CS 596 Homework #1. Xuenan Wang (xw336)

Problem 1:

(a) $(\lambda I - A) = \begin{pmatrix} \lambda-1 & -0.5 \\ 0 & \lambda-(1+\epsilon) \end{pmatrix} = (\lambda-1)[\lambda-(1+\epsilon)] = 0$

$$\lambda^2 - (2+\epsilon)\lambda + (1+\epsilon) = 0$$

we have $\lambda_1 = 1$ $\lambda_2 = 1+\epsilon$

Therefore the eigenvalue is 1 and $1+\epsilon$ (where $\epsilon \neq 0$).

when $\lambda_1 = 1$:

$$\begin{pmatrix} 1-\lambda & -0.5 \\ 0 & \lambda-(1+\epsilon) \end{pmatrix} = \begin{pmatrix} 0 & -0.5 \\ 0 & -\epsilon \end{pmatrix} = 0$$

$\xi_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ unit norm: $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

when $\lambda_2 = 1+\epsilon$:

$$\begin{pmatrix} 1-\lambda & -0.5 \\ 0 & \lambda-(1+\epsilon) \end{pmatrix} = \begin{pmatrix} \epsilon & -0.5 \\ 0 & 0 \end{pmatrix} = 0$$

$\xi_2 = \begin{pmatrix} 0.5 \\ \epsilon \end{pmatrix}$ unit norm: $\frac{1}{\sqrt{0.5^2+\epsilon^2}} \begin{pmatrix} 0.5 \\ \epsilon \end{pmatrix}$

(b) Since $\epsilon \neq 0$ and $\lambda_1 \neq \lambda_2$, we know ~~that~~ matrix A can be diagonalize.

$$P^{-1} A P = \Lambda$$

$$\begin{pmatrix} 1 & 0 \\ 0 & 1+\epsilon \end{pmatrix}$$

(c) While $\epsilon \to 0$, λ_1 and λ_2 become more and more close. When $\epsilon = 0$, we will have two identical eigenvalue, which will result that matrix A can no longer be diagonalized.

Problem 2:

(a) we have:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \\ a_{k1} & a_{k2} & \cdots & a_{km} \end{pmatrix} \qquad B = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1m} \\ b_{21} & b_{22} & \cdots & b_{2m} \\ \vdots & & & \\ b_{k1} & b_{k2} & \cdots & b_{km} \end{pmatrix}$$

easily we can show:

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{k1} \\ a_{12} & a_{22} & \cdots & a_{k2} \\ \vdots & & & \\ a_{1m} & a_{2m} & \cdots & a_{km} \end{pmatrix} \qquad B^T = \begin{pmatrix} b_{11} & b_{21} & \cdots & b_{k1} \\ b_{12} & b_{22} & \cdots & b_{k2} \\ \vdots & & & \\ b_{1m} & b_{2m} & \cdots & b_{km} \end{pmatrix}$$

$$\text{trace}(AB^T) = (a_{11}b_{11} + a_{12}b_{12} + \cdots + a_{1m}b_{1m}) + (a_{21}b_{21} + a_{22}b_{22} + \cdots + a_{2m}b_{2m})$$
$$+ \cdots + (a_{k1}b_{k1} + a_{k2}b_{k2} + \cdots + a_{km}b_{km}) = \sum_{i=1}^{k}\sum_{j=1}^{m} X_{ij} Y_{ij}$$

$$\text{trace}(B^TA) = (b_{11}a_{11} + b_{21}a_{21} + \cdots + a_{k1}b_{k1}) + (b_{12}a_{12} + b_{22}a_{22} + \cdots + b_{k2}a_{k2})$$
$$+ \cdots + (b_{1m}a_{1m} + b_{2m}a_{2m} + \cdots + b_{km}a_{km}) = \sum_{i=1}^{k}\sum_{j=1}^{m} X_{ij} Y_{ij}$$

$$\text{trace}(BA^T) = (b_{11}a_{11} + b_{12}a_{12} + \cdots + b_{1m}a_{1m}) + (b_{21}a_{21} + b_{22}a_{22} + \cdots + b_{2m}a_{2m})$$
$$+ \cdots + (b_{k1}a_{k1} + b_{k2}a_{k2} + \cdots + b_{km}a_{km}) = \sum_{i=1}^{k}\sum_{j=1}^{m} X_{ij} Y_{ij}$$

$$\text{trace}(A^TB) = (a_{11}b_{11} + a_{21}b_{21} + \cdots + a_{k1}b_{k1}) + (a_{12}b_{12} + a_{22}b_{22} + \cdots + a_{k2}b_{k2})$$
$$+ \cdots + (a_{1m}b_{1m} + a_{2m}b_{2m} + \cdots + a_{km}b_{km}) = \sum_{i=1}^{k}\sum_{j=1}^{m} X_{ij} Y_{ij}$$

Therefore, we have $\text{trace}(AB^T) = \text{trace}(B^TA) = \text{trace}(BA^T) = \text{trace}(A^TB)$

(b) Since $E[xx^T] = Q$ and the trace of a scalar is the scalar itself.

$$E(x^TAx) = \text{trace}(E(x^TAx))$$

Because trace and expectation is linear operation.

$$\text{trace}(E(x^TAx)) = E(\text{trace}(x^TAx))$$
$$= E(\text{trace}(Axx^T))$$
$$= \text{trace}(E(Axx^T))$$

$$= \text{trace}(A E(x x^T))$$
$$= \text{trace}(AQ)$$
$$= Q\, \text{trace}(A)$$

(c) $\text{trace}(UAU^{-1}) = \text{trace}(UU^{-1}A) = \text{trace}(IA) = \text{trace}(A)$

Therefore, the trace does not change if we apply a similarity transformation.

