

## Homework #2

Due: 2024-10-21 23:59 | 7 Problems, 100 Pts

Name: 徐靖, ID: 2200012917

**Problem 1 (20').** Consider a cube  $C$  with side length 1 in  $d$ -dimensional space.

- (1) (4') Write down the radius of a ball  $B$  whose volume is equal to the cube's. You don't need to prove your result.
- (2) (16') When the centers of the cube  $C$  and the ball  $B$  are both at the origin, calculate the volume of their intersection when  $d \rightarrow +\infty$ .

[Hint: Try to calculate  $\mathbb{E}_{X \sim C}[\|X\|_2^2]$  and  $\text{Var}_{X \sim C}[\|X\|_2^2]$ . Then use Chebyshev's Inequality to analyze the concentration of  $\|X\|_2^2$ . You may use the Stirling's approximation

$$\lim_{n \rightarrow +\infty} \frac{\Gamma(n+1)}{\sqrt{2\pi n}(n/e)^n} = 1.$$

/

◀

**Answer.**

- (1) Let  $r$  donate the radius.

$$r = \pi^{-\frac{1}{2}} \left( \frac{d\Gamma(\frac{d}{2})}{2} \right)^{\frac{1}{d}}$$

- (2) First we calculate  $\mathbb{E}_{X \sim C}[\|X\|_2^2]$  and  $\text{Var}_{X \sim C}[\|X\|_2^2]$  step by step. Let  $X_i$  denote the components of  $X$ , where  $X \sim C$ . Given that  $X_i \stackrel{i.i.d.}{\sim} C, i \in [n]$ , we have:

$$\begin{aligned} \mathbb{E}_{X \sim C} [X_i^2] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} X_i^2 dX_i = \frac{1}{12} \\ \mathbb{E}_{X \sim C} [X_i^4] &= \int_{-\frac{1}{2}}^{\frac{1}{2}} X_i^4 dX_i = \frac{1}{80} \\ \mathbb{E}_{X \sim C} [\|X\|_2^2] &= \sum_{i=1}^d \mathbb{E}_{X \sim C} [X_i^2] = \frac{d}{12} \\ \mathbb{E}_{X \sim C} [\|X\|_2^4] &= \mathbb{E}_{X \sim C} \left[ \left( \sum_{i=1}^{d-1} X_i^2 \right)^2 \right] + 2\mathbb{E}_{X \sim C} [X_d^2] \cdot \mathbb{E}_{X \sim C} \left[ \sum_{i=1}^{d-1} X_i^2 \right] + \mathbb{E}_{X \sim C} [X_d^4] \\ &= \mathbb{E}_{X \sim C} \left[ \left( \sum_{i=1}^{d-1} X_i^2 \right)^2 \right] + \frac{d-1}{72} + \frac{1}{80} \\ &= \sum_{i=1}^d \frac{d-1}{72} + \frac{1}{80} = \frac{d^2}{144} + \frac{d}{180} \\ \text{Var}_{X \sim C} [\|X\|_2^2] &= \mathbb{E}_{X \sim C} [\|X\|_2^4] - \mathbb{E}_{X \sim C} [\|X\|_2^2]^2 = \frac{d}{180} \end{aligned}$$

Then we analyze the concentration of  $\|X\|_2^2$ . Given that,

$$\begin{aligned}
 \mathbb{P}_{X \sim C}[X \in B] &= \mathbb{P}_{X \sim C} [\|X\|_2^2 < r^2] \\
 &= \mathbb{P}_{X \sim C} \left[ \|X\|_2^2 - \mathbb{E} [\|X\|_2^2] < r^2 - \frac{d}{12} \right] \\
 &\leq \mathbb{P}_{X \sim C} \left[ \left| \|X\|_2^2 - \mathbb{E} [\|X\|_2^2] \right| > \left| \frac{d}{12} - r^2 \right| \right] \\
 &\leq \frac{\text{Var}_{X \sim C} [\|X\|_2^2]}{\left( r^2 - \frac{d}{12} \right)^2} \\
 &= \frac{\frac{d}{180}}{\left( \pi^{-1} \left( \frac{d\Gamma(\frac{d}{2})}{2} \right)^{\frac{2}{d}} - \frac{d}{12} \right)^2} \\
 &\rightarrow 0 \quad (d \rightarrow +\infty) \quad (\text{using Stirling's approximation})
 \end{aligned}$$

