Nicole Fleming
A02079019

STAT 5810

September 19, 2020

Homework 1

1) Show that for any n × n invertible matrices A and B, the following properties are true:
(a) $(A^{-1})^T = (A^T)^{-1}$
(b) $A^{-1}B^{-1} = (BA)^{-1}$

1a) Consider matrix A where $A\epsilon\ R^{nxn}$ invertible matrix. The inverse of any matrix is the determinant multiplied by The identity matrix of the matrix's dimensions, i.e.

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \text{ then } A^{-1} = \frac{1}{a_{11}a_{22}-a_{21}a_{12}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

When taking the inverse of the Identity matrix the math is simply $1 * I^{nxn}$, or the Identity matrix.

The transpose of a matrix is a reordering of the matrix, making the columns become the rows, and the rows become the columns. For example in a R^{2x2} the Transpose of A is as follows

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \text{ then } A^T = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix}.$$

The transpose of the Identity matrix of any nxn dimension is the identity matrix. This is due to the fact that is is a pure symmetric matrix.

The statement asks for the inverse of A or the inverse of A^T. When the inverse is taken for any R^{nxn} matrix, as defined above, it is the following, where det is the determinant, and I_{nxn} is the square matrix.

$$\frac{1}{det}(I_{nxn})$$

Using this knowledge, the scalar, $\frac{1}{det}$ is just being multiplied into the Identity matrix. The transpose of the Identity matrix as stated above is the Identity matrix for any I_{nxn} matrix. Therefore, the statement is true since, $(A^{-1})^T$ results in a scalar multiplied by I^T, where $(A^T)^{-1}$ is reorganizing rows by column, and then taking the inverse. However, the determinant will be the same and the transpose of I is still I. Therefore the statement is equal. Thus, this is true.

1b) Suppose $A, B\epsilon R^{nxn}$ and are invertible, this is to say A, B cannot be populated with zeros, or a matrix that is non-invertible.
Let C =$(AB)^{-1}$. Therefore, it follows that

$$A * B = C^{-1} \text{ or } A * B = \frac{1}{C}.$$

After solving for C it leaves the following

$$\frac{1}{A} * \frac{1}{B} = C \text{ or } A^{-1} * B^{-1} = C.$$

Substituting C for the assinged value, the result is

$$A^{-1} * B^{-1} =(AB)^{-1}.$$

---

2) Derive properties (5.29) and (5.30) from the book using Definition 5.2 (definition of the derivative).

a) The Product Rule: (fg)' = fg' + f'g
Starting with the definition of the derivative given in definition 5.2, $\lim_{h \to 0} \frac{fx+h-fx}{h}$.
To get something more usable, add and subtract zero in the form $f(x+h)g(x)$. The argument then becomes

$$\lim_{h \to 0} \frac{fx+hgx+h-fx+hgx+fx+hgx-fxgx}{h}$$
$$= \lim_{h \to 0} \frac{fx+h*<gx+h-gx>}{h} + \lim_{h \to 0} \frac{gx*<fx+h-fx>}{h} **$$

Using the properties of limits,

$$\lim_{h \to 0} fx+h * \lim_{h \to 0} \frac{gx+h-gx}{h} + \lim_{h \to 0} gx * \frac{fx+h-fx}{h}.$$

With the following assignments,

g(x) = g(x)

g'(x) = $\lim_{h \to 0} \frac{gx+h-gx}{h}$

f(x) = $\lim_{h \to 0} fx+h$

f'(x) = $\lim_{h \to 0} \frac{fx+h-fx}{h}$

Which makes the expression become

f(x)*g'(x) + f'(x)*g(x).

**forgive my use of ¡¿ as parenthesis, actual parenthesis were not showing up in the math mode in Latex.

---

b) The Quotient Rule: $(\frac{f}{g})$' = $\frac{f'g-fg'}{g^2}$.

