Ilhar Kdayev — MA8473 Final Exam — ID: 20190788

1) $a \in \mathbb{R}^{m\times 1}$, $b \in \mathbb{R}^{n\times 1}$, $B \in \mathbb{R}^{m\times n}$

We have to first evaluate the derivative $\frac{\partial}{\partial B} a^T B b = \frac{\partial}{\partial B_{ij}} a^T B b$ and find shorter expression for that

Since the scalar resulting from the product $a^T B b$ is given by

$$a^T B b = \sum_{j=1}^{m} \sum_{k=1}^{n} b_{jk} a_j b_k \text{ where } B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ & & \vdots & \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

If we take the differentiation wrt B_{ij} element

$$\frac{\partial}{\partial B_{ij}} (a^T B b) = \frac{\partial}{\partial B_{ij}} \left(\sum_{j=1}^{m} \sum_{k=1}^{n} b_{jk} a_j b_k \right) = a_i b_j, \text{ with } B_{ij} = b_{ij}$$

(i,j^{th} element of B being b_{ij})

Since the other b_{jk}'s do not depend on b_{ij}, it's clear that derivative wrt ij^{th} element of B will be $\frac{\partial}{\partial B_{ij}} \left(\sum_{j=1}^{m} \sum_{k=1}^{n} b_{jk} a_j b_k \right) = a_i b_j$ and constructing the matrix

$$a b^T = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} \begin{bmatrix} b_1 & b_2 & \dots & b_n \end{bmatrix} = \begin{bmatrix} a_1 b_1 & a_1 b_2 & \dots & a_1 b_n \\ a_2 b_1 & a_2 b_2 & \dots & a_2 b_n \\ \vdots & \vdots & & \vdots \\ a_m b_1 & a_m b_2 & & a_m b_n \end{bmatrix}$$

meaning derivative wrt ij^{th} elem. of B will be ij^{th} element of matrix multiplication ab^T, where $(ab^T)_{ij} = a_i b_j$; Thus,

$$\boxed{\frac{\partial}{\partial B} a^T B b = \frac{\partial}{\partial B_{ij}} (a^T B b) = a b^T}$$

b) $a \in \mathbb{R}^{m\times 1}$, $b \in \mathbb{R}^{n\times 1}$, $B \in \mathbb{R}^{n\times m}$

$$B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ & & \vdots & \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix}, \quad a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$Ba = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ & & \vdots & \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix} = \begin{bmatrix} b_{11}a_1 + b_{12}a_2 + \dots + b_{1m}a_m \\ \vdots \\ b_{n1}a_1 + b_{n2}a_2 + \dots + b_{nm}a_m \end{bmatrix}$$

b) $L = \frac{1}{n} \sum_{i=1} \|x_i - x_{i,\epsilon}\|^2 = \frac{1}{n} \sum_{i=1} (x_i - x_{i,\epsilon})^T (x_i - x_{i,\epsilon})$, considering from given conditions

$$x_{i,\epsilon} = U y_i = U W x_i \Rightarrow L = \frac{1}{n} \sum_{i=1}^{n} (x_i - UWx_i)^T (x_i - UWx_i) = \frac{1}{n} \sum_{i=1}^{n} (x_i^T - x_i^T W^T U^T) \cdot (x_i - UWx_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(x_i^T x_i - x_i^T W^T U^T x_i - x_i^T U W x_i + x_i^T W^T U^T U W x_i \right).$$

Taking derivative wrt matrix U

$$\frac{\partial L}{\partial U} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\partial}{\partial U} \left(-x_i^T W^T U^T x_i \right) - \frac{\partial}{\partial U} \left(x_i^T U W x_i \right) + \frac{\partial}{\partial U} \left(x_i^T W^T U^T U W x_i \right) \right)$$

$$\frac{\partial L}{\partial U} = \frac{1}{n} \sum_{i=1}^{n} \left(-\frac{\partial}{\partial U} \left(x_i^T W^T U^T x_i \right) - \frac{\partial}{\partial U} \left(x_i^T U W x_i \right) + \frac{\partial}{\partial U} \left(x_i^T W^T U^T U W x_i \right) \right)$$

From the previous problem, $\frac{\partial}{\partial U} (x_i^T U W x_i) = \frac{\partial}{\partial U} (x_i^T U (W x_i)) = x_i (W x_i)^T$ $= x_i x_i^T W^T$. Similarly, $\frac{\partial}{\partial U} (x_i^T W^T U^T U W x_i) = \frac{\partial}{\partial U} \left((W x_i)^T U^T U (W x_i) \right) =$

$$= 2U(Wx_i)(Wx_i)^T = 2U(Wx_i)x_i^T W^T$$

Claim: $\frac{\partial}{\partial X} (a^T X^T b) = b a^T$

Proof: $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$, $X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ & \vdots & & \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$, $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$, plugging in back the results, $a^T X^T b =$

$$= \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{m1} \\ x_{12} & & x_{m2} \\ \vdots & & \vdots \\ x_{1n} & & x_{mn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} \underbrace{a_1 x_{11} + a_2 x_{12} + \dots + a_n x_{1n}}_{c_1} & \dots & \underbrace{a_1 x_{m1} + a_2 x_{m2} + \dots + a_n x_{mn}}_{c_m} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} =$$

$$= \begin{bmatrix} c_1 & \dots & c_m \end{bmatrix} \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix} = c_1 b_1 + \dots + c_m b_m = (a_1 x_{11} + \dots + a_n x_{1n}) b_1 + \dots + (a_1 x_{m1} + \dots + a_n x_{mn}) \cdot b_m$$