Thus we find that:

$$\lim_{d \rightarrow +\infty} \mathbb{P}_{X \sim C}[X \in B] = 0$$

Hence,

$$V(C \cap B) = 0$$

◁

**Problem 2 (16').** You select a PE class this term, so you need to complete a total of 85km of extracurricular exercise. There are only  $n$  days before the deadline, but you still have  $m$  kilometers remaining. To simplify your work, you can run at most 10km every day, but there is no lower bound. You are mindless when running, so the distance you run every day is uniformly random. In other words, the number of kilometers you run every day is a real number uniformly selected from range  $[0, 10]$ . Prove that, the probability that you can complete  $m$  kilometers in  $n$  days, i.e., the probability that your total distance in  $n$  days is greater than or equal to  $m$  kilometers is

$$1 - \sum_{i=0}^{\min(\lfloor m/10 \rfloor, n)} \frac{(-1)^i}{i!(n-i)!} \left( \frac{m}{10} - i \right)^n.$$

◀

**Answer.** Let  $\{x_i\}$  denote the numbers of kilometers divided by 10, we have,

$$\mathbb{P}_{x_i \sim U(0,1)} \left( 10 \sum_{i \in [n]} x_i \geq m \right) = 1 - \frac{V(x_i \sim U(0,1), \sum_{i \in [n]} x_i \leq m/10)}{V(x_i \sim U(0,1))}$$

Let  $U = \{\mathbf{x} \mid \|\mathbf{x}\|_1 \leq \frac{m}{10}, x_i \geq 0\}$ ,  $A_i = \{\mathbf{x} \mid \mathbf{x} \in U, x_i \leq 1\}$ . Naturally,  $\overline{A_i} = U \setminus A_i$ .

$$V(x_i \sim U(0,1), \sum_{i \in [n]} x_i \leq m/10) = V \left( \bigcap_{i \in [n]} A_i \right)$$

According to the inclusion-exclusion principle,

$$\begin{aligned}
 V\left(\bigcap_{i \in [n]} A_i\right) &= \sum_{k \in [n]} (-1)^{k+1} \sum_{1 \leq i_1 \leq \dots \leq i_k \leq n} V\left(\bigcap_{j \in [k]} A_{i_j}\right) \\
 &= \sum_{k \in [n]} (-1)^{k+1} \binom{n}{k} \left( V(U) - V\left(\bigcup_{j \in [k]} \overline{A_{i_j}}\right) \right) \\
 &= \sum_{k \in [n]} (-1)^k \binom{n}{k} V\left(\bigcup_{j \in [k]} \overline{A_{i_j}}\right) + V(U) \sum_{k \in [n]} (-1)^{k+1} \binom{n}{k}
 \end{aligned}$$

Note that

$$\sum_{k \in [n]} (-1)^{k+1} \binom{n}{k} = 0$$

and (The choice of  $i_j$  is arbitrary, because  $x_i$  is iid.)

$$V\left(\bigcup_{j \in [k]} \overline{A_{i_j}}\right) = \begin{cases} 0, & k \geq m/10 \\ \frac{1}{n!} \left(\frac{m}{10} - k\right)^n, & k < m/10 \end{cases}$$

Thus,

$$V\left(\bigcap_{i \in [n]} A_i\right) = \sum_{i \in [n], k < m/10} \frac{(-1)^i}{i!(n-i)!} \left(\frac{m}{10} - i\right)^n$$

which means,

$$\mathbb{P}_{x_i \sim U(0,1)}\left(10 \sum_{i \in [n]} x_i \geq m\right) = 1 - \sum_{i=0}^{\min(\lfloor m/10 \rfloor, n)} \frac{(-1)^i}{i!(n-i)!} \left(\frac{m}{10} - i\right)^n$$

◁

**Problem 3 (10').** If  $A$  is square, show that  $AA^\top$  and  $A^\top A$  are similar.

[Hint: Use the SVD of  $A$ .]

◀

**Answer.** Let  $A$  be an  $n \times n$  matrix. By the Singular Value Decomposition (SVD), we have:

$$A = UDV^\top$$

where  $U$  and  $V$  are orthogonal matrices, and  $D$  is a diagonal matrix containing the singular values of  $A$ . Then we have:

$$\begin{aligned}
 AA^\top &= (UDV^\top)(UDV^\top)^\top = UDV^\top V D^\top U^\top = U D^2 U^\top \\
 A^\top A &= (UDV^\top)^\top (UDV^\top) = (V D^\top U^\top)(UDV^\top) = V D^2 V^\top
 \end{aligned}$$

Let  $P = VU^{-1}$ , we have:

$$PAA^\top P^{-1} = VU^{-1}UD^2U^\top UV^{-1} = VD^2V^\top$$

Hence  $AA^\top$  and  $A^\top A$  are similar.

