Assignment 1

Lokesh Mohanty (SR no: 21014)

September 2022

Problem 1

$$\mathbf{A} \in \mathbb{R}^{m \times n}, \mathbf{B} \in \mathbb{R}^{n \times p}$$
$$R(.) \rightarrow \text{ range/column space of matrix}$$
$$N(.) \rightarrow \text{ null space of matrix}$$
$$dim(.) \rightarrow \text{ dimension of vector space}$$

(a) **Prove**: $dim[R(\mathbf{AB})] \leq dim[R(\mathbf{A})]$

We can consider $\mathbf{B}$ as a group of column vectors i.e., $\{\mathbf{b_1}|\mathbf{b_2}|...|\mathbf{b_p}\}$ where $\mathbf{b_i} \in \mathbb{R}^n$, which implies that, $\mathbf{AB} = \{\mathbf{Ab_1}|\mathbf{Ab_2}|...|\mathbf{Ab_p}\}$ where $\mathbf{Ab_i} \in \mathbb{R}^m$ and we know that every $\mathbf{Ab_i}$ is a linear combination of the columns of $\mathbf{A}$.

Hence, all the columns of $\mathbf{AB}$ are some linear combinations of the columns of $\mathbf{A}$.

$\therefore$ The maximum dimension they can span is the dimension of $\mathbf{A}$.

$\implies$ Column space/Range of $\mathbf{AB} \leq$ Column space/Range of $\mathbf{A}$
$\implies$ $\boxed{dim[R(\mathbf{AB})] \leq dim[R(\mathbf{A})]}$

(b) Given that the matrix $\mathbf{B}$ is non singular, **prove** $dim[R(\mathbf{AB})] = dim[R(\mathbf{A})]$

From problem 1.a, we know that

$$dim[R(\mathbf{AB})] \leq dim[R(\mathbf{A})] \tag{1}$$

Since $\mathbf{B}$ is non singular, its inverse exists hence from equation 1, if we replace $\mathbf{A}$ with $\mathbf{AB}$ and $\mathbf{B}$ with $\mathbf{B}^{-1}$, we get

$$dim[R((\mathbf{AB})\mathbf{B}^{-1})] \leq dim[R(\mathbf{AB})]$$
$$dim[R(\mathbf{A})] \leq dim[R(\mathbf{AB})] \tag{2}$$

$\therefore$ From equations 1 and 2 we can confirm that

$$\boxed{dim[R(\mathbf{AB})] = dim[R(\mathbf{A})]}$$

(c) **Prove** $dim[N(\mathbf{AB})] \leq dim[N(\mathbf{A})] + dim[N(\mathbf{B})]$

We know that for a matrix $\mathbf{A}$ its null space is the set of $\mathbf{x}$ that satisfy $\mathbf{Ax} = 0$. Hence the null space of $\mathbf{A}$ is the set of $\mathbf{x}$ that satisfy $\mathbf{ABx} = 0$

Which can be written as $\mathbf{A(Bx)} = \mathbf{0}$ which implies that all $\mathbf{x}$ in $N(\mathbf{B})$ are also in $N(\mathbf{AB})$ i.e., $N(\mathbf{B}) \subseteq N(\mathbf{AB})$ as $\forall \mathbf{A}, \mathbf{A0} = \mathbf{0}$. And $\forall \mathbf{x} \notin N(\mathbf{B}), \mathbf{x} \in N(\mathbf{AB})$, $\mathbf{Bx}$ must belong to $N(\mathbf{A})$

$$\implies N(\mathbf{AB}) - N(\mathbf{B}) \subseteq N(\mathbf{A})$$
$$\implies N(\mathbf{AB}) \subseteq N(\mathbf{A}) + N(\mathbf{B})$$
$$\implies \boxed{dim(N(\mathbf{AB})) \leq dim(N(\mathbf{A})) + dim(N(\mathbf{B}))}$$

