Midterm CS 596 Yan Gu (yg369)

Problem 1

a) we define two operations:

$\oplus : V + V \to V$, $\odot : R \times V \to V$

$P_i = P(x)_i = a_{0i} + a_{1i} X + a_{2i} X^2 + \cdots + a_{ni} X^n$ for fix $n \in I$

$a_{0i}, a_{1i}, a_{2i}, a_{3i}, \cdots, a_{ni} \in R$

So Assume we have $P_1, P_2, P_3 \in V$ (random pick).

it is easy to know:

1) $P_1 \oplus (P_2 \oplus P_3)$

$= (a_{01} + a_{11} X + \cdots + a_{n1} X^n) + [(a_{02} + a_{03}) + (a_{12} + a_{13}) X + \cdots + (a_{n2} + a_{n3}) X^n]$

$= (a_{01} + a_{02} + a_{03}) + (a_{11} + a_{12} + a_{13}) X + \cdots + (a_{n1} + a_{n2} + a_{n3}) X^n$

$= [(a_{01} + a_{02}) + (a_{11} + a_{12}) X + \cdots + (a_{n1} + a_{n2}) X^n] + (a_{03} + a_{13} X + \cdots + a_{n3} X^n)$

$= (P_1 \oplus P_2) \oplus P_3$

In the same way we know:

2) $P_1 \oplus P_2 = P_2 + P_1$

3) $P_1 \oplus 0 = a_{01} + a_{11} X + a_{21} X^2 + \cdots + a_{n1} X^n + 0 = P_1$

4) $(-P_1) \oplus P_1 = -(a_{01} + a_{11} X + \cdots a_{n1} X^n) + (a_{01} + a_{11} X + \cdots + a_{n1} X^n)$

$= 0$

Suppose we have $\lambda_1, \lambda_2 \in R$. we have..

5) $\lambda_1 \odot (\lambda_2 \odot P_1) = \lambda_1 \cdot (\lambda_2 a_{01} + \lambda_2 a_{11} X + \cdots + \lambda_2 a_{n1} X^n)$

$= \lambda_1 \lambda_2 a_{01} + \lambda_1 \lambda_2 a_{11} X + \cdots + \lambda_1 \lambda_2 a_{n1} X^n$

$= (\lambda_1 \cdot \lambda_2) \odot P_1$

6) $1 \odot p_1 = 1 \cdot (a_{01} + a_{11}x + \cdots + a_{n1}x^n) = p_1$

7) $\lambda_1 \odot p_1 = \lambda_1 a_{01} + \lambda_1 a_{11} x + \lambda_1 a_{21} x^2 + \cdots + \lambda_1 a_{n1} x^n$

Suppose we have a ~~p_v where~~ : p_j where :

$a_{0j} = \lambda_1 a_{01}$. $a_{1j} = \lambda_1 a_{11}$ $\cdots$ for all $i = 0, 1, \cdots, n$

so $\lambda_1 \odot p_1 = p_j \in V$

8) $\lambda_1 \odot (p_1 \oplus p_2)$

$= \lambda_1 \cdot \left[(a_{01} + a_{02}) + (a_{11} + a_{12})x + (a_{21} + a_{22})x^2 + \cdots + (a_{n1} + a_{n2})x^n \right]$

$= \lambda_1 a_{01} + \lambda_1 a_{02} + \lambda_1 a_{11} x + \lambda_1 a_{12} x + \cdots + \lambda_1 a_{n1} x^n + \lambda_1 a_{n2} x^n$

$= \lambda_1 (a_{01} + a_{11} x + \cdots + a_{n1} x^n) + \lambda_1 (a_{02} + a_{12} x + \cdots + a_{n2} x^n)$

$= (\lambda_1 \odot p_1) \oplus (\lambda_1 \odot p_2)$

9) $(\lambda_1 + \lambda_2) \odot p_1 = (\lambda_1 + \lambda_2) \cdot (a_{01} + a_{11} x + \cdots + a_{n1} x^n)$

$= \lambda_1 (a_{01} + a_{11} x + \cdots + a_{n1} x^n) + \lambda_2 (a_{01} + a_{11} x + \cdots + a_{n1} x^n)$