Scanned with CamScanner

$a^TB^TBa = (Ba)^TBa = (b_{11}a_1 + b_{12}a_2 + \dots + b_{1m}a_m)^2 + \dots + (b_{n1}a_1 + b_{n2}a_2 + \dots + b_{nm}a_m)^2$ since Ba - column vector. Now, taking derivative wrt b_{ij}, if we consider the terms containing b_{ij}, $(Ba)^TBa = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} b_{ij}a_j\right)^2$

$$\frac{\partial}{\partial B_{ij}}(Ba)^TBa = 2\left(\sum_{j=1}^{m} b_{ij}a_j\right)\frac{\partial}{\partial B_{ij}}(b_{ij}a_j) = \left(2\sum_{j=1}^{m} b_{ij}a_j\right)\cdot a_j$$ since no other elements depend on b_{ij}

In the end, derivative of $(Ba)^TBa$ wrt B can be found as in the following:

$\frac{\partial}{\partial B_{ij}}(Ba)^TBa = (2a_j)\sum_{j=1}^{m} b_{ij}a_j$, meaning ij^{th} element of derivative will be $2\left(\sum_{j=1}^{m} b_{ij}a_j\right)a_j = 2(b_{i1}a_1 + \dots + b_{im}a_m)a_j$. If we take the matrix multiplication Baa^T,

$$Baa^T = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1m} \\ b_{21} & b_{22} & \dots & b_{2m} \\ & \vdots & & \\ b_{n1} & b_{n2} & \dots & b_{nm} \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}\begin{bmatrix} a_1 & a_2 & \dots & a_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} b_{ij}a_j \\ \vdots \\ \sum_{j=1}^{m} b_{nj}a_j \end{bmatrix}\begin{bmatrix} a_1 & a_2 & \dots & a_m \end{bmatrix}$$

$$= \begin{bmatrix} \left(\sum_{j=1}^{m} b_{ij}a_j\right)a_1 & \dots & \left(\sum_{j=1}^{m} b_{1j}a_j\right)a_m \\ \vdots & & \vdots \\ \left(\sum_{j=1}^{m} b_{nj}a_j\right)a_1 & & \left(\sum_{j=1}^{m} b_{nj}a_j\right)a_m \end{bmatrix}$$

It's easily seen that ij^{th} elem. of Baa^T is $\left(\sum_{j=1}^{m} b_{ij}a_j\right)a_j$. Thus, ij^{th} element of derivative where derivative is taken as $\frac{\partial}{\partial B_{ij}}(Ba)^TBa$ will be $2\left(\sum_{j=1}^{m} b_{ij}a_j\right)a_j$ which is also ij^{th} elem of $2\cdot Baa^T$; Thus,

$$\boxed{\frac{\partial}{\partial B_{ij}} a^TB^TBa = \frac{\partial}{\partial B_{ij}}(Ba)^TBa = 2Baa^T}$$ ✓.

$a^TX^Tb = (a_1x_{11} + \dots + a_nx_{1n})b_1 + \dots + (a_1x_{m1} + \dots + a_nx_{mn})b_m$, taking derivative wrt ij^{th} element of $X \Rightarrow a^TX^Tb = \sum_{i=1}^{m}\sum_{j=1}^{n}(a_jx_{ij})b_i$

$a^TX^Tb = \sum_{k=1}^{m}\sum_{p=1}^{n}(a_px_{kp})b_k$ and $\boxed{\frac{\partial}{\partial x_{ij}} a^TX^Tb = a_jb_i}$ Taking the matrix multiplication ba^T,

$$ba^T = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}\begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix} = \begin{bmatrix} b_1a_1 & \dots & b_1a_n \\ b_2a_1 & & b_2a_n \\ \vdots & & \vdots \\ b_ma_1 & & b_ma_n \end{bmatrix},$$

ij^{th} element of ba^T is b_ia_j. In conclusion, derivative of a^TX^Tb wrt X where ij^{th} element of derivative $= b_ia_j$, where ij^{th} element of ba^T is $b_ia_j \Rightarrow \boxed{\frac{\partial}{\partial X} a^TX^Tb = ba^T}$ ✓

Using this claim, $\frac{\partial}{\partial U}(x_i^TW^TU^Tx_i) = \frac{\partial}{\partial U}((Wx_i)^TU^Tx_i) = x_i(Wx_i)^T = x_ix_i^TW^T \Rightarrow \frac{\partial L}{\partial U} = \frac{1}{n}\sum_{i=1}^{n}(-x_ix_i^TW^T - x_ix_i^TW^T + 2UWx_ix_i^TW^T)$

$$= \frac{1}{n}\sum_{i=1}^{n}(2UWx_ix_i^TW^T - 2x_ix_i^TW^T) = \frac{2}{n}\sum_{i=1}^{n}(UWx_ix_i^TW^T - x_ix_i^TW^T)$$

Assume that $\Sigma = \frac{1}{n}\sum_{i=1}^{n} x_ix_i^T$, then making derivative to be zero,

$\frac{\partial L}{\partial U} = 0 \Rightarrow \sum_{i=1}^{n} UWx_ix_i^TW^T = \sum_{i=1}^{n} x_ix_i^TW^T$, or in other words $\Rightarrow$ $UW\Sigma W^T = \Sigma W^T$, meaning $\boxed{U = (\Sigma W^T)(W\Sigma W^T)^{-1}}$ ✓

Also, $\left(\frac{\partial L}{\partial U}\right)'$ second derivative > 0 $\boxed{\text{derivative of } \frac{\partial L}{\partial U} > 0}$