Homework 1

3.(a) $(\Rightarrow)$ If K is positive semi-definte,
for any eigenvalue λ and corresponding eigenvector $\vec{v}$ of K, we have

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

Then $\vec{v}^T K \vec{v} = \lambda \vec{v}^T \vec{v} \geq 0$ since K is positive semi-definite.

$$\vec{v}^T \vec{v} = \|\vec{v}\|^2 \geq 0$$

$$\therefore \lambda \geq 0$$

$(\Leftarrow)$ If the eigenvalues of K are all non-negative,
define the eigenvalues and corresponding eigenvectors as $\lambda_1, \cdots \lambda_n$, $\vec{v_1}, \cdots \vec{v_n}$

we have $\vec{v_i}^T K \vec{v_i} = \lambda_i \vec{v_i}^T \vec{v_i} \geq 0$

Since the eigenvectors are orthogonal and span $\mathbb{R}^n$

$\forall x \in \mathbb{R}^n$, $\vec{x} = c_1\vec{v_1} + \cdots + c_n\vec{v_n}$

$$\vec{x}^T K \vec{x} = c_1 \vec{v_1}^T K \vec{v_1} + \cdots + c_n \vec{v_n}^T K \vec{v_n} \geq 0$$

(b) $K_{ii} + K_{jj} - 2K_{ij} = (e_i - e_j)^T K (e_i - e_j)$

or equivently $(u_i - u_j)^T A (u_i - u_j) = d_A(u_i, u_j) = \|u_i - u_j\|^2$

for $A = (u_i - u_j)^T (e_i - e_j)^T K (e_i - e_j)(u_i - u_j)$

(c) $\forall x \in \mathbb{R}^n$

$$x^T \beta_\alpha x = -\frac{1}{2} x^T H_\alpha D H_\alpha^T x$$
$$= -\frac{1}{2} x^T (I - e\alpha^T) D (I - e\alpha^T)^T x$$
$$= -\frac{1}{2} \|x\|^2 \langle D, -e\alpha^T \rangle = \frac{1}{2} \|x\|^2 \langle D, e\alpha^T \rangle \geq 0$$

β_α is positive semi-definte.

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

$\forall x \in \mathbb{R}^n$, $x^T A x \geq 0$ and $x^T B x \geq 0$

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

$\therefore A+B \succeq 0$

$$x^T (A \circ B) x = x^T (A_{ij} B_{ij})_{ij} x$$
$$= x^T (A_i \circ B_j)(A_i \circ B_j)^T x$$
$$= \left(\sum_k A_{ik} B_{jk} x_k\right)^2 \geq 0$$

$\therefore A \circ B \succeq 0$

4. (a) 1. $d(x,x) = 0 \Rightarrow d^2(x,x) = 0$

2. $d(x,y) > 0 \Rightarrow d^2(x,y) > 0$

3. $d(x,y) = d(y,x) \Rightarrow d^2(x,y) = d^2(y,x)$

4. If $d(x,z) \leq d(x,y) + d(y,z)$

let $d(x,z) = 5$, $d(x,y) = d(y,z) = 3$

$$d(x,z)^2 = 25 > 9 + 9 = d(x,y)^2 + d(y,z)^2$$

$\therefore d^2$ is not a distance function.

(b) 1, 2, 3 properties are satisfied.

If $d(x,z) \le d(x,y) + d(y,z)$

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

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

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

5. (a) $A \in \mathbb{R}^{m\times n}$, we have

$$AA^T = VTV^T$$

$$T = V^T AA^T V$$

T is a positive semi-definte Hermitian matrix and T is upper triangular according to Schur-decomposition.

$$V_1^{m\times n} = A^T V T^{-\frac{1}{2}}$$

$$\Sigma^{m\times n} = [T^{-\frac{1}{2}}\ 0]$$

we have

$$V_1^T V_1 = I$$

$$V^{m\times m} = [V_1\ V_2]$$

$$V\Sigma V^T = V[T^{\frac{1}{2}}\ 0][V_1 V_2]^T = VT^{\frac{1}{2}}T^{-\frac{1}{2}}V^T A = A$$

Therefore, the singular value decomposition exists.

The singular value matrix Σ is unique, the unitary matrices V and V are not unique.

(b) $\|A-A_k\|_2^2 = \sigma_{k+1}^2$

let $A=\sum_{i=1}^{r}\sigma_i u_i v_i^T$ be the singular value decomposition of A. Then $A_k=\sum_{i=1}^{k}\sigma_i u_i v_i^T$ and $A-A_k=\sum_{i=k+1}^{r}\sigma_i u_i v_i^T$

let v be the top singular vector of $A-A_k$.

Express v as a linear combination of $v_1, v_2, \cdots v_r$,

$v=\sum_{i=1}^{r}\alpha_i v_i$. Then

$$|(A-A_k)v| = \left|\sum_{i=k+1}^{r}\sigma_i u_i v_i^T \sum_{j=1}^{r}\alpha_j v_j\right|$$

$$= \left|\sum_{i=1}^{r}\alpha_i \sigma_i u_i v_i^T v_i\right|$$

$$= \left|\sum_{i=k+1}^{r}\alpha_i\sigma_i u_i\right| = \sqrt{\sum_{i=k+1}^{r}\alpha_i^2\sigma_i^2}$$

The v maximizing the last quantity subject to the constraint that $|v|^2=\sum_{i=1}^{r}\alpha_i^2=1$, occurs when $\alpha_{k+1}=1$ and the rest of $\alpha_i=0$. Thus $\|A-A_k\|^2=\sigma_{k+1}^2$.