$= (\lambda_1 \odot p_1) \oplus (\lambda_2 \odot p_1)$

Hence: V is a vector space. with

$\oplus : V + V \Rightarrow V$, $\odot : R \times V \Rightarrow V$

b): The dimension is $n+1$:

Since every polynomial ~~is~~ is of form $a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$

where $a_i \in R$, $i = 0, 1, \cdots, n$, $n \in R$ ~~so~~

so for example it can has basis of $\{x^n, x^{n-1}, \cdots, x^2, x^1, 1\}$

c). From b): we already defined.

so we can have

$\{1, x, x^2, x^3, \cdots, x^{n-1}, x^n\}$ as a basis of V

d) ① Suppose we define $a_0 = a_1 = a_2 = \cdots = a_n = 0$.

So we assume subspace $W_1 = \{ p(x) \in V \mid p(x) = 0, \text{ for all } x \}$

Proof:

I. for all x, zero vector is $p(x)$ since $p(x) = 0 \in V$ when $a_0 = a_1 = \cdots = a_n = 0$

~~so~~ $p(x) \in W_1$

so ~~it~~ condition I met

II. for u and v in W_1, $u = v = 0$ they are zero vector, so

$u, v \in V$ and $u + v = 0$

so $u + v \in W_1$

II met.

III. for scalar $c \in R$, $c \cdot u = 0 \Rightarrow$ it is zero vector

so $c \cdot u \in W_1$

Hence W_1 is linear subspace

② Suppose we have $p(x) = x$,

$W_2 = \{ p(x) \in V \mid p(x) = x \; \cancel{\text{for}} \}$

Proof:

I. for ~~all~~ $x = 0$, $p(0) = 0$ ~~it~~ is a zero vector in V,

and $p(0) = 0 \in W_2 \Rightarrow$ I met

II. for $u, v \in W_2$, $u = p(x_1) = x_1$, $v = p(x_2) = x_2$

$u + v = x_1 + x_2$

Suppose $x_3 = x_1 + x_2 \Rightarrow p(x_3) = x_1 + x_2 \Rightarrow p(x_3) \in W_2$

III. for $c \in R$, $c \cdot u = c \cdot p(x_1) = cx_1$

suppose $x_4 = c \cdot x_1 \Rightarrow p(x_4) = c \cdot p(x_1) \in W_2$

Hence, W_2 is a linear subspace.

③: let W_3 be the following subset of ~~of~~ P_n, where P_n is a degree of ~~n~~ ~~polynia~~ polynomial ~~space in V~~ subspace in V and it ~~is~~ is easy to show as ~~in~~ the same method in ②

where P_n is V,

so we define

$$W_3 = \{ p(x) \in V \mid p'(-1) = 0 \text{ and } p''(1) = 0 \}$$

$p'(x)$ is $\frac{d(p(x))}{dx}$, $p''(x) = \frac{d^2[p(x)]}{dx^2}$

I. Note that zero vector in V is zero polynomial we denote as $\theta(x) \Rightarrow \theta(x) = 0$ for all x.

so the derivate of $\theta(x)$ is

$$\theta'(x) = 0 \text{ and } \theta''(x) = 0.$$

$\Rightarrow \theta'(-1) = 0$ and $\theta''(1) = 0 \Rightarrow \theta(x)$ in W_3

I. met

II. let $f(x), g(x) \in W_3 \Rightarrow f'(-1) = g'(-1) = 0$, $f''(1) = g''(1) = 0$.

let $h(x) = f(x) + g(x)$.

~~$\Rightarrow h'(x) = (f(x) + g(x))' = f'(x) \cdot g(x) + f(x) \cdot g'(x)$~~

~~$\Rightarrow h'(-1) = f'(-1) \cdot g(-1) + g'(-1) f(-1) \cdot g'(-1) = 0$.~~

~~$h''(x) =$~~

$h'(x) = (f(x) + g(x))' = f'(x) + g'(x) \Rightarrow h'(-1) = 0$

The same way

$h''(x) = f''(x) + g''(x) \Rightarrow h''(1) = 0$

$\Rightarrow h(x) \in W_3$

III. $c \in R$, $f(x) \in W_3$, $g(x) = c \cdot f(x)$

we have $f'(-1) = 0$, $f''(1) = 0$

(See next Page)

(Continuing)

we note that

$g'(x) = (c \cdot f(x))' = c \cdot f'(x) \Rightarrow g'(-1) = c \cdot f'(-1) = 0$