(d) $P^{-1}AP = \Lambda \iff P\Lambda P^{-1} = A$

$\text{trace}(P\Lambda P^{-1}) = \text{trace}(\Lambda) = \text{trace}(A)$

trace of Λ is the sum of diagonal elements which is also sum of eigenvalues.

(e) According to the defination of trace, if matrix A is real then trace(A) must be real. Even though eigenvalue can be complex, they show up in a pair. Complex parts will disppear when doing summation.

(f) Assume $\lambda^k + C_{k-1}\lambda^{k-1} + \cdots + C_0 = 0$ and the eigenvalues are $\lambda_1, \lambda_2 \cdots \lambda_k$

we have $(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_k) = 0$

$$C_{k-1} = -\left(\sum_{i=1}^{k} \lambda_i\right) = -\text{trace}(A)$$
$$C_0 = (-1)^k \lambda_1 \lambda_2 \cdots \lambda_k = (-1)^k \det(A)$$

Problem 3.

(a) Assume ~~that~~ λ is a eigenvalue of A. x $(x \neq 0)$ is the eigenvector respectly. we have:

$$Ax = \lambda x$$
$$A^r x = \lambda^r x = 0$$
$$\therefore \lambda^r = 0 \quad \Rightarrow \quad \lambda = 0$$

(b) No. Because $\lambda_i = 0$, matrix A has i eigenvalues that all equal to each other, this matrix can not be diagonlized.

(c) $\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}$

(d) Assume $B = P^{-1}AP$

we have $A^r = 0$, $B^r = (P^{-1}AP)^r$

$= P^{-1}AP \cdot P^{-1}AP \cdots P^{-1}AP$

$= P^{-1}A^rP = 0$

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ is a second example of a nilpotent matrix.

Problem 4.

(a) $\min\limits_{A} \sum\limits_{n=1}^{t} (y_n - X_n^T A)^2 \rightarrow \min \| y_n - X_n^T A \|^2$

$w = (y_n - X_n^T A)^T (y_n - X_n^T A) = y_n^T y_n - A^T X_n^T y_n - y_n^T X_n A + A^T X_n^T X_n A$

$\frac{\partial w}{\partial A} = -2X_n y_n + 2X_n X_n^T A = 0$

$X_n X_n^T A = y_n X_n$

$A = (X_n X_n^T)^{-1} y_n X_n$

Since $R_t = \sum\limits_{n=1}^{t} X_n X_n^T$, $U_t = \sum\limits_{n=1}^{t} y_n X_n$ and $A_t = R_t^{-1} U_t$.

we have $A_t = \left(\sum\limits_{n=1}^{t} X_n X_n^T \right)^{-1} \cdot \sum\limits_{n=1}^{t} y_n X_n$

$= \sum\limits_{n=1}^{t} (X_n X_n^T)^{-1} y_n X_n$ this is same as $(X_n X_n^T)^{-1} y_n X_n$.

That's what we have above.

(b) $A_t = R_t^{-1} U_t$ $\quad R_t \rightarrow \sum\limits_{n=1}^{t} X_n X_n^T : k^2 + t$ $\quad R_t \rightarrow R_t^{-1} : k^3 + k^2 + t$

$U_t \rightarrow k + t$

$A_t = k^3 + k^2 + t + k + t + k^3 \Rightarrow O(k^3)$

(c) $R_{t+1} = \sum_{n=1}^{t+1} X_n X_n^T = \sum_{n=1}^{t} X_n X_n^T + X_{t+1} X_{t+1}^T$

$X_{t+1} y_{t+1} = \sum_{n=1}^{t+1} X_n y_n = \sum_{n=1}^{t} X_n y_n + X_{t+1} y_{t+1} = X_t y_t + X_{t+1} y_{t+1}$

$A_{t+1} = R_{t+1}^{-1} U_{t+1}$

$A_t = R_t^{-1} (X_t y_t) \qquad R_t A_t = X_t y_t = U_t$

$A_{t+1} = R_{t+1}^{-1} (R_t A_t + X_{t+1} y_{t+1})$

$= R_{t+1}^{-1} (R_{t+1} A_t - X_{t+1} X_{t+1}^T A_t + X_{t+1} y_{t+1})$

$= A_t + R_{t+1}^{-1} X_{t+1} (y_{t+1} - X_{t+1}^T A_t)$

$e_{t+1} = y_{t+1} - X_{t+1}^T A_t$

$K_{t+1} = Q_t X_{t+1}$

$r_{t+1} = 1 + X_{t+1}^T K_{t+1}$

$A_{t+1} = A_t + e_{t+1} K_{t+1}$

$Q_{t+1} = Q_t - \frac{1}{r_{t+1}} K_{t+1} K_{t+1}^T$

$(A+BCD)^{-1} = A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1}$

$A = R_t \qquad B = X_{t+1} \qquad C = 1 \qquad D = X_{t+1}^T$

$R_{t+1} = R_t + X_{t+1} X_{t+1}^T$

$R_{t+1}^{-1} = (R_t + X_{t+1} X_{t+1}^T)^{-1}$

$= A^{-1} - A^{-1} B (C^{-1} + D A^{-1} B)^{-1} D A^{-1}$

$= R_t^{-1} - R_t^{-1} X_{t+1} (1 + X_{t+1}^T R_t^{-1} X_{t+1})^{-1} X_{t+1}^T R_t^{-1}$

$= R_t^{-1} - \frac{R_t^{-1} X_{t+1} X_{t+1}^T R_t^{-1}}{1 + X_{t+1}^T R_t^{-1} X_{t+1}^T}$

$= Q_{t+1}$

$Q_t = R_t^{-1}$

Therefore, we can compute A_{t+1} by using the previous A_t.

Complexity: $O(k^2)$