresponding calculation time using Sketched OLS algorithm.

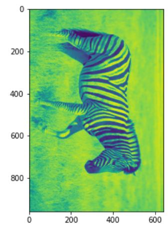
Note that compared to 0.0846 seconds of calculation time for the normal OLS estimator $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$, Sketched-OLS provides a 84% to 99% decrease in calculation time.

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.

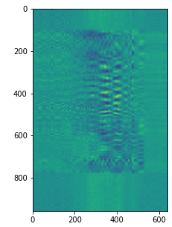
- We pledge that we are honest students with academic integrity and we have not cheated on this homework. ✓
- These answers are our own work. ✓
- We did not give any other students assistance on this homework. ✓
- We understand that to submit work that is not our own and pretend that it is our is a violation 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
Grant Gambetta

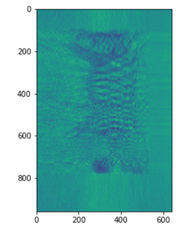
Team Member 2
Ignat Kulinka



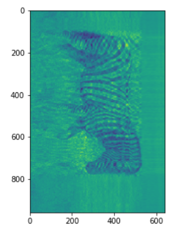
(a) Actual



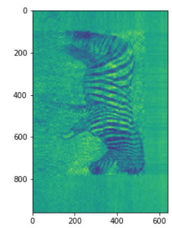
(b) $r = 20$



(c) $r = 50$



(d) $r = 100$



(e) $r = 200$

Figure 1: Plots for the actual and estimated matrices.

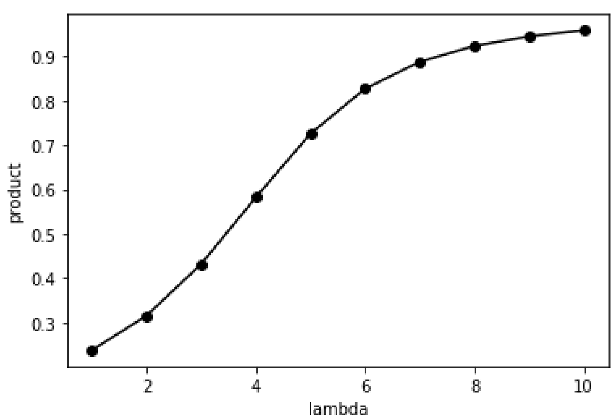


Figure 2: Plot of the inner product between the estimated eigenvector and true eigenvector vs λ.