Linear Algebra

1. Vector notations (same as the one in high school)

e.g. $\vec{x} = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} = (1\ 4\ 3)^T \Rightarrow \vec{x}^T = ((1\ 4\ 3)^T)^T = (1\ 4\ 3)$

$(\vec{x}^T)^T = \vec{x}$

2. Matrix notations

e.g. $A = \begin{pmatrix} 1 & 7 \\ 4 & -6 \\ 3 & 0 \end{pmatrix} \Leftrightarrow A \in M(3,2)$ where M is the space of 2x3 matrices

Generally, in terms of arbitrary $m \times n$ matrices, $M(m,n)$ is the space of $m\times n$ matrices

we don't use the notations $A \in \mathbb{R}^{m\times n}$ (usually)

3. Vector & Matrix Element notations

e.g. $A = \begin{pmatrix} 1 & 7 \\ 4 & -6 \\ 3 & 0 \end{pmatrix}$ $[A]_2 = (4,-6)$, $[A]_{3,1} = 3$

$[A]_{(2:3):(1:2)} = \begin{pmatrix} 4 & -6 \\ 3 & 0 \end{pmatrix}$ $[A]_{:,2} = \begin{pmatrix} 7 \\ -6 \\ 0 \end{pmatrix}$

4. Dot product

e.g. $\vec{x} = (1\ 4\ 3)^T$, $\vec{y} = (7, -6, 0)^T$

$\Rightarrow \vec{x}\cdot\vec{y} = \begin{pmatrix} 1 \\ 4 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 7 \\ -6 \\ 0 \end{pmatrix} = 1\times 7 + 4\times(-6) + 3\times 0 = -17$

generally, $\vec{x}\cdot\vec{y} = \sum_{i=1}^{n} [\vec{x}]_i [\vec{y}]_i$, notice that vectors of $\vec{x}$ and $\vec{y}$ must be in the same size

we can rewrite in $\vec{x}^T \cdot \vec{y} = (1\ 4\ 3)\begin{pmatrix} 7 \\ -6 \\ 0 \end{pmatrix} = 7 - 24 + 0 = -17$ (1x1 matrix)

in linear algebra, we can consider a 1x1 matrix as a number

5. Matrix Multiplication

e.g.1 Given $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \\ 4 & 4 & 4 \end{pmatrix}$, $B = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$

$\Rightarrow AB = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \\ 4 & 4 & 4 \end{pmatrix}}_{4\times 3} \underbrace{\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}}_{3\times 1} = \begin{pmatrix} 6 \\ 12 \\ 18 \\ 24 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + 2\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} = 6\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$

4×1

e.g.2 Prove: $(A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) = \vec{x}^T A^T A \vec{x} + \vec{b}^T\vec{b} - 2\vec{b}^T A\vec{x}$

where $A \in M(m,n)$, $\vec{x} \in M(n,1)$ and $\vec{b} \in M(m,1)$

$$
\begin{aligned}
(A\vec{x}-\vec{b})^T(A\vec{x}-\vec{b}) &= [(A\vec{x})^T - \vec{b}^T](A\vec{x}-\vec{b}) \\
&= (\vec{x}^T A^T - \vec{b}^T)(A\vec{x}-\vec{b}) \\
&= \vec{x}^T A^T A \vec{x} - \vec{x}^T A^T \vec{b} - \vec{b}^T A \vec{x} + \vec{b}^T\vec{b} \\
&= \vec{x}^T A^T A\vec{x} + \vec{b}^T\vec{b} - 2\vec{b}^T A\vec{x}
\end{aligned}
$$

$1\times n$ $n\times m$ $m\times 1$

1×1

$1\times m$ $m\times n$ $n\times 1$

1×1

how to determine whether $\vec{x}^T A^T \vec{b} = \vec{b}^T A\vec{x}$ or not ?

Just obtain the size of each matrix (vector) appeared in two terms, respectively, and derive the updated sizes after multiplication to see if they are both 1x1.

6. Gradient

e.g.1 $f(\vec{x}) = \vec{b}^T\vec{x} = \vec{x}^T\vec{b} \Rightarrow f(\vec{x})$ is symmetric

$\Rightarrow \vec{b}^T\vec{x}$ and $\vec{x}^T\vec{b}$ are necessarily square matrices

let $\vec{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in M(m,1)$ and $\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \in M(m,1)$

$\Rightarrow \vec{b}^T\vec{x} = (b_1\ b_2 \cdots b_m)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} = \sum_{i=1}^{m} b_i x_i$

