Name: AYUSH SAHNI
Roll no.: 21111019

Quiz 5

1(a) Minimize $\|AX - C\|_F$ {$\|\cdot\|_F$ denotes Frobenius norm}
subject to $\exists Y \in \mathbb{R}^{d\times q}$ s.t. $X = YB$

, $A \in \mathbb{R}^{n\times d}$, $X \in \mathbb{R}^{d\times p}$, $C \in \mathbb{R}^{n\times p}$, $B \in \mathbb{R}^{q\times p}$

$\Rightarrow$ Minimise $\|AYB - C\|_F$

$X = (A^-)C$ {A^- = pseudoinverse A}
$Y = (A^-)C(B^-)$

Why pseudoinverse gives correct answer to minimisation problem?
Consider the problem minimize $\|Ax - b\|$.
I will show that $x_0 = (A^-)b$ minimises $\|Ax - b\|$.

$Ax - b = Ax \underbrace{- AA^-b + AA^-b}_{\text{adding \& subtracting } AA^-b} - b$

$= A(x - A^-b) + (I - AA^-)(-b)$

So, $\|Ax - b\|^2 = \|A(x - A^-b) + (I - AA^-)(-b)\|^2$
$= \|A(x - A^-b)\|^2 + \|(I - AA^-)(-b)\|^2 + 2(A(x - A^-b))^T((I - AA^-)(-b))$

Orthogonal vectors

$\Rightarrow \|Ax - b\|^2 = \|A(x - A^-b)\|^2 + \|(I - AA^-)(-b)\|^2 + 0$
$= \|A(x - x_0)\|^2 + \|Ax_0 - b\|^2$
$\geq \|Ax_0 - b\|^2$ {hence x_0 minimizes $\|Ax - b\|$}

Note that A is tall and thin matrix.
Doing the same thing for each column of $X = [x_1, x_2, \dots x_p]$
& $C = [c_1, c_2, \dots c_p]$ we can show that $X = (A^-)C$
minimises $\|AX - C\|$

Date
Page

Putting $Y = A^{-}CB^{-}$ in $\|AYB - C\|_F$:-

$\|A(A^{-}CB^{-})B - C\|_F$

Substituting $A = U_A \Sigma_A V_A^T$ & $B = U_B \Sigma_B V_B^T$

$= \|(U_A \Sigma_A V_A^T)(U_A \Sigma_A V_A^T)^{-} C (U_B \Sigma_B V_B^T)^{-}(U_B \Sigma_B V_B^T) - C\|_F$

$= \|(U_A U_A^T) C (V_B V_B^T) - C\|_F$

$U_A U_A^T$ & $V_B V_B^T$ are projector matrices. This means we are measuring distance of C from its projection on space. This distance between C and its projection on space will be the minimum distance between C & the space.

Hence, $Y = A^{-}CB^{-}$ minimises $\|AYB - C\|_F$

1 (b) $\min\limits_{Y \in R^{d \times q}} \|AYB - C\|_F$

AY → the columns of this matrix are in the column space of A

YB → the rows of this matrix are in the row space of B.

Notations used :-

$X^* = A^{-}C$

$C^* = AA^{-}C - C = (I - AA^{-})(-C)$

$Y^* = A^{-}CB^{-}$

$B^* = AY^*B - AX^* = -AA^{-}C(I - B^{-}B)$

Let S be $m \times n$ matrix such that each element of S is sampled from D_1. Let R be $p \times l$ matrix such that each element of R is sampled from distribution D_2

~~Let S be~~ Consider the events:-

E_A^{subemb} : Under D_1, column space of A is approx. preserved.

$$P[\forall x \in R, \|SAx\| \in (1 \pm \epsilon/20)\|Ax\|] \geq 1 - \frac{\delta}{20}$$

E_B^{subemb} : Under D_2, row space of B is approx. preserved.

$$P[\forall y^T \in R^q ; \|y^T BR\| \in (1 \pm \frac{\epsilon}{20}) \|y^T B\|] \geq 1 - \frac{\delta}{20}$$

$$E_A^{ip-subsp} : P_{S|R}\{\forall Q \in R^{d \times p}, |(S(AQ - AX^*)R)_j^T S(AX^*R - CR)_j - ((AYB - AX^*)R_j)^T (AY^* - C)R_j| \leq \frac{\epsilon}{20} \|(AYB - AX^*)R_j\| \|(AX^* - C)R_j\|\} \geq 1 - \frac{\delta}{20}$$

$$E_B^{ip-subsp} : P_{R|S}\{\forall Q \in R^q, |Q^T BR(SB^*)_i R^T - Q^T B(SB^*)_i| \leq \left(\frac{\epsilon}{20}\right)\|Q^T B\|_2 \|(SB^*)_i\|_2\} \geq 1 - \frac{\delta}{20}$$

$$E_{C^*}^{frob} : P\{\|C^*R\|_F \in (1 \pm \frac{\epsilon}{20})\|C^*\|_F\} \geq 1 - \frac{\delta}{20}$$

$$E_{B^*}^{frob} : P\{\|B^*R\|_F \in (1 \pm \frac{\epsilon}{20})\|B^*\|_F\} \geq 1 - \frac{\delta}{20}$$

$$E_B^{Aff-emb} : P[\forall Q \in R^{n \times q}, \|QBR - AX^*R\|_F \in (1 \pm \epsilon)\|QB - AX^*\|_F] \geq 1 - \frac{\delta}{20}$$

I will assume $\epsilon \leq \frac{1}{4}$.