&lt;

**Problem 4 (12').** Calculate the SVD of matrix  $A$ , where

$$A = \begin{pmatrix} -4 & -6 \\ 3 & -8 \end{pmatrix}.$$

You need to calculate it in two different ways.

[Hint: Use the definition of singular vectors, or consider  $A^\top A$ .]

◀

**Answer.**

**Method 1**

Assume  $\mathbf{v} = (x, y)^\top$ , then we have  $\|A\mathbf{v}\| = \sqrt{25x^2 + 100y^2}$ . Next, we look for singular vectors one by one.

$$\mathbf{v}_1 = \arg \max_{|\mathbf{v}|=1} \|A\mathbf{v}\| = (0, 1)^\top, \sigma_1 = \sqrt{\|A\mathbf{v}_1\|^2} = 10, \mathbf{v}_2 = \arg \max_{|\mathbf{v}|=1, \mathbf{v} \perp \mathbf{v}_1} \|A\mathbf{v}\| = (1, 0)^\top, \sigma_2 = \sqrt{\|A\mathbf{v}_2\|^2} = 5$$

Then we have,

$$\mathbf{u}_1 = \frac{A\mathbf{v}_1}{\sigma_1} = (-0.6, -0.8)^\top, \mathbf{u}_2 = \frac{A\mathbf{v}_2}{\sigma_2} = (-0.8, 0.6)^\top$$

Hence,

$$A = \begin{pmatrix} -0.6 & -0.8 \\ -0.8 & 0.6 \end{pmatrix} \text{diag}\{10, 5\} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

**Method 2**

Consider

$$A^\top A = \begin{pmatrix} 25 & 0 \\ 0 & 100 \end{pmatrix}.$$

Solve the equation  $|A^\top A - \lambda I| = 0$ , we get  $\lambda = 25, 100$ , which means  $\sigma_1 = 10, \sigma_2 = 5$ . Then solve equations  $(A^\top A - \lambda_i I)\mathbf{v} = 0, |\mathbf{v}| = 1$  separately. We get the eigenvector  $\{\mathbf{v}_i\}$ , then left singular vectors  $\{\mathbf{u}_i\}$ , the solution is obviously same as Method 1.

&lt;

**Problem 5 (16').** Let  $A = \sum_{i=1}^r \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$  be the SVD of a rank- $r$  matrix  $A$ , let  $A_k := \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$  be the rank- $k$  approximation of  $A$  for some  $k < r$ .

- (1) (4') Express the following quantities in terms of the singular values  $\{\sigma_i, 1 \leq i \leq r\}$ . You don't need to prove your result.

(a) (2')  $\|A_k\|_F^2$ .

(b) (2')  $\|A_k\|_2^2$ .

(2) (3') Prove that,  $\sigma_k \leq \frac{\|A\|_F}{\sqrt{k}}$  for  $k = 1, 2, \dots, r$ .

(3) (9') Suppose  $A \in \mathbb{R}^{m \times m}$ . Let  $B = \sum_{i=1}^r \frac{1}{\sigma_i} v_i u_i^\top$ .

(a) (6') Show that  $BAx = x$  for all  $x$  in the span of the right singular vectors of  $A$ . For this reason  $B$  is sometimes called the pseudo inverse of  $A$  and can play the role of  $A^{-1}$  in many applications.

(b) (3') Does  $BAx = x$  always hold for all  $x \in \mathbb{R}^m$ ? Prove or give a counterexample.

◀

**Answer.**

(1) (a)  $\|A_k\|_F^2 = \sum_{i \in [k]} \sigma_i^2$

(b)  $\|A_k\|_2^2 = \sigma_1^2$

(2) This seems obvious because the singular values are decreasing. And we have shown that the square of the Frobenius norm is equal to the sum of the squares of the singular values.

$$\|A\|_F = \sqrt{\sum_{j \in [n]} |a_j|^2} = \sqrt{\sum_{j \in [n], i \in [r]} (a_j \cdot v_i)^2} = \sqrt{\sum_{i \in [r]} \sigma_i^2(A)} \geq \sqrt{k \sigma_k^2(A)}$$