$g''(x) = (c \cdot f'(x))' = c \cdot f''(x) \Rightarrow \cancel{g''(} g''(1) = c \cdot f''(1) = 0$

$\Rightarrow g(x) \in W_3$

Hence W_3 is a linear subspace in V

e) Suppose we have ~~a~~ vector $p = [p_0 \ p_1 \ p_2 \cdots p_n]^T$

$p' = [p_0' \ p_1' \ p_2' \cdots p_n']^T$

So we have inner product $\langle p, p' \rangle$

I. $\langle p, p' \rangle = p^T \cdot p' = p_0 p_0' + p_1 p_1' + \cdots + p_n p_n'$

$\langle p', p \rangle = p'^T \cdot p = p_0 p_0' + p_1 p_1' + \cdots + p_n p_n'$

$\Rightarrow \langle p, p' \rangle = \langle p', p \rangle$

II. $\alpha \in R$

$\langle \alpha p, p' \rangle = (\alpha p^T) \cdot p' = \alpha p_0 p_0' + \alpha p_1 p_1' + \cdots + \alpha p_n p_n'$

$= \alpha \langle p, p' \rangle$

III. for $p_u(x) = u_0 + u_1 x + \cdots + u_n x^n \Rightarrow u = [u_0 \ u_1 \cdots u_n]^T$

$\langle p + u, p' \rangle = p_0 p_0' + p_1 p_1' + \cdots + p_n p_n' + u_0 p_0' + \cdots + u_n p_n'$

$= \langle p, p' \rangle + \langle u, p' \rangle$

So it is an inner product for V

The norm of polynomials $p(x)$ is

$||p(x)|| = \sqrt{\langle p, p \rangle}$

$= \sqrt{p_0^2 + p_1^2 + \cdots + p_n^2}$

Problem 2:

a) Q is real symmetric matrix, so we have:

$$Q = Q^T = Q^* \quad , \quad Q^* \text{ is conjugate transpose of } Q$$

$$\Rightarrow Q^* \cdot Q = Q \cdot Q^* \Rightarrow Q = (U^*)^T \cdot \Lambda \cdot U$$

where Λ is $\begin{bmatrix} p_1 & & & 0 \\ & p_2 & & \\ & & \ddots & \\ 0 & & & p_k \end{bmatrix}$

U is orthogonal matrix whose columns are the eigenvectors of Q

$$\frac{(x^*)^T Q x}{(x^*)^T x} = \frac{(x^*)^T (U^*)^T \Lambda U x}{(x^*)^T x}$$

$$= \frac{((Ux)^*)^T \Lambda (Ux)}{((Ux)^*)^T (Ux)} \quad \text{Suppose } Y = Ux$$

$$= \frac{(Y^*)^T \Lambda Y}{(Y^*)^T Y} = \frac{\sum_{i=1}^{n} p_i y_i^2}{\sum_{i=1}^{n} y_i^2}$$

Since $p_1 \geq p_2 \geq \cdots \geq p_k \Rightarrow p_1 \sum_{i=1}^{n} y_i^2 \geq p_2 \sum_{i=1}^{n} y_i^2 \geq \cdots \geq p_k \sum_{i=1}^{n} y_i^2$

$$\Rightarrow \sum_{i=1}^{n} p_1 y_i^2 \geq \cdots \geq \sum_{i=1}^{n} p_k y_i^2$$

So

$$p_1 \geq \frac{\sum_{i=1}^{n} p_i y_i^2}{\sum_{i=1}^{n} y_i^2} \geq p_k$$

which is $p_1 \geq \dfrac{(x^*)^T Q x}{(x^*)^T x} \geq p_k$

b) Since we have $\lambda_1, \lambda_2, \cdots, \lambda_k$ be eigenvalue and
& $\delta_1, \delta_2, \delta_3, \cdots, \delta_k$ be singular value of A
Assume for each λ_i we have eigenvector X_i such that,

$$A X_i = \lambda_i X_i \Rightarrow (A X_i)^T = X_i^T A^T = \lambda_i X_i^T$$

such that we have:

$$\cancel{X_i^t A^t X_i}$$

$$\frac{X_i^T A^T A X_i}{X_i^T X_i} = \frac{\lambda_i^2 \cdot X_i^T X_i}{X_i^T X_i} = \lambda_i^2$$