Now,

$$\|SAYBR - SCR\|_F^2$$

$$= \|SA(Y - Y^*)BR\|_F^2 + \|SA(Y^*B - X^*R)\|_F^2 + \|SAX^*R - SCR\|_F^2$$
$$+ 2 \operatorname{tr} SA(Y - Y^*)BR(SA(Y^*B - X^*)R)^T$$
$$+ 2 \operatorname{tr}(S(AYB - AX^*)R)^T(SAX^*R - SCR) \quad \text{— (1)}$$

, tr denotes trace of matrix.

Proof : $\|SAYBR - SCR\|_F^2$

$$= \|S(AYB - AX^*)R + SAX^*R - SCR\|_F^2$$
$$= \|S(AYB - AX^*)R\|_F^2 + \|SAX^*R - SCR\|_F^2$$
$$+ 2 \operatorname{tr}(S(AYB - AX^*)R)^T(SAX^*R - SCR)$$

$$\|SA(Y^*B - X^*R\|_F^2$$

$$= \cancel{\|SA(Y} \; \|SA(Y-Y^*)BR\|_F^2 + \|SA(Y^*B - X^*)R)^T$$
$$+ 2\,tr\, SA(Y-Y^*)BR(SA(Y^*B - X^*)R)^T (SAX^*R - S(R)$$
$$+ \|SAX^*R - SCR\|_F^2 + 2\,tr\,(S(AYB - AX^*)R)^T (SAX^*R - S(R))$$

$$= \|SA(Y-Y^*)BR\|_F^2 + \|SA(Y^*B - X^*R)\|_F^2 + \|SAX^*R - S(R\|_F^2$$
$$+ 2\,tr\, SA(Y-Y^*)BR(SA(Y^*B - X^*)R)^T.$$
$$+ 2\,tr\,(S(AYB - AX^*)R)^T(SAX^*R - S(R))$$

Now, I will find bound on each term of this eqn.

step 1 $\|SA(Y-Y^*)BR\|_F^2 \in \left(1 \pm \frac{\epsilon}{20}\right)^2 \|A(Y-Y^*)BR\|_F^2$ {if event E_A^{subemb} holds}

$\in \left(1 \pm \frac{\epsilon}{20}\right)^4 \|A(Y-Y^*)B\|_F^2$ {if event E_B^{subemb} holds

$\in (1 \pm \epsilon)\|A(Y-Y^*)B\|_F^2$

step 2 If Events $E_A^{ip\text{-}subsp}$ & $E_{C^*}^{frob}$ holds:-

$$\forall\, Y \in R^{d \times q},\ |tr(S(A\cancel{Y}B - AX^*)R)^T(SAX^*R - SCR)|$$
$$\leq \frac{\epsilon}{10}\|(\cancel{A}AYB - AX^*)\|_F \|C^*\|_F$$

This is because,

$$tr(S(AYB - AX^*)R)^T(SAX^*R - SCR)$$
$$= \sum_{j=1}^{l} (S(AYB - AX^*)R)_j^T\, S(AX^*R - CR)_j$$

So, by triangle inequality

$$|tr(S(AYB - AX^*)R)^T(SAX^*R - SCR)$$
$$\leq \sum_{j=1}^{l} |(S(AYB - AX^*)R)_j^T\, S(AX^*R - CR)_j$$

$$\leq \sum_j |S(AYB - AX^*)R)_j^T\, S(AX^*R - CR)_j$$
$$- ((AYB - AX^*)R_j)^T(AX^* - C)R_j|$$

$$+ \sum_{j=1}^{l} |((AYB - AX^*)R_j)^T(AX^* - C)R_j| \quad \{\text{triangle ineq.}\}$$

$$\leq \frac{\epsilon}{20}\left(\sum_j \|(AXB - AX^*)R_j\|_F^2\right)^{1/2}\left(\sum_{j=1}^{l} \|(AX^* - C)R_j\|_F^2\right)^{1/2}$$

{Cauchy-swartz ineq}

$$= \frac{\epsilon}{20}\|(A\cancel{X}B - AX^*)R\|_F \|C^*R\|_F$$

Now using $E_B^{Aff\text{-}emb}$, $E_{C^*}^{frob}$ & $AX^* - C = C^*$, we will get

$$|\text{tr}(S(AYB - AX^*)R)^T(SAX^*R - S(R))|$$

$$\leq \frac{\epsilon}{20}\left(1+\frac{\epsilon}{20}\right)\|(AYB - AX^*)\|_F \|C^*\|_F$$

$$\leq \left(\frac{\epsilon}{10}\right)\|AYB - AX^*\|_F \|C^*\|_F$$

Step 3 Analysing term $\text{tr}\, SA(Y-Y^*)BR(SA(Y^*B - X^*)R)^T$
If events $E_B^{ip\text{-}subsp}$, E_A^{subemb} holds then

$$|\text{tr}\, SA(Y-Y^*)BR(SA(Y^*B - X^*)R)^T|$$

$$= |\text{tr}\, SA(Y-Y^*)BR(SA(Y^*B - X^*)R)^T - \text{tr}\, SA(Y-Y^*)B(SA(Y^*B - X^*))|$$

$$\leq \sum_i \left(\frac{\epsilon}{20}\right)\|(SA(Y-Y^*)B)_i\| \, \|SA(Y^*B - X^*)_i\|$$

