596 Midterm Yujia Fan (yf198)

Problem 1

(a) V is a vector space.

For addition, denote $P_1(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k$, $k \le n$, $P_1(x) \in V$

$P_2(x) = b_0 + b_1 x + b_2 x^2 + \cdots + b_m x^m$, $m \le n$, $P_2(x) \in V$

then $P_3(x) = P_1(x) + P_2(x) = C_0 + C_1 x^1 + C_2 x^2 + \cdots + C_p x^p$, $p = \max\{k, m\} \le n$, $C_0 = a_0 + b_0$, $C_1 = a_1 + b_1$, $\cdots$ and so on.

so $P_3(x)$ still belongs to V.

For multiplication, if there exists a scalar λ, $\lambda P_1(x) = \lambda a_0 + a_1 \lambda x^1 + a_2 \lambda x^2 + \cdots + a_k \lambda x^k$, $k \le n$

$= \alpha_0 + \alpha_1 x^1 + \alpha_2 x^2 + \cdots + \alpha_k x^k$, $k \le n$

$\lambda P_1(x)$ still belongs to V.

(b) V is finite dimensional and its dimension is $(n+1)$.

(c) $(1, x, x^2, \cdots, x^n)$

(d) For vector space V, the set $\{0\}$ and V itself are subspaces of V.

Also, if W denotes the space of all polynomials $p(x)$ of order up to $(n-1)$, then W is a subspace of V.

(e) Define an inner product on V by $\langle p(x), q(x) \rangle = \int_0^1 p(x) q(x) dx$.

where $p(x) = p_0 + p_1 x + p_2 x^2 + \cdots + p_n x^n$

$q(x) = q_0 + q_1 x + q_2 x^2 + \cdots + q_n^n x^n$

Positivity: $\langle p(x), p(x) \rangle = \int_0^1 p(x) p(x) dx = \int_0^1 |p(x)|^2 dx$ is nonnegative

Definiteness: $\langle p(x), p(x) \rangle = \int_0^1 |p(x)|^2 dx$ is zero iff $|p(x)| = 0$, that's iff p is zero polynomial.

Additivity, homogeneity: follow from the distributivity of polynomial multiplication and linearity of the integral.

Conjugate Symmetry: $\langle q(x), p(x) \rangle = \int_0^1 q(x) p(x) dx = \int_0^1 p(x) q(x) dx = \langle p(x), q(x) \rangle$

the norm for polynomials $\| p(x) \| = \| \langle p(x), p(x) \rangle \| = \| \int_0^1 |p(x)|^2 dx \| = \sqrt{\int_0^1 |p(x)|^2 dx}$

Problem 2

a) Since Q is real symmetric, there exists an unitary matrix U which satisfies:

$$Q = UAU^T, \quad \text{where } A = diag(\lambda_1, \lambda_2, \cdots, \lambda_k)$$

So $$(\bar{X})^T Q X = (\bar{X})^T U A U^T (X) = (U^T\bar{X})^T A (U^T X)$$

Denote $U^T = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ u_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ u_{k1} & \cdots & & u_{kk} \end{bmatrix}$, $\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_k \end{bmatrix}$, $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix}$

So $$(U^T\bar{X})^T A (U^T X) = \begin{bmatrix} u_{11}\bar{x}_1 + u_{12}\bar{x}_2 + \cdots + u_{1k}\bar{x}_k \\ u_{21}\bar{x}_1 + u_{22}\bar{x}_2 + \cdots + u_{2k}\bar{x}_k \\ \vdots \\ u_{k1}\bar{x}_1 + u_{k2}\bar{x}_2 + \cdots + u_{kk}\bar{x}_k \end{bmatrix}^T \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_k \end{bmatrix} \begin{bmatrix} u_{11}x_1 + u_{12}x_2 + \cdots + u_{1k}x_k \\ u_{21}x_1 + u_{22}x_2 + \cdots + u_{2k}x_k \\ \vdots \\ u_{k1}x_1 + u_{k2}x_2 + \cdots + u_{kk}x_k \end{bmatrix}$$

$$\begin{aligned} &= (u_{11}\bar{x}_1 + u_{12}\bar{x}_2 + \cdots + u_{1k}\bar{x}_k)\lambda_1(u_{11}x_1 + u_{12}x_2 + \cdots + u_{1k}x_k) \\ &+ (u_{21}\bar{x}_1 + u_{22}\bar{x}_2 + \cdots + u_{2k}\bar{x}_k)\lambda_2(u_{21}x_1 + u_{22}x_2 + \cdots + u_{2k}x_k) \\ &+ \cdots \\ &+ (u_{k1}\bar{x}_1 + u_{k2}\bar{x}_2 + \cdots + u_{kk}\bar{x}_k)\lambda_k(u_{k1}x_1 + u_{k2}x_2 + \cdots + u_{kk}x_k) \end{aligned}$$

let $M_i = u_{i1}x_1 + u_{i2}x_2 + \cdots + u_{ik}x_k$

$\Rightarrow \bar{M}_i = u_{i1}\bar{x}_1 + u_{i2}\bar{x}_2 + \cdots + u_{ik}\bar{x}_k$

so $(\bar{X})^T Q X = (U^T\bar{X})^T A (U^T X) = \sum_{i=1}^{k} \lambda_i \bar{M}_i M_i$