$\therefore$ A_k is the best rank k 2-norm approximation to A.

Let A be an $n\times d$ matrix. For any matrix B of rank at most k. $\|A-A_k\|_2 \le \|A-B\|_2$.

If A is of rank k or less, the theorem is true since $\|A-A_k\|_2=0$. Assume A is of rank greater than k. $\|A-A_k\|_2^2=\sigma_{k+1}^2$. Suppose there is some matrix B of rank at most k such that B is a better

2-norm approximation to A than A_k. That is, $\|A-B\|_2 < \sigma_{k+1}$. The null space of B, $Null(B)$ has dimension at least $d-k$. Let $v_1, \cdots v_{k+1}$ be the first $k+1$ singular vectors of A.

$\exists z \neq 0$ in $Null(B) \cap Span\{v_1 \cdots v_{k+1}\}$

Scale z so that $|z|=1$

$$\|A-B\|_2^2 \geq |(A-B)z|^2$$

$$Bz=0, \quad \|A-B\|_2^2 \geq |Az|^2$$

Since z is in the span $\{v_1 \cdots v_{k+1}\}$,

$$|Az|^2 = \left|\sum_{i=1}^{n} \sigma_i u_i v_i^T z\right|^2 = \sum_{i=1}^{k} \sigma_i^2 (v_i^T z)^2 = \sigma_{k+1}^2$$

$\|A-B\|_2^2 \geq \sigma_{k+1}^2$ contradicts $\|A-B\|_2 < \sigma_{k+1}$.

(c) Let a be an arbitrary row vector. Since the v_i are orthonormal, the projection of the vector a onto V_k is given by $\sum_{i=1}^{k} (a \cdot v_i) v_i^T$. Thus, the matrix whose rows are the projections of the rows of A onto V_k is given by $\sum_{i=1}^{k} A v_i v_i^T$.

$$\sum_{i=1}^{k} A v_i v_i^T = \sum_{i=1}^{k} \sigma_i u_i v_i^T = A_k$$

For any matrix B of rank at most k

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

Let B minimize $\|A-B\|_F^2$ among all rank k or less matrices. Let V be the space spanned by the

rows of B. The dimension of V is at most k.
Since B minimizes $\|A-B\|_F^2$, it must be that each row of B with the projection of corresponding row of A onto V.

(d) To show that the Schatten p-norm is unitary invariant, we need to show that for any unitary matrix V and any matrix A, we have $\|VAV\|_p = \|A\|_p$.
Note that any matrix A can be decomposed using SVD as $A = U\Sigma V^T$, where U and V are unitary matrices.

$$VAV = V(U^TAV)V\Sigma V^T$$

$$\|VAV\|_p = \|V(U^TAV)V\Sigma V^T\|_p$$

Note that U^TAV is a Hermitian matrix, so it has a spectral decomposition as $U^TAV = Q\Lambda Q^T$.
Using this decomposition, we can write

$$V(U^TAV)V = VQ\Lambda Q^TV = (VQ)(Q^TV)$$

$$\|VAV\|_p = \|(VQ)(QU)\Sigma U^T\|_p$$

$$\|(VQ)(Q^TV)\Sigma V\|_p = \|\Sigma\|_p = \|A\|_p$$

(e) We need to show that for any matrix A and any integer k, the best rank-k approximation of A in terms of any unitarily invariant norm is given by the truncated SVD of A, i.e. $A_k = V_k \Sigma_k V_k^T$, let B be any rank-k matrix that approximates A, i.e. $\text{rank}(B) \leq k$ and $\|A-B\| = \min\{\|A-X\| : \text{rank}(X) \leq k\}$. Using the SVD of A, we can write A as $A = U\Sigma V^T$. Since B has rank at most k, we can write B as $B = XY^T$, where $X, Y \in \mathbb{R}^{m \times n}$.

$$\|A-B\| = \|U\Sigma V^T - XY^T\| = \|\Sigma - X^T U Y\|$$

we have $\|\Sigma - X^T V Y\|_F = \|\Sigma_k - X^T V_k Y_k^T\|_F$

$$\|A-B\|_F = \|\Sigma_k - X^T V_k Y_k^T\|_F$$

$$\|\Sigma_k - X^T V_k Y_k^T\|_F = \|\Sigma_k - (X^T Q)(Q^T V_k) Y_k^T\|_F$$

$X^T V_k = I_k$ and $Y_k V_k = I_k$.

we have $\|A-B\|_F = \|\Sigma_k - I_k\|_F$

$$= (\Sigma_{k+1}^2 + \cdots + \Sigma_r^2)^{\frac{1}{2}}$$

Using this expression, we can write

$$\|A-B\| = \|VAV - VBV\| = \|AV - BV\|$$