(d) **Prove** $dim[R(\mathbf{A})] + dim[N(\mathbf{A})] = n$

We know that $\mathbf{Ax}$ can be considered as a linear combination of the columns of $\mathbf{A}$ with their coefficients as the elements of $\mathbf{x}$. Assuming there are a linearly independent columns in A, we know that other than the trivial case, for all other coefficients, the linear combination will be non zero. That is, for any values of the corresponding a elements(coefficients) in $\mathbf{x}$ the linear combination will not be zero. Only by modifying the other $n - a$ elements in $\mathbf{x}$ can we make the resultant zero as those are the coefficients of the $n - a$ columns which are linearly dependent on the a columns.

From this we can say that the dimension of null space of $\mathbf{A}$ is $n - a$ as we can select any values for the $n - a$ elements and we can select the other a elements such that their linear combination with the corresponding a columns of $\mathbf{A}$ sums up to be the negative of the linear combination of the $n-a$ elements with the $n-a$ columns which sum up to zero.

Hence if there are a linearly independent columns in $\mathbf{A}$, then $dim[R(\mathbf{A})] = a$ and $dim[N(\mathbf{A})] = n - a$ which implies that $\boxed{dim[R(\mathbf{A})] + dim[N(\mathbf{A})] = n}$

(e) **Prove** $rank(\mathbf{A}) + rank(\mathbf{B}) - n \leq rank(\mathbf{AB}) \leq min(rank(\mathbf{A}), rank(\mathbf{B}))$

From 1.c and 1.d we know that,

$$dim[N(\mathbf{AB})] \leq dim[N(\mathbf{A})] + dim[N(\mathbf{B})]$$
$$\implies r - dim[R(\mathbf{AB})] \leq n - dim[R(\mathbf{A})] + r - dim[R(\mathbf{B})]$$
$$\implies dim[R(\mathbf{A})] + dim[R(\mathbf{B})] \leq dim[R(\mathbf{AB})]$$

And since $rank(\mathbf{A}) = dim[R(\mathbf{A})]$,

$$\implies \boxed{rank(\mathbf{AB}) \geq rank(\mathbf{A}) + rank(\mathbf{B})} \tag{3}$$

From the 1.c and 1.d we know that,

$$dim[N(\mathbf{AB})] \geq dim[N(\mathbf{B})]$$
$$\implies r - dim[R(\mathbf{AB})] \geq r - dim[R(\mathbf{B})]$$
$$\implies dim[R(\mathbf{AB})] \leq dim[R(\mathbf{B})]$$

But from 1.a we know that,

$$dim[R(\mathbf{AB})] \leq dim[R(\mathbf{A})]$$
$$\implies dim[R(\mathbf{AB})] \leq min(dim[R(\mathbf{A})], dim[R(\mathbf{B})])$$

And since $rank(\mathbf{A}) = dim[R(\mathbf{A})]$,

$$\implies \boxed{rank(\mathbf{AB}) \leq min(rank(\mathbf{A}), rank(\mathbf{B}))} \tag{4}$$

(f) **Prove** $rank(\mathbf{uu}^T) = 1, \mathbf{u} \in \mathbb{R}^n$

From 1.a, we know that $dim(R(\mathbf{uu}^T)) \leq dim(R(\mathbf{u}))$ and since $dim(R(\mathbf{u})) = 1$, $dim(R(\mathbf{uu}^T)) \leq 1$, which implies that $rank(\mathbf{uu}^T) \leq 1$ but for a vector to have rank $= 0$, all its elements should be 0.
Hence $\boxed{\forall \mathbf{u} \in \mathbb{R}^n - \{\mathbf{0}\}, rank(\mathbf{uu}^T) = 1}$ and for $\mathbf{u} = \mathbf{0}$, $rank(\mathbf{uu}^T) = 0$