By writing this form in the limit definition,

$$\lim_{h \to 0} \frac{\frac{fx+h}{gx+h} - \frac{fx}{gx}}{h} = \lim_{h \to 0} \frac{fx+hgx-fxgx+h-fxgx+fxgx}{gx+hgx}$$
$$= \lim_{h \to 0} \frac{1}{gx+hgx} * \lim_{h \to 0} \frac{fx+hgx-fxgx+fxgx-fxgx+h}{h}$$
$$= \lim_{h \to 0} \frac{1}{gx+hgx} * \lim_{h \to 0} \frac{fx+hgx-fxgx}{h} + \lim_{h \to 0} \frac{fxgx-fxgx+h}{h}$$
$$= \lim_{h \to 0} \frac{1}{gx+hgx} * \lim_{h \to 0} gx * \frac{fx+h-fx}{h} + \lim_{h \to 0} fx * \frac{gx-gx+h}{h}.$$

With the following being true due to the definition, f(x) = $\lim_{h \to 0} fx$, f'(x) = $\lim_{h \to 0} \frac{\frac{fx+h}{-} fx}{h}$

g(x) = $\lim_{h \to 0} gx$ and $\lim_{h \to 0} gx+h$, g'(x) = $\lim_{h \to 0} \frac{\frac{gx+h}{-} gx}{h}$

Therefore this can be rewritten in the following form:

$$\frac{f'xgx-g'xfx}{gx^2}.$$

---

3) Show for matrices A and B with elements a_{ij} and b_{ij} respectively, that: .

$$\sum_{ij} a_{ij} b_{ij} = tr(AB^T)$$

To show this to be true, it is first important to note that the trace of a matrix as defined in the textbook, and in Linear Algebra, can happen only with a square matrix, R^{nxn}. This means that the matrix $A \epsilon R^{nxn}$ or if $A \epsilon R^{nxk}$ then matrix $B \epsilon kxn$. If A is R^{nxn} then B is R^{nxn}. It is beneficial to note a property of the Transpose of a square matrix that the transpose of a symmetrical square matrix is the same matrix, i.e. the Identity Matrix. However, for a matrix that is nxk this flips the dimensions when ordering row by column to order colun by row making a kxn matrix.

For the two cases, the square will be proved first. Where both A and B are some matrices ϵR^{nxn}. The transpose of the matrix B since it is a square, if it is symmetric is the same as B. Though, if it is not symmetric the Transpose will just change the rows into columns of the answer if one takes the $tr(AB^T)$.

The definition of the trace is as follows

$$\mathrm{tr}(\boldsymbol{A}) := \sum_{i=1}^{n} a_{ii}$$

where the two elements inside the trace are being multiplied the result written using the summation form is as follows. (Forgive the $\mathcal{P}$ the summation symbol is not cooperating).

$$tr(AB^T) = \mathcal{P}_{i=1}^{n} a_{ij}b_{ij}.$$

Where in the symmetric and square case, b_{ii} = b_{ii}, the diagonal remains the same. If the matrix is not symmetric, the diagonals remain the same when taking the transpose of the square matrix, thus the non-symmetry doesn't matter.

However, for a non-square case, where $A \epsilon R^{nxn}$ and $B \epsilon R^{kxn}$ when the transpose of B is taken, B's dimensions become nxk where the resulting matrix will be an nxn matrix. This holds true as well if A were in a mxn dimension with B^T in the nxk. The resulting matrix would still be square. The summation that would happen is as follows

$$\mathcal{P}_{i=1}^{n} a_{ij} * b_{ij}$$

Where i is the index of rows from A and j is the index of rows from B (columns from B^T), where the dimensions i,j = n. The resulting matrix would be an nxn matrix. Therefore, this is true since both A and B are not square, the trace would be taken of the respective indexes. Thus $\mathcal{P}_{i=1}\, a_{ij} * b_{ij}$ is true.

