Math5473 hw1

Li Jiayi student id: 20749286

Feb 2021

1 Problem1

1.1 (d)

Plot eigenvalue curve.

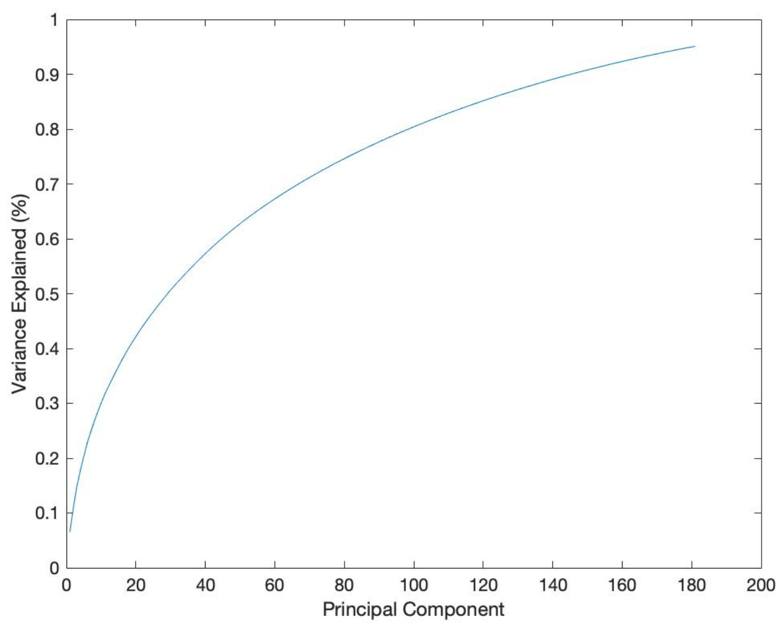


Figure 1: eigenvalue curves

1.2 (e)

Visualize the mean and top-k principle components. I just show the first 12 corresponding principle components pictures here.