(g) **Prove** row rank = column rank

We know that a matrix $\mathbf{A}$ can be written as $\mathbf{A} = \mathbf{CR}$, where the columns of $\mathbf{C}$ are the linearly independent columns of $\mathbf{A}$ and the rows of $\mathbf{R}$ are the respective coefficients which generate the matrix $\mathbf{A}$. We also know that every matrix multiplication can viewed in 2 ways i.e., linear combinations of the columns of the first matrix and also as the linear comibnations of the rows of the second matrix. Which implies that all the columns in the resultant matrix (here it's $\mathbf{A}$) are some linear combinations of the columns of the first matrix (here it's $\mathbf{C}$) and all the rows in the resultant matrix are some linear combinations

of the columns of the second matrix (here it's $\mathbf{R}$). This implies that dimension of range of rows of $\mathbf{A}$ is equal to the dimension of range of rows of $\mathbf{R}$ and dimension of range of columns of $\mathbf{A}$ is equal to the dimension of range of columns of $\mathbf{C}$. And since for it to be possible to multiply the matrices $\mathbf{C}$ and $\mathbf{R}$, the number of columns in $\mathbf{C}$ should be equal to the number of rows in $\mathbf{R}$. And since all rows in $\mathbf{R}$ and all columns in $\mathbf{C}$ are linearly independent, dimension of range of columns in $\mathbf{C}$ is equal to the dimension of range of rows in $\mathbf{R}$.

$$\therefore \boxed{\text{row rank of a matrix} = \text{column rank of a matrix}}$$

Problem 2

(a) Given that, there always exists a set of real coefficients $c_1, c_2, c_3, ...c_{10}$ for any set of real numbers $d_1, d_2, d_3, ...d_{10}$, such that

$$\sum_{j=1}^{10} c_j f_j(i) = d_i, \text{ where } i \in \{1, 2, ...10\}$$

This can be framed as a matrix vector multiplication, i.e.,

$$\text{Let } \mathbf{F} = \begin{pmatrix} f_1(1) & f_2(1) & ... & f_{10}(1) \\ f_1(2) & f_2(2) & ... & f_{10}(2) \\ ... & ... & ... & ... \\ f_1(10) & f_2(10) & ... & f_{10}(10) \end{pmatrix}, \mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ ... \\ c_{10} \end{pmatrix}, \mathbf{d} = \begin{pmatrix} d_1 \\ d_2 \\ ... \\ d_{10} \end{pmatrix}$$

$$\implies \mathbf{Fc} = \mathbf{d}$$

Since there always exists a set of real coefficients $c_1, c_2, ...c_{10}$ for any set of real numbers $d_1, d_2, ..., d_{10}$ (i.e., $\exists$ $\mathbf{c}$ for every $\mathbf{d}$). It implies that the range of the matrix $\mathbf{F}$ is atleast 10. And since $\mathbf{F}$ is a 10×10 matrix, for it to have a range of atleast 10, there should be atleast 10 linearly independent columns. Hence we can say that all the columns of $\mathbf{F}$ are linearly independent (i.e., full rank matrix) and since it is a square matrix, we can also say that it is a non-singular matrix.

We know that for a full rank matrix, the null space is 0. Hence for every unique $\mathbf{c}$ there exists a unique $\mathbf{d}$. And since $\mathbf{F}$ is non-singular, it has an inverse which can be used to uniquely determine $\mathbf{c}$ when $\mathbf{d}$ is known.

$$\text{i.e., } \boxed{\mathbf{c} = \mathbf{F}^{-1}\mathbf{d}}, \text{ where } \mathbf{F}_{i,j} = f_j(i), \mathbf{c}_j = c_j, \mathbf{d}_i = d_i$$

(b) Given that $\mathbf{A}$ is 10×10 matrix representing linear mapping from $\mathbf{d}$ (i.e., $d_1, d_2, ..., d_{10}$) to $\mathbf{c}$ (i.e., $c_1, c_2, ..., c_{10}$)

$$\implies \mathbf{Ad} = \mathbf{c} \implies \mathbf{d} = \mathbf{A}^{-1}\mathbf{c}$$
$$\implies \mathbf{A}^{-1}\mathbf{c} = \mathbf{d}$$