---

4) Book problems (5.5, 5.6, 5.7, 5.8(a,b))
5.5)

Consider the following functions:

$$f_1(\boldsymbol{x}) = \sin(x_1)\cos(x_2)\,, \quad \boldsymbol{x} \in \mathbb{R}^2$$
$$f_2(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top \boldsymbol{y}\,, \quad \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$$
$$f_3(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{x}^\top\,, \quad \boldsymbol{x} \in \mathbb{R}^n$$

a) What are the dimensions of $\frac{df_i}{dx}$?
Where the the dimensions are defined above, with x$_{f_1} \epsilon R^2$, though x_1 is the first index in the vector and x_2 is a separate index in the vector. Then for x_{f_2} since it is specifically asking for x the dimensions for x_{f_2} $\epsilon R^{(n}$].Then for x_{f_3} the dimensions are defined to be x_{f_3} ϵR^n.
Then the dimensions of $\frac{df_i}{dx}$ are R^{2xnxn}.
b) Compute the Jacobians.
From what I understand the Jacobians of f$_i xareasfollows.Foreachf_{1-3}are:$

f_1: $\frac{d}{dx}(sin(x_1)*cos(x_2))$
Using the product rule, with the vector x_{f_1} then the result is
$sin(x_1)-sin(x_2)$ + $cos(x_1)*cos(x_2)$.
f_2: $\frac{d}{dx}(x^Ty)$ where the multiplications of a row vector by a column vector results in a scalar number. The derivative of this new scalar number, is 0. Therefore the jacobian for f_2 is 0 for all entires.
f_3: $\frac{d}{dx}(xx^T)$ where x and x^T are of the same dimensions and are multiplied to result in a scalar, since they are the same vector. The jacobian of this would also result in zero.

Thus the jacobian of the defined $\frac{df_i}{dx}$ would be the 3 dimensional matrix with the jacobian of f_1 on the main diagonal, with f_2 above with their jacobian vector of all zeros, and f_3 below with their jacobian of zeros. The dimensions of the Jacobian R^{nxnx2}.

5.6)

Differentiate f with respect to $\boldsymbol{t}$ and g with respect to $\boldsymbol{X}$, where

$$f(\boldsymbol{t}) = \sin(\log(\boldsymbol{t}^\top \boldsymbol{t})), \qquad \boldsymbol{t} \in \mathbb{R}^D$$
$$g(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{AXB}), \qquad \boldsymbol{A} \in \mathbb{R}^{D\times E}, \boldsymbol{X} \in \mathbb{R}^{E\times F}, \boldsymbol{B} \in \mathbb{R}^{F\times D},$$

where $\operatorname{tr}(\cdot)$ denotes the trace.

For f'(t) with respect to t, the chain rule is used. In this problem I used $log_e(t^T t)$ or $ln(t^T t)$. For f(t) the derivative is as follows:
$f'(t) = sin(ln(t^T t))$ applying the chain rule twice, as $t^T t$ is an operation of its own. However, the vector t when multiplied by it's transpose, results in a scalar as it is the result of a dot product. If the equation was $t * t^T$, then the result would be a matrix that is DxD in dimensions. The derivative of a scalar is zero. Thus, the chain rule involves the two functions that act upon this statement. This results in $\cos(ln(t^T t)) * \frac{d}{dx}\ (ln(t^T t)) = \cos(ln(t^T t)) * \frac{1}{t^T t}$.
For g(X) the derivative is taken with respect to the matrix X. With derivatives of matrices, the operations are similar to those with scalar variables as long as the dimensions match. For g(X) the derivative is as follows:
g'(X) $= tr(AXB) = tr(AB)$. Where $tr(AB)\ \epsilon R^{DxE}$.

5.7 Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions by using the chain rule. Provide the dimensions of every single partial derivative. Describe your steps in detail.

a.

$$f(z) = \log(1 + z), \quad z = \boldsymbol{x}^\top \boldsymbol{x}, \quad \boldsymbol{x} \in \mathbb{R}^D$$

b.

$$f(\boldsymbol{z}) = \sin(\boldsymbol{z}), \quad \boldsymbol{z} = \boldsymbol{Ax} + \boldsymbol{b}, \quad \boldsymbol{A} \in \mathbb{R}^{E\times D}, \boldsymbol{x} \in \mathbb{R}^D, \boldsymbol{b} \in \mathbb{R}^E$$

where $\sin(\cdot)$ is applied to every element of $\boldsymbol{z}$.

a) I used a source on Math Stack Exchange to help with the actual distribution of $\frac{d}{dt}(x^T x)$.
The dimensions will be included after the derivative of f(z) is solved for.I used log_e or ln as the log. In using the chain rule to find the derivative the steps were the following.