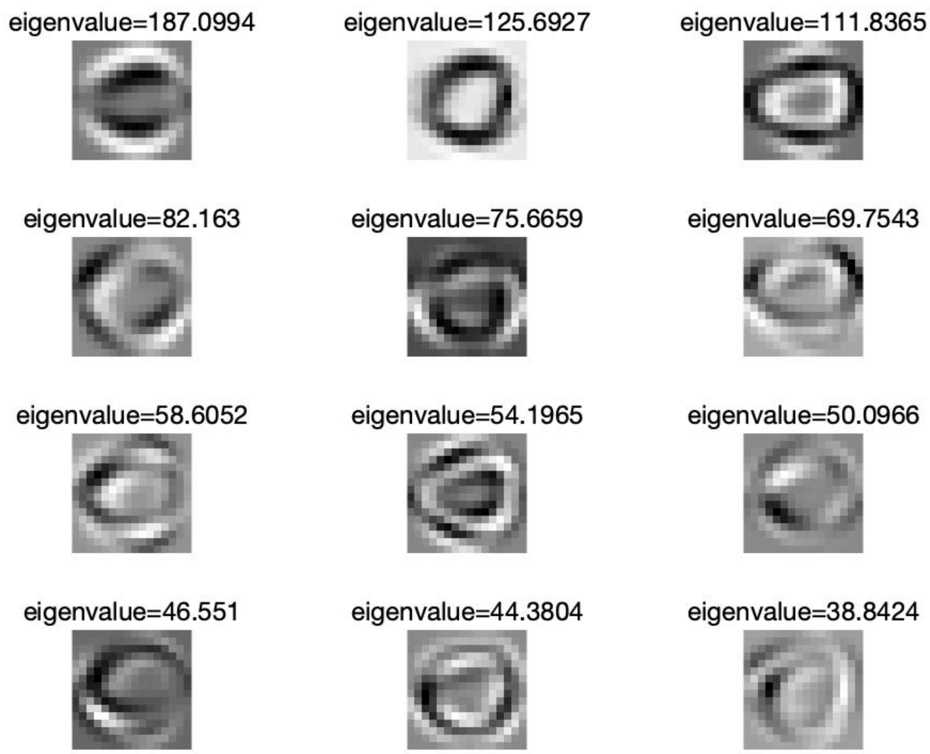


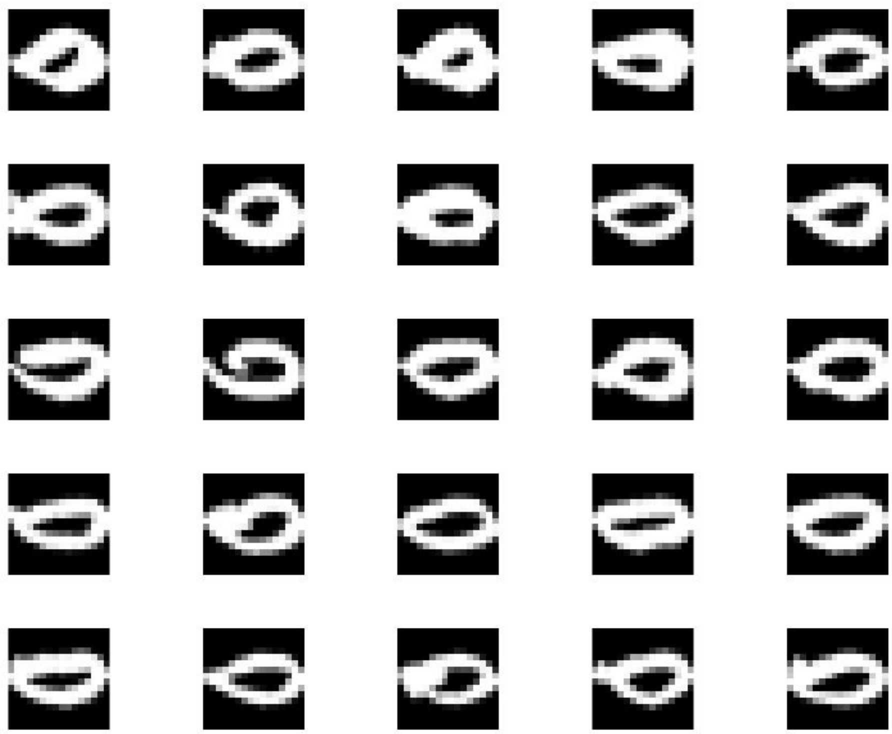
Figure 2: top k principle components

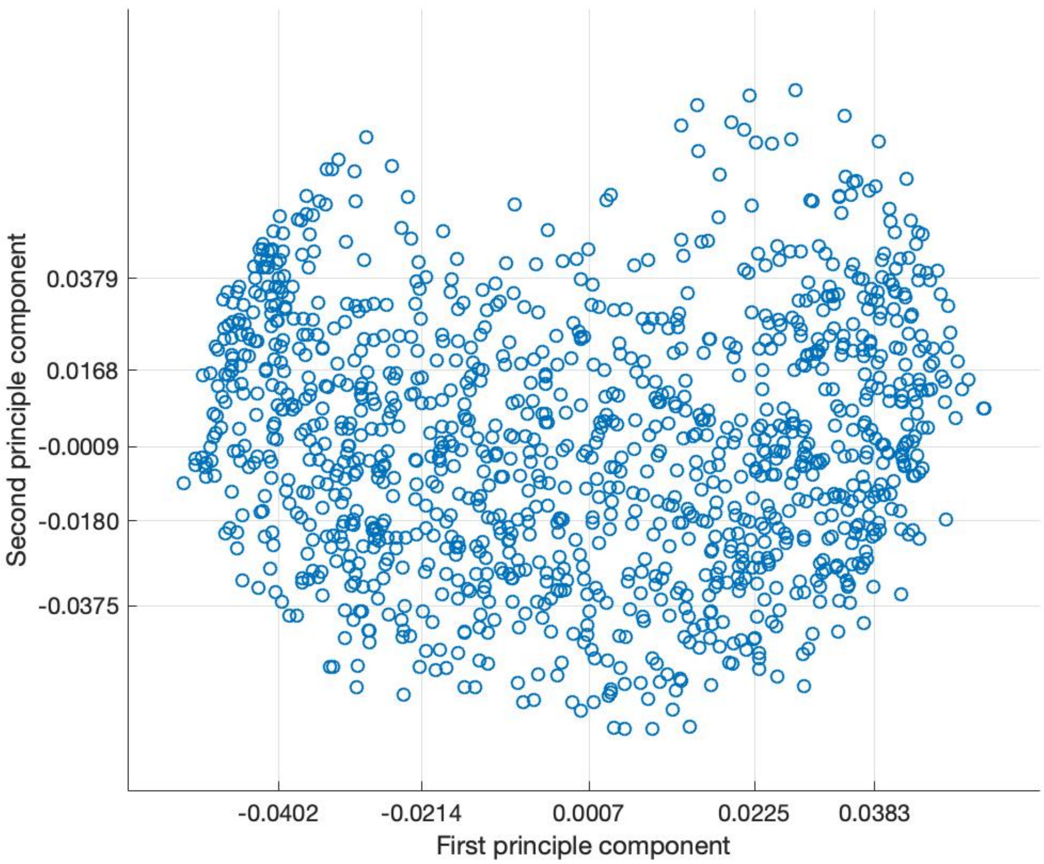
1.3 (f)