In total, $$\frac{(\bar{X})^T Q X}{(\bar{X})^T X} = \frac{\sum_{i=1}^{k} \lambda_i \bar{M}_i M_i}{\sum_{i=1}^{k} \bar{x}_i x_i} \quad (*)$$

and $\lambda_1 \sum_{i=1}^{k} \bar{M}_i M_i \le \sum_{i=1}^{k} \lambda_i \bar{M}_i M_i \le \lambda_k \sum_{i=1}^{k} \bar{M}_i M_i$

$$\sum_{i=1}^{k} \bar{M}_i M_i = \sum_{j=1}^{k} \sum_{n=1}^{k} \left(\sum_{i=1}^{k} \bar{u}_{ni} u_{ij} \right) \bar{x}_j x_n$$

because $U^T U = I \Rightarrow I_{jn} = \sum_{i=1}^{k} \overline{u_{ji}}\, u_{in}$, $I_{jn} = 0$ if $j \ne k$; $I_{jn} = 1$ if $j = k$.

So we can get $\sum_{i=1}^{k} \bar{M}_i M_i = \sum_{i=1}^{k} \bar{x}_i x_i$

substitute $\sum_{i=1}^{k} \bar{M}_i M_i = \sum_{i=1}^{k} \bar{x}_i x_i$ in $(*)$:

$$\lambda_1 \le \frac{(\bar{X})^T Q X}{(\bar{X})^T X} \le \lambda_k$$

(b) A^TA is a real, symmetric matrix & σ_i^2 are eigenvalues of A^TA

From (a) we know, for any complex vector X:

$$\sigma_k^2 \leq \frac{(\bar{X})^T A^T A X}{(\bar{X})^T X} \leq \sigma_1^2$$

For matrix A, if X is a eigenvector of A, then $AX = \lambda_i X$ (λ_i is a eigenvalue of A)

also, $AX = \lambda_i X \Rightarrow A\bar{X} = \bar{\lambda}_i \bar{X}$ (A is a real matrix)

So $\frac{(\bar{X})^T A^T A X}{(\bar{X})^T X} = \frac{(A\bar{X})^T (AX)}{(\bar{X})^T X} = \frac{(\bar{\lambda}_i \bar{X})^T (\lambda_i X)}{(\bar{X})^T X} = \frac{\bar{\lambda}_i \lambda_i (\bar{X}^T X)}{(\bar{X}^T X)} = \bar{\lambda}_i \lambda_i$

If $\lambda_i = a_i + b_i i$, then $\bar{\lambda}_i = a_i - b_i i$, so $\bar{\lambda}_i \lambda_i = (a_i - b_i i)(a_i + b_i i) = a_i^2 + b_i^2$

so
$$\sigma_k^2 \leq (a_i^2 + b_i^2) \leq \sigma_1^2$$
$$\sigma_k \leq \sqrt{a_i^2 + b_i^2} \leq \sigma_1$$
$$\sigma_k \leq |\lambda_i| \leq \sigma_1$$

Also, if λ_i is a real number, $\sigma_k \leq |\lambda_i| \leq \sigma_1$ still holds.

Problem 3

(a) since $P=P^2$, $PX=\lambda X \Rightarrow P^2X=\lambda X$

also $P^2X = P(PX) = P\cdot\lambda X = \lambda PX = \lambda(\lambda X) = \lambda^2 X = \lambda X$

$\Rightarrow (\lambda^2-\lambda)X = 0$, X is a non-zero vector.

$\lambda^2-\lambda=0 \Rightarrow \lambda = 0$ or 1

(b) $(I-P)^2 = I-2P+P^2 = I-2P+P = I-P$

so $(I-P)$ is also a projection.

(c) $(X-PX)^TPX = X^TPX - X^TP^TPX = X^TPX - X^TP^2X = X^TPX - X^TPX = 0$

so $X-PX$ and PX are orthogonal.

(d) $P = A(B^TA)^{-1}B^T$

B^TA is a $n\times n$ matrix, if $(B^TA)^{-1}$ exists, B^TA should be nonsingular

If $\underset{(n\times m)}{B^T}\underset{(m\times n)}{A}$ is nonsigular, there must be $n<m$.

so the condition is : $n\leq m$ and B^TA is a nonsingular matrix.

then $P^2 = A(B^TA)^{-1}(B^TA)(B^TA)^{-1}B^T = A(B^TA)^{-1}B = P$

If P is also symmetric,

$P^T = [A(B^TA)^{-1}B^T]^T = B[(B^TA)^{-1}]^TA^T = B[(B^TA)^T]^{-1}A^T = B(A^TB)^{-1}A^T$

So we need $P^T = B(A^TB)^{-1}A^T = BB^{-1}(A^T)^{-1}A^T = I = P$

If so, the condition is : $m=n$ & A, B are both nonsingular. or $A=B$

(e) With $\hat{b}=Ax$, the goal is to find optimum $\hat{x}$ where $\min\limits_{\hat{x}} \|b-A\hat{x}\|$

Define the column space $C(A) = \{Ax \mid x\in R^n\}$, A is $m\times n$.