{if E_B^{subemb} holds}

$$\leq \frac{\epsilon}{20}\|SA(Y-Y^*)B\|_F \|SA(Y^*B - X^*)\|_F$$

If we do furthur analysis and assume E_A^{subemb} holds then,

$$\leq \frac{\epsilon}{10}\|A(Y-Y^*)B\|_F \|A(Y^*B - X^*)\|_F$$

Putting these bounds in eqn (1) :-

$\|SAYBR - SCR\|_F^2$

$= \left(1 \pm \frac{\epsilon}{10}\right)\|A(Y-Y^*)B\|_F^2 + \left(1 \pm \frac{\epsilon}{10}\right)\|AB^*\|_F^2 + \left(1 \pm \frac{\epsilon}{10}\right)\|C^*\|_F^2$

$+ 2\left(\frac{\epsilon}{16}\right)\|AYB - AX^*\|_F \|C^*\|_F + 2\left(\frac{\epsilon}{16}\right)\|A(Y-Y^*)B\|_F \|AB^*\|_F$

$= \left(1 \pm \frac{\epsilon}{10}\right)\left[\|A(Y-Y^*)B\|_F^2 + \|AB^*\|_F^2 + \|C^*\|_F^2\right]$

$+ \frac{\epsilon}{16}\left[\|AYB - AX^*\|_F^2 + \|C^*\|_F^2 + \|C^*\|_F^2 + \|A(Y-Y^*)B\|_F^2 + \|AB^*\|_F^2\right]$

Now,

$\left|\|SAYBR - SCR\|_F^2 - \|AYB - C\|_F^2\right|$

$\leq \frac{\epsilon}{10}\left[\|A(Y-Y^*)B\|_F^2 + \|AB^*\|_F^2 + \|C^*\|_F^2\right]$

$+ \frac{\epsilon}{16}\left[\|AYB - AX^*\|_F^2 + \|C^*\|_F^2 + \|A(Y-Y^*)B\|_F^2 + \|AB^*\|_F^2\right]$

$= \left(\frac{\epsilon}{10} + \frac{\epsilon}{16}\right)\|AYB - C\|_F^2 + \frac{\epsilon}{16}\|AYB - C\|_F^2$

$\leq \left(\frac{\epsilon}{4}\right)\|AYB - C\|_F^2$

$\hat{Y} = (SA)^-(SCR)(BR)^-$

$\underbrace{\hat{Y}B}_{\hat{X}} = (SA)^-(SCR)(R)^-$

1(c) Minimise $\|AX - C\|_F$ = Minimise $\|AYB - C\|_F$

We don't need to preserve the whole column space of [A C]. ~~A is long matrix with random matrix but not~~

~~We preserve the column space of~~ A
~~Condition~~

$$\|SAYBR - SCR\|_F^2$$

Condition of S

S is an $m \times n$ matrix sampled from distribution D_1.
It should preserve whole column space of A
By γ-net argument, no. of vectors in the col. space of A $\leq 5^d$ {$\because$ A is $n \times d$ matrix}

S should also preserve the norms of all vectors in C (columns of C).
no. of columns in C = p

So, $m = O\left(\frac{1}{\varepsilon^2} \log \frac{5^d}{\delta} + \frac{1}{\varepsilon^2} \log \frac{p}{\delta}\right)$

$= O\left(\frac{d}{\varepsilon^2} \log \frac{1}{\delta} + \frac{1}{\varepsilon^2} \log \frac{p}{\delta}\right)$

Condition on R

R is an $p \times l$ matrix sampled from distribution D_2
It should preserve the row space of B.
no. of vectors in row space of B $\leq 5^q$
{$\because$ B is $q \times p$ matrix}

R should also preserve the norms of all rows in C.
no. of rows in C = n

So, $l = O\left(\frac{1}{\varepsilon^2}\log\frac{5^q}{\delta} + \frac{1}{\varepsilon^2}\log\frac{n}{\delta}\right)$

$= O\left(\frac{q}{\varepsilon^2}\log\frac{1}{\delta} + \frac{1}{\varepsilon^2}\log\frac{n}{\delta}\right)$

Provided,
E_A^{subemb}, E_B^{subemb}, $E_A^{ip-subsp}$, $E_B^{ip-subsp}$, $E_{C^*}^{Frob}$, $E_{B^*}^{Frob}$

and $E_B^{Aff-Emb}$ events hold. $S_{ij} \sim N(0, \frac{1}{m})$

$R_{ij} \sim N(0, \frac{1}{l})$