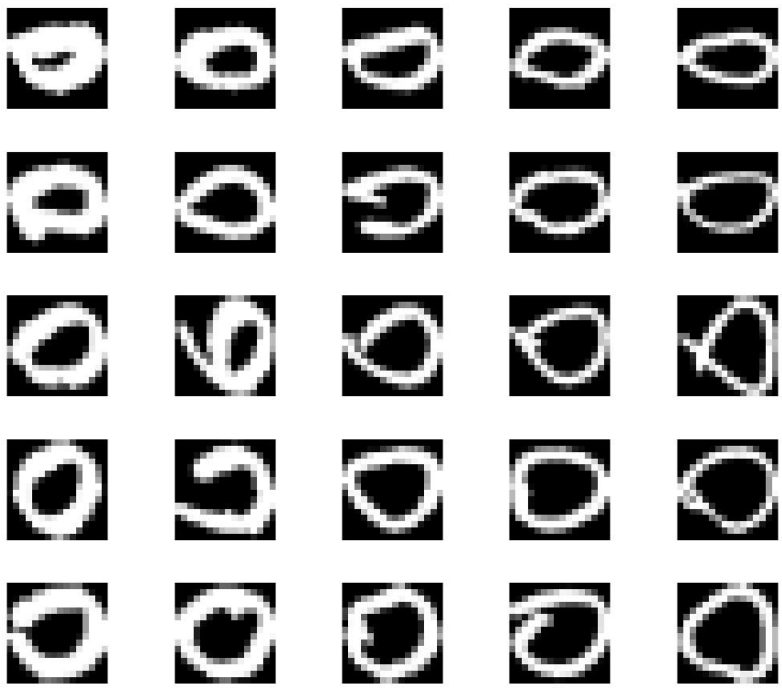
For k = 1, order the images , I present the first 25 pictures here.

1.4 (g)

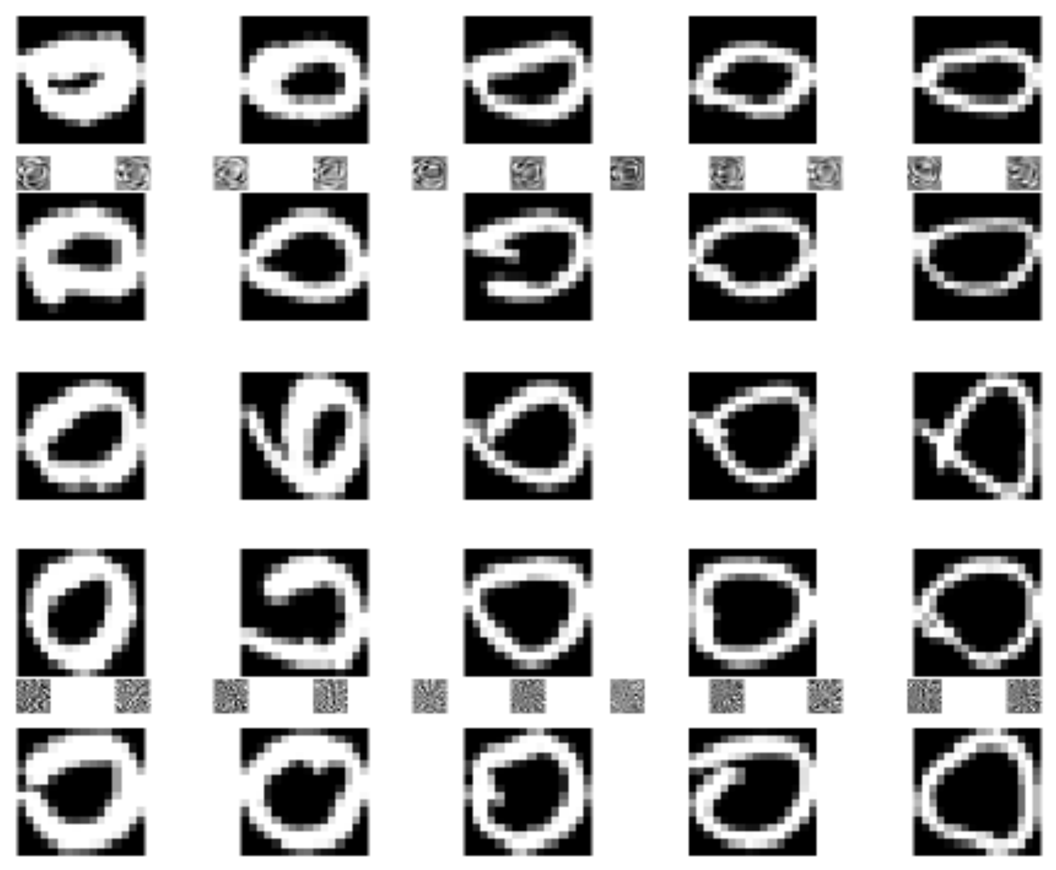
For $k = 2$, scatter plot $(v1, v2)$ with grids. And also show those images on the grids.