We want to map $\hat{x}\in R^n$ to $p = A\hat{x}\in C(A)$, that is $\hat{x}\xrightarrow{A}p$

So the minimum error $e = b-A\hat{x} = b-p$ should be orthogonal to p

Let P be the orthogonal projection matrix of vector of b so that $p=Pb$

According to the properties of projection matrix, $P^2=P=P^T$:

$Pp = PPb = P^2b = Pb = p \Rightarrow P(b-p) = p-p = 0$

And also, $e=b-A\hat{x}=b-p$ belongs to orthogonal complement of $C(A)$, $C(A)^\perp = N(A^T)$

$\Rightarrow A^Te = A^T(b-A\hat{x}) = 0$

the normal equation is $A^TA\hat{x}=A^Tb$ and A^TA is invertible

the optimum $\hat{x} = (A^TA)^{-1}A^Tb$

so $\hat{b} = p = A\hat{x} = A(A^TA)^{-1}A^Tb$

the orthogonal projection matrix $P = A(A^TA)^{-1}A^T$.

Problem 4

(a) $\hat{x} = az + bw$, we want to minimize $x - \hat{x}$.

According to orthogonal principle, denote optimum $\hat{x}$ as $\hat{x}^* = a^*z + b^*w$.

$$\left.\begin{array}{l} \langle x-\hat{x}^*, z\rangle = 0 \\ \langle x-\hat{x}^*, w\rangle = 0 \end{array}\right\} \Rightarrow \left.\begin{array}{l} \langle x - a^*z - b^*w,\ z\rangle = 0 \\ \langle x - a^*z - b^*w,\ w\rangle = 0 \end{array}\right\} \Rightarrow \begin{array}{l} \langle x, z\rangle - \langle a^*z + b^*w, z\rangle = 0 \\ \langle x, w\rangle - \langle a^*z + b^*w, w\rangle = 0 \end{array}$$

$$\Rightarrow \begin{array}{l} \langle x, z\rangle = a^*\langle z, z\rangle + b^*\langle w, z\rangle \\ \langle x, w\rangle = a^*\langle z, w\rangle + b^*\langle w, w\rangle \end{array}$$

$$\Rightarrow \text{so } a^* = \frac{\langle w, z\rangle\langle w, x\rangle - \langle w, w\rangle\langle x, z\rangle}{\langle z, w\rangle\langle w, z\rangle - \langle z, z\rangle\langle w, w\rangle}, \quad b^* = \frac{\langle z, w\rangle\langle x, z\rangle - \langle z, z\rangle\langle x, w\rangle}{\langle w, z\rangle\langle z, w\rangle - \langle z, z\rangle\langle w, w\rangle}$$

also, $$a^* = \frac{E(zw)E(xw) - E[w^2]E[xz]}{E^2(zw) - E[z^2]E[w^2]}, \quad b^* = \frac{E[zw]E[xz] - E[z^2]E[xw]}{E^2[zw] - E[z^2]E[w^2]}$$

(b) The optimum approximation

$$\hat{x}^* = a^*z + b^*w = \frac{E(zw)E(xw) - E[w^2]E[xz]}{E^2(zw) - E[z^2]E[w^2]}\cdot z + \frac{E[zw]E[xz] - E[z^2]E[xw]}{E^2[zw] - E[z^2]E[w^2]}\cdot w$$

To find the minimum distance $\min d = \|x - \hat{x}^*\|$,

we first consider $\|x - \hat{x}^*\|^2 = \langle x - \hat{x}^*, x - \hat{x}^*\rangle = \langle x - a^*z - b^*w, x - \hat{x}^*\rangle = \langle x, x - \hat{x}^*\rangle$

And
$$\begin{aligned} \langle x, x - \hat{x}^*\rangle &= \langle x, x\rangle - \langle x, \hat{x}^*\rangle \\ &= \langle x, x\rangle - \langle x, a^*z + b^*w\rangle \\ &= \langle x, x\rangle - a^*\langle x, z\rangle - b^*\langle x, w\rangle \\ &= E[x^2] - a^*E[xz] - b^*E[xw] \end{aligned}$$

so min distance
$$= \sqrt{E[x^2] - a^*E[xz] - b^*E[xw]}$$

$$= \sqrt{E[x^2] - \frac{E(zw)E(xw) - E[w^2]E[xz]}{E^2(zw) - E[z^2]E[w^2]}\cdot E[xz] - \frac{E[zw]E[xz] - E[z^2]E[xw]}{E^2[zw] - E[z^2]E[w^2]}\cdot E[xw]}$$