We know from the solution of *Problem 2 (a)* that,

$$\mathbf{Fc} = \mathbf{d} \implies \mathbf{A}^{-1} = \mathbf{F}$$

Since we know that

$$\mathbf{F}_{i,j} = f_j(i) \implies \mathbf{A}_{i,j}^{-1} = f_j(i)$$
$$\boxed{\therefore \text{ i,j th entry of } \mathbf{A}^{-1} = f_j(i)}$$

Problem 3

(a) Prove that $\mathbf{Q}$ is orthogonal given that ,

$$\mathbf{S}^T = -\mathbf{S}, \ \mathbf{Q} = (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S}) \tag{5}$$

We know that for an orthogonal matrix $\mathbf{Q}$,

$$\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I} \tag{6}$$

Computing $\mathbf{Q}^T$,

$$
\begin{aligned}
\mathbf{Q}^T &= ((\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S}))^T \\
&= (\mathbf{I}^T + \mathbf{S}^T)(\mathbf{I}^T - \mathbf{S}^T)^{-1} \\
\implies \mathbf{Q}^T &= (\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S})^{-1}
\end{aligned}
\tag{7}
$$

Multiplying equations 5 and 7,

$$
\begin{aligned}
\mathbf{Q}\mathbf{Q}^T &= (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S})(\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S})^{-1} \\
&= (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S} - \mathbf{S} - \mathbf{S}\mathbf{S})(\mathbf{I} + \mathbf{S})^{-1} \\
&= (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I}(\mathbf{I} + \mathbf{S}) - \mathbf{S}(\mathbf{I} + \mathbf{S}))(\mathbf{I} + \mathbf{S})^{-1} \\
&= (\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S})(\mathbf{I} + \mathbf{S})^{-1} \\
\implies \mathbf{Q}\mathbf{Q}^T &= I
\end{aligned}
\tag{8}
$$

Similarly multiplying equations 7 and 5,

$$
\begin{aligned}
\mathbf{Q}^T\mathbf{Q} &= (\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S})^{-1}(\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S}) \\
&= (\mathbf{I} - \mathbf{S})((\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S}))^{-1}(\mathbf{I} + \mathbf{S}) \\
&= (\mathbf{I} - \mathbf{S})(\mathbf{I} + \mathbf{S} - \mathbf{S} - \mathbf{S}\mathbf{S})^{-1}(\mathbf{I} + \mathbf{S}) \\
&= (\mathbf{I} - \mathbf{S})(\mathbf{I}(\mathbf{I} - \mathbf{S}) + \mathbf{S}(\mathbf{I} - \mathbf{S}))^{-1}(\mathbf{I} + \mathbf{S}) \\
&= (\mathbf{I} - \mathbf{S})((\mathbf{I} + \mathbf{S})(\mathbf{I} - \mathbf{S}))^{-1}(\mathbf{I} + \mathbf{S}) \\
&= (\mathbf{I} - \mathbf{S})(\mathbf{I} - \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S})^{-1}(\mathbf{I} + \mathbf{S}) \\
\implies \mathbf{Q}^T\mathbf{Q} &= I
\end{aligned}
\tag{9}
$$

Hence as per equations 6, 8 and 9, it is proved that $\mathbf{Q}$ is orthogonal

(b) Prove that $\mathbf{u}^T\mathbf{A}\mathbf{u} = 0, \forall \mathbf{u} \in \mathbb{R}^m$ if and only if $\mathbf{A} = 0$, given that $\mathbf{A}$ is a symmetric matrix and it can be written as

$$\mathbf{A} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T, \ \text{where } \mathbf{D} \text{ is a diagonal matrix} \tag{10}$$

6

From equation (10),

$$\mathbf{u}^T \mathbf{A} \mathbf{u} = 0$$
$$\implies \mathbf{u}^T \mathbf{Q} \mathbf{D} \mathbf{Q}^T \mathbf{u} = 0$$