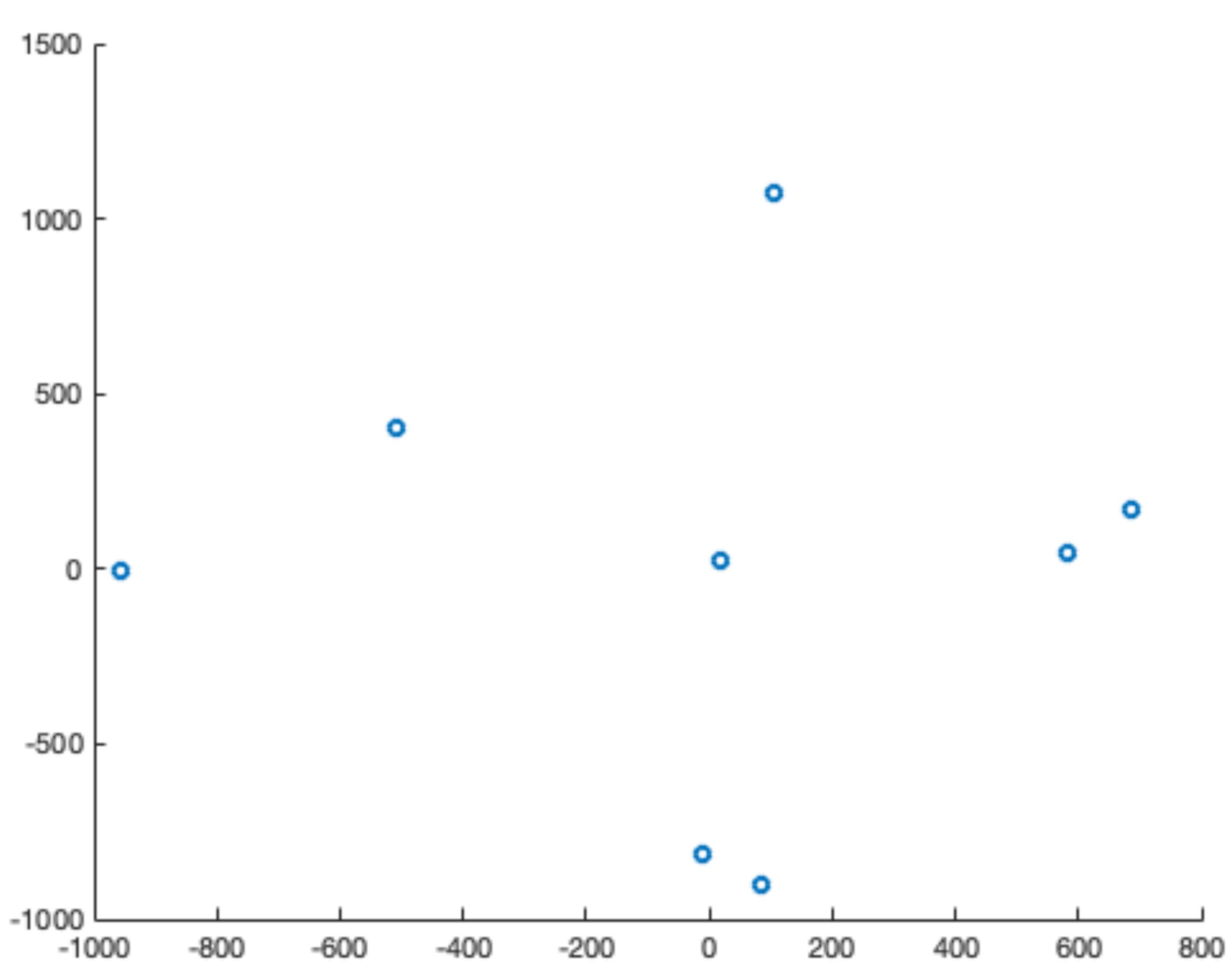

0.0379
0.0168
-0.0009
-0.0180
-0.0375
Second principle component
-0.0402
-0.0214
0.0007
0.0225
0.0383
First principle component



```
%Problem1
%(a)
A = transpose(train); %uploade the train data set p = 256, n = 1194

%(b)
miu_hat = mean(A,2);

e = ones(1,1194);
x = A- miu_hat*e;
%(c)
s = svd(x); %svd of x
total_var = sum(s);
%choose top k-svd, choose k such that explains 95% total variations in data

percent_threshold = 0.95;
percents=0;
for n=1:256
    percents=percents+s(n)/total_var;
    if percents>percent_threshold
        break;
    end
end

% we have
k = 181 ;

% (d)
s_hat = s(1:k,1);
percents= 0;
eigens= zeros(k,1);
for n = 1:k
    percents = percents+s_hat(n)/total_var;
    eigens(n,1) = percents;
end

plot(eigens);
xlabel('Principal Component');
ylabel('Variance Explained (%)');

%(e)
[U,S,V] = svd(x);
%top k components
uk = U(:,1:k);
for i = 1:k
    outpic = uk(:,i);

    min_num = min(outpic);
    max_num = max(outpic);
    outpic = reshape(outpic,16,16);
    %outtitle = ['eigenvalue=' num2str(s_hat(i))];
    subplot(18,11,i),imshow(outpic,[min_num max_num]),%title(outtitle);
end

imshow(reshape(mean(uk,2),16,16),[])

%(f)
v1 = V(:,1);
[out,idx] = sort(v1,'ascend');
sortedX=A(:,idx);
for i = 1:25
    outpic = sortedX(:,i);

    min_num = min(outpic);
    max_num = max(outpic);
    outpic = reshape(outpic,16,16);
    subplot(5,5,i),imshow(outpic,[min_num max_num]);
end

%(g)
v2 = V(:,2);
scatter(v1,v2);
xlabel('First principle component');
ylabel('Second principle component');

a = zeros(1,5);
b = zeros(1,5);

for i = 1:5
    a(i) = prctile(v1,10+(i-1)*20);
    b(i) = prctile(v2,10+(i-1)*20);
end
scatter(v1,v2);
set(gca,'YTickMode','manual','YTick',b);

set(gca,'XTickMode','manual','XTick',a);
grid;
xlabel('First principle component');
ylabel('Second principle component');

[ m, n ] = meshgrid( a,b );