$\frac{d}{dx}(ln(\text{h})) = \frac{1}{h} \implies \frac{1}{1+z}$, the dimensions of this is $R^D \to R$.
With the chain rule, this derivative of ln(h) is multiplied by h', h' is as follows
The easiest way to think about it is with the gradient. Where $\frac{dx^T x}{dx} = \nabla(x^T x) = \mathcal{P}_{j=1}^{D} \frac{\partial x_j^2}{\partial^2 x_i}$
$\implies \partial_{ji} x_j + x_j \partial_{ij}$,
where in a transposed vector, the ith or jth term is the same, then it becomes
$x_i + x_i = 2x_i$. Thus the derivative of $x^T x$ is $\mathcal{P}_{i=1}^{D} 2x_i$, and the dimension is R.
f'(x) $= (\frac{1}{1+x^T x})*(\mathcal{P}_{i=1}^{D} 2x_i)$, the dimensions of this end as ϵR.

b) Similarly, by using the chain rule where I first solve with z before substituting the values the derivative and the dimensions are as follows

f'(z) $= \cos(z) * z'$ where $z = Ax + b$, with A as a matrix, and x, b are vectors.
$z = Ax + b \implies z' = A$, dimensions go $R^{ExD} * R^D + R^E$ to R^{DxE}. When substituted into the exuation above, the result is
$\cos(Ax + b)* A$.
The dimensions of the result of f'(z) are R^{DxExD}.

5.8)

Compute the derivatives $\mathrm{d}f/\mathrm{d}\boldsymbol{x}$ of the following functions. Describe your steps in detail.

a. Use the chain rule. Provide the dimensions of every single partial derivative.

$$f(z) = \exp(-\tfrac{1}{2}z)$$
$$z = g(\boldsymbol{y}) = \boldsymbol{y}^\top \boldsymbol{S}^{-1}\boldsymbol{y}$$
$$\boldsymbol{y} = h(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\mu}$$

where $\boldsymbol{x}, \boldsymbol{\mu} \in \mathbb{R}^D$, $\boldsymbol{S} \in \mathbb{R}^{D\times D}$.

b.

$$f(\boldsymbol{x}) = \mathrm{tr}(\boldsymbol{x}\boldsymbol{x}^\top + \sigma^2 \boldsymbol{I}), \quad \boldsymbol{x} \in \mathbb{R}^D$$

Here $\mathrm{tr}(\boldsymbol{A})$ is the trace of $\boldsymbol{A}$, i.e., the sum of the diagonal elements A_{ii}.
Hint: Explicitly write out the outer product.

a) The following shows the work done to solve for this problem. The dimensions are written to the right of the work done after each step.

a) $f(z) = e^{-\frac{1}{2}z}$

$z = g(y) = y^T S^{-1} y$

$y = h(x) = x - \mu$

$f'(z) = -\frac{1}{2} e^{-\frac{1}{2}z}$

$y'(x) = x' \rightarrow \frac{d}{dx}\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) \rightarrow \begin{bmatrix} \frac{dx_1}{dx} \\ \frac{dx_2}{dx} \\ \vdots \\ \frac{dx_n}{dx} \end{bmatrix}$, $\in \mathbb{R}^D$

$g'(y) = (x^T)' S^{-1} (x')$, $\in \mathbb{R}^{D\times D\times D}$

$\underbrace{[dx_1 \dots dx_n] \begin{bmatrix} s_{11} & \cdots & s_{n1} \\ \vdots & \ddots & \vdots \\ s_{1n} & \cdots & s_{nn} \end{bmatrix}^{-1} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix}}_{\mathbb{R}^{D\times D\times D}}$, $\mathbb{R}^{D\times D}$

$f'(x) = -\frac{1}{2} e^{-\frac{1}{2}((x^T)' S^{-1} x')}$, where $-\frac{1}{2} e^{-\frac{1}{2}\cdots} \in \mathbb{R}$ and $(x^T)' S^{-1} x' \in \mathbb{R}^{D\times D}$