(3) (a) Note that

$$\begin{aligned} BA &= \left( \sum_{k=1}^r \frac{1}{\sigma_k} v_k u_k^\top \right) \left( \sum_{k=1}^r \sigma_k u_k v_k^\top \right) \\ &= \sum_{k=1}^r v_k u_k^\top u_k v_k^\top + \sum_{k=1}^r \sum_{j \in [r] \setminus \{k\}} v_k u_k^\top u_j v_j^\top \\ &= \sum_{k=1}^r v_k v_k^\top \end{aligned}$$

and  $v_1, \dots, v_r$  is a basis of the span of the right-singular vectors of  $A$ . For any  $1 \leq i \leq r$ ,

$$BA v_i = \sum_{k=1}^r v_k v_k^\top v_i = v_i$$

thus for any  $x$  in the span of the right-singular vectors of  $A$ ,

$$BAx = x$$

(b) Not necessarily true. If  $A$  is not rank-full, then there exists  $x \neq 0$  such that  $Ax = 0$ , then  $BAx = 0 \neq x$ . Naturally, when  $A$  is full-rank, the set of orthonormal vectors  $\{v_i\}$  spans the space  $\mathbb{R}^m$ , so the conclusion holds.

◀

**Problem 6 (16').** In class, we introduced the best-fit subspace for a set of points. We extend this definition to probability densities instead of a set of points. If  $P$  is a probability density in  $\mathbb{R}^d$ , then we define the best-fit  $k$ -dimensional subspace ( $k \leq d$ )

$$V_k = \arg \max_{V, \dim(V)=k} \mathbb{E}_{X \sim P}(|\text{proj}(X, V)|^2),$$

where  $\text{proj}(X, V)$  is the orthogonal projection of  $X$  onto  $V$ .

- (1) (6') Find out the best-fit 1-dimensional subspace (i.e., a best-fit line through the origin) for probability density  $\mathcal{N}(0, 0, 1, 2, \frac{1}{2})$ .

[Hint: Suppose random variable  $(X_1, X_2) \sim \mathcal{N}(0, 0, 1, 2, \frac{1}{2})$ , then we have

$$X_1 \sim \mathcal{N}(0, 1), X_2 \sim \mathcal{N}(0, 2), \rho := \frac{\mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)}{\sqrt{\text{Var}(X_1)\text{Var}(X_2)}} = \frac{1}{2}. \quad ]$$

- (2) (10') Prove that, for  $d$ -dimensional Gaussian distribution  $\mathcal{N}(\mu, I_d)$ , a  $k$ -dimensional subspace is a best-fit subspace if and only if it contains  $\mu$ .

[Hint: Suppose random variable  $(X_1, \dots, X_d) \sim \mathcal{N}(\mu, I_d)$ , then we have  $X_i \sim \mathcal{N}(\mu_i, 1)$  and  $X_1, \dots, X_d$  are independent random variables.]

◀

**Answer.**

- (1) Assume unit vector  $\mathbf{v}_0 = (v_1, v_2)^\top$  spans the best-fit 1-dimensional subspace. Then

$$\begin{aligned} \mathbf{v}_0 &= \arg \max_{|\mathbf{v}|=1} \mathbb{E}_{X \sim \mathcal{N}(0, 0, 1, 2, \frac{1}{2})} (|X \cdot \mathbf{v}|^2) \\ &= \arg \max_{|\mathbf{v}|=1} \mathbb{E}((X_1 v_1 + X_2 v_2)^2) \\ &= \arg \max_{|\mathbf{v}|=1} v_1^2 + \sqrt{2} v_1 v_2 + 2 v_2^2 \\ &= \left( \sqrt{\frac{3 - \sqrt{3}}{6}}, \sqrt{\frac{3 + \sqrt{3}}{6}} \right) \end{aligned}$$

Hence the subspace is  $\{t\mathbf{v}_0 | t \in \mathbb{R}\}$ .

(2) Let  $\mathbf{v}_i$  be a set of standard orthogonal bases of  $V$ . Note that

$$\begin{aligned}
 \mathbb{E}_{X \sim \mathcal{N}(\mu_i, I_d)}(|\text{proj}(X, V)|^2) &= \mathbb{E}_{X \sim \mathcal{N}(\mu_i, I_d)} \left( \left| \sum_{i \in [k]} (X \cdot \mathbf{v}_i) \mathbf{v}_i \right|^2 \right) \\
 &= \mathbb{E}_{X \sim \mathcal{N}(\mu_i, I_d)} \left( \sum_{i \in [k]} \left( \sum_{j \in [d]} X_j v_{ij} \right)^2 \right) \\
 &= \mathbb{E}_{X \sim \mathcal{N}(\mu_i, I_d)} \left( \sum_{i \in [k]} \sum_{j \in [d]} X_j^2 v_{ij}^2 \right) \\
 &= \sum_{i \in [k]} \sum_{j \in [d]} v_{ij}^2 \mathbb{E}(X_j^2) \\
 &= \sum_{i \in [k]} \sum_{j \in [d]} v_{ij}^2 (\mu_i^2 + 1) \\
 &= k + |\text{proj}(\mu, V)|^2
 \end{aligned}$$