[ res(:,1),res(:,2) ] = deal( reshape(m,[],1), reshape(n,[],1) );

n = size(res,1);
v = [v1,v2];

indx = zeros(n,1);
for i = 1:n
    c = v-res(i,:);
    dis2 = (c(:,1).^2 + c(:,2).^2);
    [~,ind] = min(dis2);
    indx(i) = ind;
end
% find the items near the grid points.

% print these pictures

pic = A(:,indx);
for i = 1:n
    outpic = pic(:,i);

    min_num = min(outpic);
    max_num = max(outpic);
    outpic = reshape(outpic,16,16);
    subplot(5,5,i),imshow(outpic,[min_num max_num]);
end
```



```
% Problem 2
% Input data, in kms.
%distance matrix for: Hangzhou,Shanghai,
%Beijing,Guangzhou,Xian,Wuhan,Chengdu and Hongkong.
D = airdistancekms;

% Multidimensional Scaling algorithm for D
DSquare = D.^2;
[N,~]=size(D);
H = eye(N)-ones(N)/N;
K = -0.5*H*DSquare*H;

[eigVec, eigVal] = eig(K);
d = zeros(N,1);
for n = 1:N
   d(n,1) = eigVal(n,n);
end

% Plot the normalized eigenvalues
total_val = sum(d);
eigens= zeros(N,1);
for n = 1:N
    percents = d(n)/total_val;
    eigens(n,1) = percents;
end

plot(eigens);
% yes we have negative eigenvalues, because those correspoing eigenvectors
% are less informative.
% the first three eigenvectors are more important for the data.

X=eigVec(:,1:3)*eigVal(1:3,1:3).^(1/2);
scatter(-X(:,2),X(:,1));
```



HW1 MATH 5473

3. (a) Recall the definition of an eigenvalue λ

$A\vec{v} = \lambda\vec{v}$

For a matrix to be positive semi-definite, $x^T K x \geq 0$ for $\forall x \in \mathbb{R}^n$.

then $\vec{v}^T A \vec{v} = \vec{v}^T(\lambda\vec{v}) = \lambda v^T v \geq 0$ since $v^T v \geq 0$, then $\lambda \geq 0$

For all eigenvalues $\lambda \geq 0$

then A symmetric. we have $A = UDU^T$ where U is orthogonal and D diagonal with all diagonal entries eigenvalues $\lambda_i \geq 0$.

then we can write $D = C^2$ where C is also a diagonal matrix,

then we have $A = UCCU^T = UC(UC)^T = V^T V$

then $x^T A x = x^T V^T V x = (Vx)^T(Vx) \geq 0$ □.