~~Suppose~~ ~~Suppose the max and min eigenvalue is λ_{max} and λ_{min}~~
~~vte~~
~~So we have.~~

And we have: eigenvalue of $(A^T A)$ is δ_i^2
From a), assume $B = A^T A$, we have:

$$\delta_1^2 \geq \frac{X_i^T B X_i}{X_i^T X_i} \geq \delta_k^2, \quad \frac{X_i^T B X_i}{X_i^T X_i} = \frac{X_i^T A^T A X_i}{X_i^T X_i} = \lambda_i^2$$

~~$\delta_1 \geq \delta_2 \geq \cdots \geq \delta_k \geq 0$~~

such that:

$$\delta_1^2 \geq \lambda_i^2 \geq \delta_k^2 \quad \text{and} \quad \delta_1 \geq \delta_2 \geq \cdots \geq \delta_k \geq 0$$

$$\Rightarrow \quad \delta_1 \geq |\lambda_i| \geq \delta_k.$$

Considering λ_i be complex numbers. we have:

$$[x_1^* \cdots x_k^*] \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_k \end{bmatrix} = |x_1|^2 + \cdots + |x_k|^2$$ where X_i is eigenvector

and $A X_i^* = \lambda_i^* X_i^* \Rightarrow (A X_i^*)^T = (X_i^*)^T A^T = \lambda_i^* \cdot (X_i^*)^T$

The same way:

$$\frac{(X_i^*)^T A^T A X_i}{(X_i^*)^T X_i} = \lambda_i^* \cdot \lambda_i = |\lambda_i|^2$$

(Next page)

(Continuing)

such that we know that, in the same way, we have

$$\delta_1^2 \geq |\lambda_i|^2 \geq \delta_k^2$$

which is also:

$$\delta_1 \geq |\lambda_i| \geq \delta_k$$

Problem 3.

a). For P, suppose we have eigenvalue λ_i with eigenvector X_i where $i = 1, 2, 3, \ldots\ldots$

So:

$$P X_i = \lambda_i X_i \quad \cancel{\Rightarrow \quad P^2 = P \cdot P X_i}$$

$$\Rightarrow P^2 X_i = P \cdot P X_i = P \cdot \lambda_i X_i = \lambda_i^2 X_i.$$

And $P^2 = P$

such that:

$$P^2 X_i = \lambda_i X_i$$

So we have $\lambda_i^2 X_i = \lambda_i X_i \quad \cancel{\Rightarrow \lambda_i X_i = X_i}$

Since X_i ~~cank~~ cannot be zero vector, So,

$$\lambda_i^2 = \lambda_i \Rightarrow \lambda_i = 0 \text{ or } \lambda_i = 1$$

b). Since $P \cdot I = I \cdot P$

We have

$$(I - P)^2 = I^2 - 2 I \cdot P + P^2$$

And $I^2 = I$, $I \cdot P = P$, $P^2 = P$

So: $(I - P)^2 = I - 2P + P = I - P$

Hence $I - P$ is also a projection.

c) In order to show $X-PX$ and PX are orthogonal, we just need to show that:

$$\langle PX, X-PX \rangle = 0$$

which is:

$$(PX)^T \cdot (X-PX) = 0$$

$\Rightarrow X^T P^T \cdot (I-P) X = 0$ And $P^T = P$, Let $Q = P^T \cdot (I-P)$

So: $Q = P \cdot (I-P) = P - P^2$ And $P^2 = P$

$\Rightarrow Q = P - P^2 = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}$

So $X^T \cdot P^T \cdot (I-P) X = X^T \cdot Q \cdot X = 0$

That is $X-PX$ and PX are orthogonal.

d) Show $P = A(B^T A)^{-1} B^T$ is projection.

Proof: $P^2 = A(B^T A)^{-1} B^T \cdot \cancel{A(B^T A^{-1}) B^T} \left(A(B^T A)^{-1} B^T\right)$

$= A(B^T A)^{-1} \cdot \left((B^T A) \cdot (B^T A)^{-1}\right) B^T$

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

Hence P is projection.

By definition ~~of~~ and properties of projection on Vector Space

Suppose we have vector space $V \to R^m$, subspace R^n

So m, and n must satisfy that:

$$m \geq n$$

And in order to get P, $B^T \cdot A$ must be invertible.