Thus the dimensions of $f'(x) \in \mathbb{R}^{D\times D}$.

b) Shown below is the work done to solve for this problem. The dimensions are written to the right of the work done after each step. The biggest question came when performing the derivative on xx^T, where that is the outer product. In a prior problem the derivative of x^Tx was shown to be the inner product.

b) $f(x) = tr\left(xx^T + \sigma^2 \cdot I\right)$, $x \in \mathbb{R}^D$

With the lack of information of σ being a vector, $\sigma \in \mathbb{R}$.

Thus,

$$f'(x) = tr\left(xx^T + \sigma I\right) = tr\left(xx^T + \sigma^2 I\right) \cdot (xx^T)'$$

The derivative of the trace is the trace, as it is $\sum_{i=1}^{n} a_{ii}$ or the sum of the diagonals.

This leaves the expression

$$f'(x) = tr\left(xx^T + \sigma^2 I\right) \cdot (xx^T)' \quad , \quad xx^T \in \mathbb{R}^{D \times D}, \ \sigma^2 I \in \mathbb{R}^{D \times D}, \ tr(\cdot) \in \mathbb{R}$$

To solve for $(xx^T)'$ it is important to note that xx^T is the outer product or cross product of the vector x.

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} [x_1, \ldots, x_n] = xx^T = \begin{bmatrix} x_1x_1 & \cdots & x_1x_n \\ \vdots & \ddots & \vdots \\ x_nx_1 & \cdots & x_nx_n \end{bmatrix} \quad , \quad \mathbb{R}^{D^T} \cdot \mathbb{R}^{D} = \mathbb{R}^{D \times D}$$

The derivative of this new matrix $\frac{d}{dx}(xx^T)$ occurs 'D' times,

Resulting in a tensor $\in \mathbb{R}^{D \times D \times D}$.

Thus the derivative

$$\Rightarrow \underbrace{tr\left(xx^T + \sigma^2 I\right)}_{\text{Sum of all diagonal elements of this addition.}} \cdot \underbrace{(xx^T)'}_{\text{Matrix.}} \Rightarrow \in \mathbb{R}^1 \cdot \mathbb{R}^{D \times D \times D} = \in \mathbb{R}^{D \times D \times D}.$$

Therefore, the derivative is

$$\underline{\underline{f'(x) = tr\left(xx^T + \sigma^2 I\right)(xx^T)' \in \mathbb{R}^{D \times D \times D}.}}$$

5) Consider the vector $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \epsilon R^2$. Given the functions $y(x) : R^2 \to R^3$, where $y = \begin{pmatrix} x_1^2 x_2 \\ x_1 + x_2 \\ 3x_2 \end{pmatrix}$, and $w(y) : R^3 \to R^2$, where $w = \begin{pmatrix} y_1 + y_2 \\ y_1 y_3 + y_2 \end{pmatrix}$. Find $\frac{\partial w}{\partial x}$.

To make it clearer, $\frac{\partial w}{\partial x} = \begin{pmatrix} x_1^2 + 2x_1x_2 + 2 \\ 3x_1^2 x_2 + 6x_1x_2 + 5 \end{pmatrix} \epsilon R^2$.

$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \in \mathbb{R}^2$; $y(\vec{x}) : \mathbb{R}^2 \to \mathbb{R}^3$, $y = \begin{bmatrix} x_1^2 \cdot x_2 \\ x_1 + x_2 \\ 3x_2 \end{bmatrix}$, and $w(y) : \mathbb{R}^3 \to \mathbb{R}^2$, $w = \begin{bmatrix} y_1 + y_2 \\ y_1 y_3 + y_2 \end{bmatrix}$.