(b) $d_{ij}^2 = (u_i - u_j)^2 = (u_i - u_j)^T (u_i - u_j) = u_i^T u_i + u_j^T u_j - 2u_i^T u_j$

we have proved that $K \succeq 0$ then K can be written as $K = U^T U$

where $K = \begin{bmatrix} U_1^T \\ U_2^T \\ \vdots \\ U_n^T \end{bmatrix} [U_1, U_2, \cdots U_n] = \begin{bmatrix} U_1^T U_1 & U_1^T U_2 & \cdots & U_1^T U_n \\ U_2^T U_1 & & & \\ U_j^T U_1 & & U_j^T U_j & \\ & & & U_n^T U_n \end{bmatrix}$

$K = [K_{ij}]_{n\times n}$

where $u_i^T u_i = K_{ii}$ and $u_j^T u_j = K_{jj}$ $u_i^T u_j = K_{ij}$

thus $d_{ij} = K_{ii} + K_{jj} - 2K_{ij}$ is a squared distance function.

(c) let $k = \mathrm{diag}(K_{ii}) \in \mathbb{R}^n$ so.

$D = [d_{ij}]_{n\times n} = k \cdot 1^T + 1 \cdot k^T - 2K$ where $K = U^T U$

and $1 = (1, 1, \cdots, 1)^T \in \mathbb{R}^n$

$B_\alpha = -\frac{1}{2} H_\alpha D H_\alpha^T = -\frac{1}{2} H_\alpha (k \cdot 1^T + 1 \cdot k^T - 2K) H_\alpha^T$

$= -\frac{1}{2} H_\alpha (k \cdot 1^T H_\alpha^T + 1 \cdot k^T H_\alpha^T - 2K H_\alpha^T)$

$k \cdot 1^T H_\alpha^T = k \cdot (H_\alpha \cdot 1)^T = k \cdot ((I - e\alpha^T) \cdot 1)^T$

$= k \cdot (1^T (I - e\alpha^T)^T)$

$= k \cdot 1^T \cdot I - k \alpha^T \alpha 1^T$

$= k \cdot 1^T - k \cdot 1^T = 0$

Thus we have $H_\alpha k \mathbf{1}^T H_\alpha^T = H_\alpha \mathbf{1} \cdot k \mathbf{1}^T \cdot H_\alpha^T = (k \mathbf{1}^T H_\alpha^T)^T H_\alpha^T = 0$

then $B_\alpha = -\frac{1}{2} H_\alpha \cdot (-2K H_\alpha^T) = H_\alpha K H_\alpha^T$

$= H_\alpha U^T U H_\alpha^T = (U H_\alpha^T)^T (U H_\alpha^T)$

$= \tilde{U}^T \tilde{U}$

then $x^T B_\alpha x = x^T \tilde{U}^T \tilde{U} x = (\tilde{U} x)^T (\tilde{U} x) \geq 0$

$\square$

(d) If $A \succeq 0$ $B \succeq 0$

$x^T (A+B) x = x^T A x + x^T B x \geq 0 \Rightarrow A+B \succeq 0$

$$a^T (A \circ B) a = [a_1\ a_2\ a_3 \cdots a_n] \begin{bmatrix} A_{11}B_{11} & A_{12}B_{12} & \cdots & A_{1n}B_{1n} \\ A_{21}B_{21} & \cdots & \cdots & A_{2n}B_{2n} \\ \vdots & & & \\ A_{n1}B_{n1} & \cdots & \cdots & A_{nn}B_{nn} \end{bmatrix}_{n\times n} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_n \end{bmatrix}_{n\times 1}$$

$$= [a_1\ a_2\ a_3 \cdots a_n] \begin{bmatrix} \sum_{j=1}^{n} A_{1j} B_{1j} a_j \\ \vdots \\ \sum_{j=1}^{n} A_{ij} B_{ij} a_j \\ \vdots \\ \sum_{j=1}^{n} A_{nj} B_{nj} a_j \end{bmatrix}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} B_{ij} a_j a_i$$

$$= tr\left(A^T diag(a^T) B\, diag(a)\right)$$

we have proved that $A = U^T U$, $B = V^T V$ $\exists\ U, V \in \mathbb{R}^{n\times n}$

$tr(A^T diag(a^T) B\, diag(a)) = tr(U^T U\, diag(a^T) V^T V\, diag(a))$

$= tr(\underline{U^T diag(a^T) V^T} \underline{V diag(a) U})$

$= tr(\tilde{A}^T \tilde{A}) \geq 0$

for $\tilde{A} = V diag(a) U$

thus $(A \circ B)$ is positive semi-definite

4. (a) d^2 is not a distance function.

Assume $d(x,y) = |x-y|$, then $d^2(x,y) = (x-y)^2$

choose $x=3$, $y=1$ and $z=2$,

$d^2(3,1) = 4$ $\quad d^2(2,1) = 1$ $d^2(3,2)=1$, $d^2(3,2) + d^2(2,1) = 2 < d^2(3,1)$

so d^2 is not a distance function.