(Next Page)

when P is orthogonal projection, we have:

$$P^T = P$$

$$\Rightarrow P^T = \left(A(B^TA)^{-1}B^T\right)^T = B\cdot\left((B^TA)^T\right)^{-1}\cdot A^T$$

$$= B\cdot(A^T\cdot B)^{-1}\cdot A^T$$

if $P^T = P$, we have:

$$B\cdot(A^T\cdot B)^{-1}\cdot A^T = A\cdot(B^TA)^{-1}\cdot B^T \qquad (1)$$

~~we have basic orthogonal projection that is~~

$$\cancel{P = A'(A'^T\cdot A')^{-1}A'^T}$$

By definition, suppose we have vector x, y, $y = Px$

so define inner product on a positive definite matrix we have:

$$\langle x, y\rangle_{D} = y^TD\cdot x \quad \text{and } D = \begin{bmatrix} d_1 & & 0 \\ & d_2 \ldots & \\ 0 & & d_m \end{bmatrix}_{m\times m}$$

And P is given by $\quad Px = \underset{y\in \text{range}(A)}{\operatorname{argmin}} \|x-y\|_D^2$

where

$$P = A(A^TDA)^{-1}A^TD. \qquad (2)$$

Suppose $B = D\cdot A$, Combine (1) and (2) we have

if $B = D\cdot A$ and $D\cdot A' = A'\cdot D$ for all A'

$$P^T = B\cdot(A^T\cdot B)^{-1}A^T = DA\cdot(A^TDA)^{-1}\cdot A^T = A\cdot(A^TDA)^{-1}\cdot A^TD$$

$$P = A\cdot(A^TD\cdot A)^{-1}A^TD$$

$\Rightarrow$ such that $P = P^T$

if and only if $B = D\cdot A$, $D\cdot M = M\cdot D$ for all matrix M

That is vector in A is orthogonal to null space, which is the same to vector in B.

~~Problem 4.~~

e) First we show that the optimum $\hat{b}$ is the orthogonal projection of b.

Suppose we have $\hat{b} = A \cdot x$, A is size of $m \times n$

So we know that columns of A is the basis of column space $COL(A)$, regard x as scalar vectors

For all x, $A \cdot x$ denotes all vectors in the space $COL(A)$

Suppose we have optimum $\hat{b}$ as $\hat{b}_m$ and to the contrary that there is an element $\hat{b}_i \in COL(A)$ which is not orthogonal to the error $b - \hat{b}_m$. Without loss of generality we may assume that $\|\hat{b}_i\| = 1$ and that $\langle b - \hat{b}_m, \hat{b}_i \rangle = \delta \neq 0$.

Define $m \in COL(A)$, $m \equiv \hat{b}_m + \delta \hat{b}_i$

$$\|b - m\|^2 = \|b - \hat{b}_m - \delta \hat{b}_i\|^2$$

$$= \|b - \hat{b}_m\|^2 - \langle b - \hat{b}_m, \delta \hat{b}_i \rangle - \langle \delta \hat{b}_i, b - \hat{b}_m \rangle + |\delta|^2$$

$$= \|b - \hat{b}_m\|^2 - |\delta|^2 < \|b - \hat{b}_m\|^2$$

So we have contradiction: $\hat{b}_m$ is not in $COL(A)$ that have minimum error since $\|b - m\|^2$ is smaller.

Hence we know
$b - \hat{b}_m$ is orthogonal to $COL(A)$

That is: $\hat{b}_m$ is the orthogonal projection of b

(Next Page)

(Continuing)

Next, we want to get projection matrix P.