$\frac{\partial \vec{w}}{\partial \vec{x}} = \frac{\partial \vec{w}}{\partial x_1} + \frac{\partial \vec{w}}{\partial x_2}$

$\frac{\partial}{\partial x}\left(\begin{bmatrix} x_1^2 \cdot x_2 + (x_1 + x_2) \\ x_1^2 x_2 (3x_2) + (x_1 + x_2) \end{bmatrix}\right) \Longrightarrow$

$\frac{\partial w_1}{\partial x_1} = 2x_1x_2 + 1$

$\frac{\partial w_1}{\partial x_2} = x_1^2 + 1$

$\boxed{\frac{\partial w_1}{\partial x} = x_1^2 + 2x_1x_2 + 2}$

$\frac{\partial w_2}{\partial x_1} = x_1^2 x_2 \cdot 3x_2 = 2x_1 \cdot 3x_2 + 0(\cdot) = 6x_1x_2 + 1$

$\frac{\partial w_2}{\partial x_2} = 3x_1^2 x_2 + 3 + 1 = 3x_1^2 x_2 + 4$

$\frac{\partial w_2}{\partial x} = 3x_1^2 x_2 + 6x_1x_2 + 5$

Thus $\frac{\partial \vec{w}}{\partial \vec{x}}$ is $\begin{bmatrix} x_1^2 + 2x_1x_2 + 2 \\ 3x_1^2 x_2 + 6x_1x_2 + 5 \end{bmatrix}, \in \mathbb{R}^2$.

6) Find the following derivatives, where capital bold letters stand for matrices and lowercase bold letters for vectors. (The work starts on the next page for this problem).

(a) $\frac{\partial(\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x})}{\partial \boldsymbol{x}}$

(b) $\frac{\partial \; tr(\boldsymbol{X})}{\partial \boldsymbol{X}}$

(c) $\frac{\partial \; tr(\boldsymbol{X}\boldsymbol{X}^T)}{\partial \boldsymbol{X}}$

(d) $\frac{\partial \; (\sqrt{(\boldsymbol{x}-\boldsymbol{a})^T(\boldsymbol{x}-\boldsymbol{a})})}{\partial \boldsymbol{x}}$

Where for a) B is the matrix represented by a form of 2xA (refer to handwritten work below.)
Where for b) l is $\mathbb{I}$.

a) $\dfrac{d\, x^T A x}{dx}$ $\rightarrow$ Note, $x^T x$ is the inner product or the dot product. A is a matrix that multiplies in the middle so the form should be of

$$\begin{bmatrix} 2dx_1 \cdot a & dx_1 dx_2 \cdot a & \cdots \\ dx_2 dx_1 \cdot a & & \\ \vdots & & \end{bmatrix}$$

where a is terms from A.

$$x = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \in \mathbb{R}^{1\times m}, \quad A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ a_{n1} & \cdots & & a_{nm} \end{bmatrix} \in \mathbb{R}^{n\times m}$$

$$x^T = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{R}^{m\times 1}$$

So now consider

$$\frac{d}{dx}\left(x^T A\right) \Rightarrow \frac{d}{dx}\left(\begin{bmatrix} x^T \end{bmatrix}\begin{bmatrix} A \end{bmatrix}\right) \Rightarrow \mathbb{R}^m \cdot \mathbb{R}^{m\times n} \longrightarrow \mathbb{R}^n \cdot \mathbb{R}^m \rightarrow \mathbb{R}^{n\times m}$$

$$\begin{bmatrix} dx_1 \\ dx_2 \\ \vdots \\ dx_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} dx_1(a_{11} + a_{21} + \cdots + a_{n1}) = dx^T A_1 \\ dx_2(a_{12} + a_{22} + \cdots + a_{n2}) = dx^T A_2 \\ \vdots \\ dx_n(a_{1m} + a_{2m} + \cdots + a_{nm}) = dx^T A_n \end{bmatrix} \in \mathbb{R}^n$$