(b) Fix $x, y, z \in X$; ① $\sqrt{d(x,y)} \geq 0$ since $d(x,y) \geq 0$

② $\sqrt{d(x,y)} = 0$ when $x = y$ ③ $\sqrt{d(x,y)} = \sqrt{d(y,x)}$ since $d(x,y) = d(y,x)$

finally check $(\sqrt{d(x,y)})^2 = d(x,y) \leq d(x,z) + d(z,y)$

$$\leq d(x,z) + d(z,y) + 2\sqrt{d(x,z)}\sqrt{d(z,y)}$$

$$= (\sqrt{d(x,z)} + \sqrt{d(z,y)})^2$$

$$\Rightarrow \sqrt{d(x,y)} \leq \sqrt{d(x,z)} + \sqrt{d(z,y)}$$

So $\sqrt{d}$ is a distance function.

5. (a) If A is a real m-by-n matrix, then there exist orthogonal matrices

$$U = [u_1 | \cdots | u_m] \in \mathbb{R}^{m\times m} \text{ and } V = [v_1 | \cdots | v_n] \in \mathbb{R}^{n\times n}$$

such that

$$U^T A V = \Sigma = \operatorname{diag}(\sigma_1, \cdots, \sigma_p) \in \mathbb{R}^{m\times n}, \quad p = \min\{m, n\},$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$.

Proof: Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be unit 2-norm vectors that satisfy $Ax = \sigma y$ with $\sigma = \|A\|_2$

And there exist $V_2 \in \mathbb{R}^{n\times(n-1)}$ and $U_2 \in \mathbb{R}^{m\times(m-1)}$ so $V = [x | V_2] \in \mathbb{R}^{n\times n}$ and $U = [y | U_2] \in \mathbb{R}^{m\times m}$ are orthogonal then

$$U^T A V = \begin{bmatrix} y^T \\ U_2^T \end{bmatrix} A [x | V_2] = \begin{bmatrix} y^T \\ U_2^T \end{bmatrix} [\sigma y | AV_2]$$

$$= \begin{bmatrix} \sigma & w^T \\ 0 & B \end{bmatrix} \equiv A_1$$

where $w \in \mathbb{R}^{n-1}$ and $B \in \mathbb{R}^{(m-1)\times(n-1)}$. Since

$$\left\| A_1 \left(\begin{bmatrix} \sigma \\ w \end{bmatrix} \right) \right\|_2^2 \geq (\sigma^2 + w^T w)^2$$

we have $\|A_1\|_2^2 \geq (\sigma^2 + w^T w)$. But $\sigma^2 = \|A\|_2^2 = \|A_1\|_2^2$

where $\left(\|A_1\|_2^2 = \max_{\|x\|_2=1} \|A_1 x\|_2^2 = \max_{\|v\|_2=1} \|U^T A V\|^2 = \max_{v^T v=1} v^T A^T U U^T A v \right.$

$\left. = \max_{v^T v=1} v^T A^T A v = \|A\|_2^2 \right)$

Thus we must have $w=0$. Then an obvious induction argument completes the proof of the theorem.

(b) prove that the "best" rank-k approximation of a matrix in the operator norm sense is given by its SVD.

$A_k = U\Sigma_k V^T$ (where $\Sigma_k = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_k, 0, \cdots, 0)$ is a diagonal matrix containing the largest k singular values) is a rank-k matrix that satisfies

$$\|A - A_k\| = \min_{\mathrm{rank}(B)=k} \|A - B\|.$$

proof. Since $U^T A_k V = \mathrm{diag}(\sigma_1, \cdots, \sigma_k, 0, \cdots, 0)$ it follows that A_k is rank k. Moreover, $U^T(A - A_k)V = \mathrm{diag}(0, 0 \cdots 0, \sigma_{k+1}, \cdots, \sigma_p)$ and so $\|A - A_k\|_2 = \sigma_{k+1}$

Now suppose $\mathrm{rank}(B) = k$ for some $B \in \mathbb{R}^{m\times n}$. It follows that we can find orthonormal vectors $x_1, \cdots, x_{n-k}$ so $\mathrm{null}(B) = \mathrm{span}\{x_1, \cdots, x_{n-k}\}$.

A dimension argument shows that

$$\mathrm{span}\{x_1, \cdots, x_{n-k}\} \cap \mathrm{span}\{v_1, \cdots, v_{k+1}\} \neq \{0\}.$$

Let z be a unit 2-norm vector in this intersection. Since $Bz = 0$ and

$$Az = \sum_{i=1}^{k+1} \sigma_i (v_i^T z) u_i$$

we have $\|A - B\|_2^2 \geq \|(A-B)z\|_2^2 = \|Az\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 (v_i^T z)^2 \geq \sigma_{k+1}^2$

then $\min_{\mathrm{rank}(B)=k} \|A - B\| = \|A - A_k\|$

□.