Since we want

$$\min_{\hat{b}} \|b-\hat{b}\|^2$$

$\Rightarrow$ which is

$$\min_{\hat{x}} \|b-A\hat{x}\|^2, \quad b-A\hat{x} \text{ is orthogonal to } COL(A)$$

$\Rightarrow$ $b-A\hat{x}$ is orthogonal to all column vectors of A

~~denote~~

That is we have:

$$A^T(b-A\hat{x})=0 \quad \text{where we get } \cancel{\text{minimum } \hat{x}} \text{ optimum } \hat{x}.$$

$$\Rightarrow \quad A^T \cdot A\hat{x} = A^T \cdot b$$

$$\Rightarrow \quad \hat{x} = (A^TA)^{-1} \cdot A^T b \quad \text{and} \quad \cancel{\hat{x}}\ A\hat{x} = \hat{b}$$

$$\Rightarrow \quad \hat{b} = A \cdot \hat{x} = A(A^TA)^{-1}A^T b$$

Hence Projection matrix

$$P = A \cdot (A^TA)^{-1}A^T$$

Problem 4

a): (detailed $\hat{x}_*$ of problem a) is given in b) in next page)

Considering $x - \hat{x}$

By Orthogonality Principle we have:

$$\langle x - \hat{x}_*, w \rangle = 0$$
$$\langle x - \hat{x}_*, z \rangle = 0 \quad , \quad \hat{x}_* = a_* z + b_* w$$

$$\Rightarrow \langle x - a_* z - b_* w, w \rangle = 0 \qquad (1)$$
$$\langle x - a_* z - b_* w, z \rangle = 0 \qquad (2)$$

Hence we have

$$\langle x, w \rangle - \langle a_* z + b_* w, w \rangle = 0$$
$$\Rightarrow \langle x, w \rangle = a_* \langle z, w \rangle + b_* \langle w, w \rangle \qquad (3)$$

the same way

$$\langle x, z \rangle = a_* \langle z, z \rangle + b_* \langle w, z \rangle \qquad (4)$$

From (3) (4) we have

$$\begin{bmatrix} \langle x, w \rangle \\ \langle x, z \rangle \end{bmatrix} = \begin{bmatrix} \langle z, w \rangle & \langle w, w \rangle \\ \langle z, z \rangle & \langle w, z \rangle \end{bmatrix} \cdot \begin{bmatrix} a_* \\ b_* \end{bmatrix}$$

which is:

$$\begin{bmatrix} E[xw] \\ E[xz] \end{bmatrix} = \underbrace{\begin{bmatrix} E[zw] & E[w^2] \\ E[z^2] & E[wz] \end{bmatrix}}_{\text{denote as } A} \begin{bmatrix} a_* \\ b_* \end{bmatrix}$$

$\Rightarrow$ So we get $\hat{x}_*$

b): So $|A| = (E[zw])^2 - E[w^2] \cdot E[z^2] > 0$ (By Binomial Theorem)

So $$A^{-1} = \begin{bmatrix} \frac{E[zw]}{|A|} & -\frac{E[z^2]}{|A|} \\ \frac{-E[w^2]}{|A|} & \frac{E[zw]}{|A|} \end{bmatrix}$$

Such that we know

$$A^{-1} \cdot \begin{bmatrix} E[xw] \\ E[xz] \end{bmatrix} = \begin{bmatrix} a_* \\ b_* \end{bmatrix}$$

(Next Page)

(Continuing)

~~sub~~ Hence

$$a_* = \frac{E[zw]\cdot E[xw] - E[xz]\cdot E[z^2]}{(E[zw])^2 - E[z^2]\cdot E[w^2]} \quad (5)$$

$$b_* = \frac{E[zw]\cdot E[xz] - E[xw]\cdot E[w^2]}{(E[zw])^2 - E[z^2]\cdot E[w^2]} \quad (6)$$

Hence: ~~$\hat{x}_*$~~ ~~Suppose~~

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

The optimum distance is

$$\min \|x-\hat{x}\| = \|x-\hat{x}_*\| = \sqrt{\langle x-\hat{x}_*, x-\hat{x}_*\rangle}$$

And we know $x-\hat{x}_*$ is orthogonal to $\hat{x}_* \Rightarrow \langle \hat{x}_*, x-\hat{x}_*\rangle = 0$

So distance:

$$d = \sqrt{\langle x-\hat{x}_*, x-\hat{x}_*\rangle} = \sqrt{\langle x, x-\hat{x}_*\rangle} = \sqrt{\langle x,x\rangle - \langle x,\hat{x}_*\rangle}$$

$$= \sqrt{E[x^2] - \langle x, a_* z\rangle - \langle x, b_* w\rangle} \qquad \text{And } a_*, b_* \in R$$

$$= \sqrt{E[x^2] - a_*\cdot E[xz] - b_*\cdot E[xw]}$$

where a_* and b_* is represent in (5) and (6)

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