$\vec{x}^T\vec{b} = (x_1\ x_2 \cdots x_m)\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \sum_{i=1}^{m} b_i x_i$

$$\Rightarrow \nabla f(\vec{x}) = \nabla_{\vec{x}}(\vec{b}^T\vec{x}) = \nabla_{\vec{x}}(\vec{x}^T\vec{b}) = \begin{pmatrix} \sum_{i=1}^{m} \frac{\delta}{\delta x_1} b_i x_i \\ \vdots \\ \sum_{i=1}^{m} \frac{\delta}{\delta x_m} b_i x_i \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \vec{b}$$

e.g.5 Given $f(\vec{x}) = \frac{1}{2}\vec{x}^T A\vec{x} + \vec{x}^T\vec{b} + c$ (where c is constant), show that $\nabla f(\vec{x}) = A\vec{x} + \vec{b}$

where A is symmetric

$\nabla f(\vec{x}) = \frac{1}{2}\cdot 2A\vec{x} + \vec{b} + \vec{0} = A\vec{x} + \vec{b}$

e.g.2 $f(\vec{x}) = \vec{x}^T\vec{x}$ let $\vec{x} = (x_1\ x_2 \cdots x_n) \in M(1,n)$

$$\Rightarrow \vec{x}^T\vec{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}(x_1\ x_2 \cdots x_n) = \sum_{i=1}^{n} x_i^2 \Rightarrow \nabla f(\vec{x}) = \nabla_{\vec{x}}(\vec{x}^T\vec{x}) = \begin{pmatrix} \sum_{i=1}^{n} \frac{\delta}{\delta x_1} x_i^2 \\ \vdots \\ \sum_{i=1}^{n} \frac{\delta}{\delta x_n} x_i^2 \end{pmatrix} = 2\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = 2\vec{x}$$

e.g.3 $f(\vec{x}) = \vec{x}^T B\vec{x}$ (B is symmetric)

let $\vec{x} = \begin{pmatrix} x_1 \\ x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$,

$\vec{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & & \vdots \\ \vdots & & \ddots & \\ b_{n1} & \cdots & & b_{nn} \end{pmatrix} \in M(n,n)$

where $b_{pq} = b_{qp}$ ($1 \le p \le n$, $1 \le q \le n$)

$$\Rightarrow \vec{x}^T B\vec{x} = (x_1\ x_2 \cdots x_n)\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & & \vdots \\ \vdots & & \ddots & \\ b_{n1} & \cdots & & b_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= \left(\sum_{i=1}^{n} x_i b_{i1}\ \ \sum_{i=1}^{n} x_i b_{i2} \cdots \sum_{i=1}^{n} x_i b_{in}\right)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= x_1^2 b_{11} + x_2^2 b_{22} + \cdots + x_n^2 b_{nn} + 2(x_1x_2 b_{21} + x_1x_3 b_{31} + \cdots + x_1x_n b_{n1} + x_2x_3 b_{32} + x_2x_4 b_{42} + \cdots + x_2x_n b_{n2} + \cdots + x_{n-1}x_n b_{n(n-1)})$$

$$= x_1\sum_{i=1}^{n} x_i b_{i1} + x_2\sum_{i=1}^{n} x_i b_{i2} + \cdots + x_n\sum_{i=1}^{n} x_i b_{in}$$

$$= \sum_{j=1}^{n}\left(x_j\sum_{i=1}^{n} x_i b_{ij}\right)$$

$$\Rightarrow \nabla f(\vec{x}) = \nabla_{\vec{x}}\left[\sum_{j=1}^{n}\left(x_j\sum_{i=1}^{n} x_i b_{ij}\right)\right] = \begin{pmatrix} \frac{\delta}{\delta x_1}\sum_{j=1}^{n}\left(x_j\sum_{i=1}^{n} x_i b_{ij}\right) \\ \vdots \\ \frac{\delta}{\delta x_n}\sum_{j=1}^{n}\left(x_j\sum_{i=1}^{n} x_i b_{ij}\right) \end{pmatrix} = \begin{pmatrix} 2x_1 b_{11} + 2(x_2 b_{21} + x_3 b_{31} + \cdots + x_n b_{n1}) \\ \vdots \\ 2x_n b_{nn} + 2(x_{n-1} b_{n(n-1)} + x_{n-2} b_{n(n-2)} + \cdots + x_1 b_{n1}) \end{pmatrix}$$

$$= 2\begin{pmatrix} x_1 b_{11} + x_2 b_{21} + \cdots + x_n b_{n1} \\ \vdots \\ x_n b_{nn} + x_{n-1} b_{n(n-1)} + \cdots + x_1 b_{n-1} \end{pmatrix} = 2\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & \ddots & & b_{2n} \\ \vdots & & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = 2B\vec{x}$$

e.g.4 Given $f(\vec{x}) = (\vec{b} - A\vec{x})^T(\vec{b} - A\vec{x})$, show that $\nabla f(\vec{x}) = -2A^T(\vec{b} - A\vec{x})$

Proof: $f(\vec{x}) = [\vec{b}^T - (A\vec{x})^T](\vec{b} - A\vec{x})$

$= (\vec{b}^T - \vec{x}^T A^T)(\vec{b} - A\vec{x})$

$= \vec{b}^T\vec{b} - \vec{b}^T A\vec{x} - \vec{x}^T A^T\vec{b} + \vec{x}^T A^T A\vec{x}$

$= \vec{b}^T\vec{b} - 2\vec{x}^T A^T\vec{b} + \vec{x}^T A^T A\vec{x}$

$\Rightarrow \nabla f(\vec{x}) = \vec{0} - 2A^T\vec{b} + 2A^T A\vec{x} = -2A^T(\vec{b} - A\vec{x})$