Now take $dx(dx^T A)$ since $x^T A x$ is the entire formula:

$$\begin{bmatrix} dx^T A_1 \\ dx^T A_2 \\ \vdots \\ dx^T A_n \end{bmatrix} \begin{bmatrix} dx_1 & dx_2 & \cdots & dx_m \end{bmatrix} = \underbrace{\begin{bmatrix} dx_1^2(a_{11} + \cdots + a_{n1}) & dx_1 dx_2 (a_{12} + \cdots + a_{n2}) & \cdots & dx_1 dx_n (a_{1m} + \cdots + a_{nm}) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ dx_m dx_1 (a_{11} + \cdots + a_{n1}) & \cdots & \cdots & dx_m dx_n (a_{1m} + \cdots + a_{nm}) \end{bmatrix}}_{B}$$

where B is some matrix $\in \mathbb{R}^{n\times m}$, B is the result of the derivative $\dfrac{d\, x^T A x}{dx}$.

b) $\frac{\partial\, tr(X)}{\partial X}$, where X is some matrix – assumed to be square since the trace is being taken. $X \in \mathbb{R}^{n \times n}$.

$$\frac{\partial\, tr(X)}{\partial X} = \frac{\partial}{\partial X}\left(tr(X)\right) = tr(X + dX)$$

Using Einstein's summation rule

$$\partial tr(X) = tr(\underbrace{X + dX}) - tr(X)$$

$$\partial\, tr(X_{ij} + dX_{ij})$$

$$= \frac{\partial}{\partial X_{mn}} \sum_{kl} X_{kl} \rightarrow \frac{\partial}{\partial X_{mn}} \sum_{kl} \partial X_{km}\, \partial X_{nl} \Rightarrow \sum 1 = \underline{\underline{1}}$$

Thus, the derivative is simply 1.

★ I had to use an example $\frac{\partial}{\partial X} tr(AXB)$ to understand what was going on.

I also used Math Stack exchange to get Einstein's Summation Rule.

c)

$$\frac{\partial}{\partial X} tr(XX^T), \quad X \in R^{n\times n}$$

$XX^T : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n\times n}$ } Matrix algebra 2007 James Gentle

$\frac{d}{dX}(XX^T) : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$

let $X^T = A$, the inner product ?

$$\frac{\partial}{\partial X} tr(AX) = A^T$$

$$\downarrow$$

$$\frac{\partial \sum_i \sum_k a_{ik} x_{ki}}{\partial x_{ij}} = 2a_{ji} = 2X^T$$

But this problem asks for the outer product of X + X^T ($\otimes$).

$$tr(XX^T)' \Rightarrow tr\left(X\frac{\partial}{\partial X^T} + X^T \frac{\partial}{\partial X}\right)$$

$\frac{\partial}{\partial x^T} = \left(\frac{\partial}{\partial x}\right)^T$

$x^T = (x)^T$

$$\hookrightarrow \sum_i \sum_k x_{ik}\, \partial x_{ki} + x_{ki}\, \partial x_{ik} \Rightarrow 2X$$

where $\frac{\partial tr(XX^T)}{\partial X} = \underline{\underline{2X.}}$

where $2X \in \mathbb{R}^{n\times n}$.

* I used Matrix Algebra by James Gentle + class textbook for help on 6c)

d) $$\frac{\partial\left(\sqrt{(x-a)^T(x-a)}\right)}{\partial x}, \quad x \in \mathbb{R}^n, a \in \mathbb{R}^n$$

$$\frac{\partial}{\partial x}\left(\left((x-a)^T(x-a)\right)^{1/2}\right) \to \text{chain rule } \times 2$$

1 - for $\sqrt{\ }$

2 - for $\cdot$

$$(x-a)^T \cdot \frac{\partial}{\partial x-a} + (x-a) \cdot \frac{\partial}{\partial (x-a)^T}$$

$$\frac{1}{2}\left((x-a)^T(x-a)\right)^{-1/2}\left((x-a)^T\frac{\partial}{\partial x} + (x-a)\frac{\partial}{\partial x^T}\right)$$

$$\frac{1}{2\sqrt{(x-a)^T(x-a)}}\left((x-a)^T\frac{\partial}{\partial x} + (x-a)\frac{\partial}{\partial x}\right)$$

$$\Rightarrow \underline{\underline{\frac{(x-a)^T\frac{\partial}{\partial x} + (x-a)\frac{\partial}{\partial x}}{2\sqrt{(x-a)^T\cdot(x-a)}}}}$$