NMEP HW #1

1. a. $\begin{bmatrix}1\\2\end{bmatrix}+\begin{bmatrix}3\\2\end{bmatrix}=\begin{bmatrix}4\\4\end{bmatrix}$

b. $\begin{bmatrix}1\\2\\3\end{bmatrix}+\begin{bmatrix}3\\2\\1\end{bmatrix}+\begin{bmatrix}4\\2\\0\end{bmatrix}=\begin{bmatrix}8\\6\\4\end{bmatrix}$

c. $\begin{bmatrix}0&1\\1&0\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}2\\1\end{bmatrix}$

2. a. Nullspace: $A\vec{x}=\vec{0}$

$$\left[\begin{array}{ccc|c}1&2&3&0\\-1&4&2&0\\0&6&5&0\end{array}\right]$$

$R_2=R_1+R_2$ $$\left[\begin{array}{ccc|c}1&2&3&0\\0&6&5&0\\0&6&5&0\end{array}\right]$$

$R_3=R_2-R_3$ $$\left[\begin{array}{ccc|c}1&2&3&0\\0&6&5&0\\0&0&0&0\end{array}\right]$$

$R_1=R_1-\frac{1}{3}R_2$ $$\left[\begin{array}{ccc|c}1&0&\frac{4}{3}&0\\0&6&5&0\\0&0&0&0\end{array}\right]$$

$R_2=\frac{R_2}{6}$ $$\left[\begin{array}{ccc|c}1&0&4/3&0\\0&1&5/6&0\\0&0&0&0\end{array}\right]$$

$x_1+\frac{4}{3}x_3=0 \rightarrow x_1=-\frac{4}{3}x_3$

$x_2+\frac{5}{6}x^3=0 \rightarrow x_2=-\frac{5}{6}x_3$

$$\vec{x}=\begin{bmatrix}-\frac{4}{3}\\-\frac{5}{6}\\1\end{bmatrix}x_3=\text{span}\left\{\begin{bmatrix}-8\\-5\\6\end{bmatrix}\right\}$$

Columnspace: Only 2 of the vectors are linearly independent, so column space is

$$\text{span}\left\{\begin{bmatrix}1\\-1\\0\end{bmatrix},\begin{bmatrix}2\\4\\6\end{bmatrix}\right\}$$

b. $\boxed{\text{No}}$, a linear transformation can be represented by a matrix. But for 2 matrices, AB is not necessarily equal to BA. For example, if $A=\begin{bmatrix}1&2\\3&4\end{bmatrix}$ and $B=\begin{bmatrix}5&6\\7&8\end{bmatrix}$,

2. b (cont). $AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}$

$BA = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix}$. $\Big\}$ $BA \neq AB$.

Therefore, $T_1(T_2(\vec{v})) = T_2(T_1(\vec{v}))$ is not always correct.

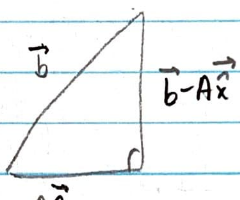
c. $\boxed{\text{No}}$, take this counterexample. If $T_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $T_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$,

then $T_1(T_2(\vec{v}))$

$= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \vec{v}$

$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \vec{v}$

$= \vec{0}$, but T_1 and T_2 both aren't all-zero matrices.

3. 

a. $\hat{x} = (A^T A)^{-1} A^T \vec{b}$

$= \left(\begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \\ 0 & 6 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ -1 & 4 & 6 \\ 1 & 2 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \\ 0 & 6 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

$= \underbrace{\begin{bmatrix} 3 & 0 & -6 \\ 0 & 24 & 24 \\ -6 & 24 & 36 \end{bmatrix}^{-1}} \begin{bmatrix} 9 \\ 12 \\ -6 \end{bmatrix}$

$\det(m) = -0 \begin{vmatrix} 0 & 24 \\ -6 & 36 \end{vmatrix} + 24 \begin{vmatrix} 3 & -6 \\ -6 & 36 \end{vmatrix} - 24 \begin{vmatrix} 3 & -6 \\ 0 & 24 \end{vmatrix}$

$= 24(108 - 36) - 24(72 - 0) = 0 \rightarrow$ not invertible

Since A does not have linearly independent columns $\left(-2\begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} + \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 6 \\ 0 \end{bmatrix} \right)$,

let $A = \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix}$.

$\hat{x} = (A^T A)^{-1} A^T \vec{b}$

$= \left(\begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 4 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & -1 & 1 \\ 2 & 4 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix}$

3. a (cont).
$$= \begin{bmatrix} 3 & 0 \\ 0 & 24 \end{bmatrix}^{-1} \begin{bmatrix} 9 \\ 12 \end{bmatrix}$$
$$= \frac{1}{72} \begin{bmatrix} 24 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 9 \\ 12 \end{bmatrix}$$
$$= \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{24} \end{bmatrix} \begin{bmatrix} 9 \\ 12 \end{bmatrix} = \begin{bmatrix} 3 \\ 0.5 \end{bmatrix}, \rightarrow \hat{\vec{x}} = \boxed{\begin{bmatrix} 3 \\ 0.5 \\ 0 \end{bmatrix}}$$

b. $A\hat{\vec{x}} = \begin{bmatrix} 1 & 2 \\ -1 & 0 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \\ 4 \end{bmatrix}$

c. $\|\vec{b} - A\hat{\vec{x}}\| = \left\| \begin{bmatrix} 3 \\ -1 \\ 5 \end{bmatrix} - \begin{bmatrix} 4 \\ -1 \\ 4 \end{bmatrix} \right\| = \left\| \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\| = \sqrt{(-1)^2 + 0^2 + 1^2} = \boxed{\sqrt{2}}$

4.
$$\min_{w} \|X\vec{w} - \vec{y}\|_2^2 + \lambda\|w\|_2^2$$
$$= \min_{w} (X\vec{w} - \vec{y})^T (X\vec{w} - \vec{y}) + \lambda \vec{w}^T \vec{w}$$
$$= \min_{w} \vec{y}^T\vec{y} + \vec{w}^T X^T X \vec{w} - \vec{y}^T X \vec{w} - \vec{w}^T X^T \vec{y} + \lambda \vec{w}^T I \vec{w}$$
$$\nabla_{\vec{w}} = (X^TX + (X^TX)^T)\vec{w} - 2X^T\vec{y} + \lambda(I + I^T)\vec{w}$$
$$= (X^TX + X^TX)\vec{w} - 2X^T\vec{y} + \lambda(2I)\vec{w}$$
$$= 2X^TX\vec{w} - 2X^T\vec{y} + 2\lambda I\vec{w} = 0$$
$$(X^TX + \lambda I)\vec{w} = X^T\vec{y}$$
$$\vec{w} = (X^TX + \lambda I)^{-1} X^T \vec{y}$$

We can tune the hyperparameter λ by trying out a bunch of different values and then checking how well the corresponding $\vec{w}$ fits our data (GridSearch).