(c) Best rank-k approximation — Frobenius norm:

$$A_k = U\Sigma_k V^T \text{ satisfies } \|A - A_k\|_F = \min_{\operatorname{rank}(B)=k} \|A - B\|_F$$

Let B minimize $\|A-B\|_F^2$ as $\operatorname{rank}(B)=k$. Let V be the space spanned by rows of B. The dimension of V is rank. Since B minimizes $\|A-B\|_F^2$, it must be that each row of B is the projection of the corresponding row of A onto V, otherwise replacing the row of B with the projection of the corresponding row of A onto V does not change V and hence $\operatorname{rank}(B)$ but would reduce $\|A-B\|_F^2$. Since each row of B is the projection of the corresponding row of A, it follows that $\|A-B\|_F^2$ is the sum of squared distances of rows of A to V. Since A_k minimizes the sum of squared distances of rows of A to any k-dimensional subspace, it follows that

$\|A - A_k\|_F \le \|A - B\|_F$

(d) $\|U^T A V\|_p = \|\Sigma\|_p = \left(\sum_i \sigma_i^p\right)^{1/p} = \|A\|_p$

$\square$

(e) According to "Matrix Analysis", Let $A, B \in M_{m,n}$ be given. Then $\|A\| \le \|B\|$ for every unitarily invariant norm $\|\cdot\|$ on $M_{m,n}$ if and only if $\|A\|_{[k]} \le \|B\|_{[k]}$ for each $k = 1, \dots, q = \min\{m,n\}$, where $\|A\|_{[k]} = \sigma_1(A) + \dots + \sigma_k(A)$.

If $B \in M_{m,n}$, and $\operatorname{rank}(B) = k$, then $\sigma_1(B) \ge \dots \ge \sigma_k(B) > 0 = \sigma_{k+1}(B) = \dots = \sigma_q(B)$. Using the fact that a unitarily invariant norm on diagonal matrices in $M_{m,n}$ is monotone norm, we have

$$\begin{aligned}
\|A - B\| &\ge \|\Sigma(A) - \Sigma(B)\| \\
&= \|\operatorname{diag}(\sigma_1(A) - \sigma_1(B), \dots, \sigma_k(A) - \sigma_k(B), \sigma_{k+1}(A), \dots, \sigma_q(A))\| \\
&\ge \|\operatorname{diag}(0, \dots, 0, \sigma_{k+1}(A), \dots, \sigma_q(A))\| \quad (*)
\end{aligned}$$

for any $B \in M_{m,n}$ s.t. $\operatorname{rank}(B) = k$. If $A = V\Sigma(A)W^*$ is a singular value decomposition, we can always attain equality in $(*)$ with $B = V\Sigma_0 W^*$ in which $\Sigma_0 \in M_{m,n}$ is a non-negative diagonal matrix with diagonal entries

$\sigma_1(A), \cdots, \sigma_k(A)$, and $q-k$ zeros. Thus, the same matrix that provides a best rank-k approximation to A in the Frobenius norm provides a best approximation in every unitarily invariant norm.

(f) Closest rotation:

$$\|A - UV^T\|_F = \min_{RR^T = R^TR = I} \|A - R\|_F$$

$\|A - R\|_F = \|U\Sigma V^T - R\|_F = \|\Sigma - V^T R V\|_F$ (by the unitarily invariance)
But the orthogonal (or the unitary) matrices form a group, so we have to minimize

$\|\Sigma - Q\|_F$ over all orthogonal matrices. we have

$$\|\Sigma - Q\|_F^2 = \sum_k (\Sigma_{kk} - Q_{kk})^2 + \sum_{j \neq k} Q_{k,j}^2$$

$$= \sum_k (\Sigma_{kk}^2 + Q_{kk}^2 - 2\Sigma_{kk} Q_{kk}) + \sum_{j \neq k} Q_{k,j}^2$$

$$= \sum_k (\Sigma_{kk}^2 - 2\Sigma_{kk} Q_{kk}) + \sum_{j,k} Q_{k,j}^2$$

$$= Tr(\Sigma^2) + Tr(Q^T Q) - 2\sum_k \Sigma_{kk} Q_{kk}$$

$$= Tr(\Sigma^2) + n - 2\sum_k \Sigma_{kk} Q_{kk}$$

Since the entries of Σ are non-negative and $Q_{kk} = 1$ or -1, choose $Q_{kk} = 1$ for all k, which makes $Q = I$,
then $\|\Sigma - I\|_F = \|U\Sigma V^T - UV^T\|_F = \|A - UV^T\|_F$ is the minimum.