That leads to

$$V_k = \arg \max_{V, \dim(V)=k} |\text{proj}(\mu, V)|^2$$

which means  $V_k$  is the subspace that maximizes the projected modulus length of  $\mu$  on  $V_k$ . In other words, it contains  $\mu$  (If and only if  $\mu$  is within  $V_k$ ,  $\text{Proj}(\mu, V)$  reaches the maximum value  $|\mu|$ )

◁

**Problem 7 (10').** Read in a photochrome and convert to a matrix. Perform a singular value decomposition of the matrix. Reconstruct the photo using only 1,2,4,16 singular values.

- (1) Print the reconstructed photo. How good is the quality of the reconstructed photo? Write down your observations.
- (2) What percent of the Frobenius norm is captured in each case? Describe the way you calculate the captured Frobenius norm.

You need to include your original photochrome and all reconstructed photos in your answer and submit the code as an attachment.

[Hint: You may want to perform SVD on each channel.]

◀

**Answer.**

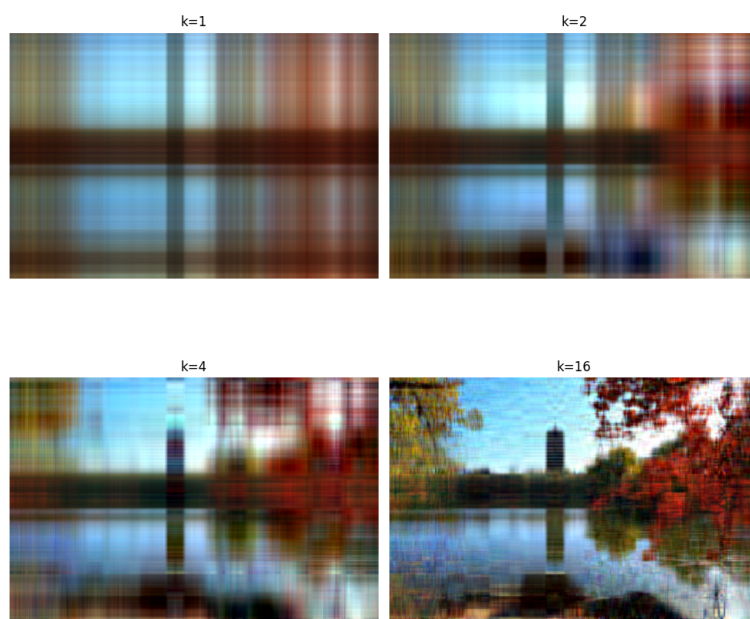


图 1: Rearranged images with different number of singular values  $k$

- (1) The image quality increases with the number of singular values. When  $k = 16$ , the image can roughly reflect the characteristics of the original image.



(2) My Frobenius norm calculation is implemented as follows (Python):

```
def frobenius_norm(matrix):
    return np.sqrt(np.sum(matrix ** 2))
```

That is, directly calculate the sum of squares of matrix elements.

表 1: Frobenius norm percentage

$k$	total percentage	percentage of channels
1	87.684%	Channel 0: 89.132% Channel 1: 87.982% Channel 2: 86.271%
2	90.882%	Channel 0: 91.354% Channel 1: 91.317% Channel 2: 90.119%
4	93.788%	Channel 0: 93.183% Channel 1: 94.147% Channel 2: 93.932%
16	96.285%	Channel 0: 95.514% Channel 1: 96.433% Channel 2: 96.747%

- My code combines the matrices of the three channels into a three-dimensional matrix and calculates the Frobenius norm together. It is actually equivalent to taking the geometric average of the norms of each channel and then calculating the percentage, which can **reflect the overall norm change**
- I once encountered an error of forgetting to do normalization when writing code. Later, I found and solved this bug through another implementation of Frobenius norm: **calculating the sum of squares of singular values**. Now the Frobenius norms calculated by the two methods are exactly the same, which shows the correctness of my SVD implementation

◁