Let $\mathbf{x} = \mathbf{Q}^T \mathbf{u}$

$$\implies \mathbf{x}^T \mathbf{D} \mathbf{x} = 0$$

Let the diagonal elements in $\mathbf{D}$ be $d_1, d_2, ..., d_m$ and the elements of $\mathbf{x}$ be $x_1, x_2, ..., x_m$. Then the above equation can be transformed as

$$\begin{pmatrix} x_1 & x_2 & ... & x_m \end{pmatrix} \begin{pmatrix} d_1 & ... & 0 \\ \vdots & \ddots & \vdots \\ 0 & ... & d_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_m \end{pmatrix} = 0$$

$$\implies \begin{pmatrix} d_1 x_1 & d_2 x_2 & ... & d_m x_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_m \end{pmatrix} = 0$$

$$\implies \begin{pmatrix} d_1 & d_2 & ... & d_m \end{pmatrix} \begin{pmatrix} x_1 & ... & 0 \\ \vdots & \ddots & \vdots \\ 0 & ... & x_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ ... \\ x_m \end{pmatrix} = 0$$

$$\implies \begin{pmatrix} d_1 & d_2 & ... & d_m \end{pmatrix} \begin{pmatrix} x_1^2 \\ x_2^2 \\ ... \\ x_m^2 \end{pmatrix} = 0$$

For the above equation to be always zero, the null space of the matrix $\begin{pmatrix} d_1 & d_2 & ... & d_m \end{pmatrix}$ should be zero i.e., for i from 1 to m, $d_i = 0$. Which implies that the diagonal matrix $\mathbf{D}$ is a zero matrix and since $\mathbf{A} = \mathbf{Q} \mathbf{D} \mathbf{Q}^T$, $\mathbf{A}$ is also a zero matrix.

$$\therefore \boxed{\mathbf{A} = 0}$$

The reverse can be easily proved as if $\mathbf{A} = 0$, then

$$\mathbf{u}^T \mathbf{A} = 0$$
$$\implies \boxed{\mathbf{u}^T \mathbf{A} \mathbf{u} = 0}$$

(c)   Prove $\mathbf{u}^T\mathbf{S}\mathbf{u} = 0 \,\forall \mathbf{u} \in \mathbb{R}^m$ if and only if $\mathbf{S} + \mathbf{S}^T = 0$

Let $\mathbf{u}^T\mathbf{S}\mathbf{u} = 0$. Then,

$$(\mathbf{u}^T\mathbf{S}\mathbf{u})^T = 0$$
$$\implies \mathbf{u}^T\mathbf{S}^T\mathbf{u} = 0 \tag{11}$$
$$\implies \mathbf{u}^T(\mathbf{S} - \mathbf{S}^T)\mathbf{u} = 0$$

We know that any matrix can be written as the sum of a symmetric and a skew-symmetric matrix

$$\implies \mathbf{u}^T \left(\left(\frac{\mathbf{S} + \mathbf{S}^T}{2} \right) + \left(\frac{\mathbf{S} - \mathbf{S}^T}{2} \right) \right) \mathbf{u} = 0$$

From equation (11)

$$\implies \mathbf{u}^T \left(\left(\frac{\mathbf{S} + \mathbf{S}^T}{2} \right) \right) \mathbf{u} = 0$$
$$\implies \mathbf{u}^T \left(\mathbf{S} + \mathbf{S}^T \right) \mathbf{u} = 0$$

From 3.b we know that if $\mathbf{u}^T(\mathbf{S} + \mathbf{S}^T)\mathbf{u} = 0$ then $\mathbf{S} + \mathbf{S}^T = 0$

$$\boxed{\therefore \text{ if } \mathbf{u}^T\mathbf{S}\mathbf{u} = 0 \text{ then } \mathbf{S} \text{ is a skew-symmetric matrix}}$$

Let $\mathbf{S}$ be a skew-symmetric matrix (i.e., $\mathbf{S} + \mathbf{S}^T = 0$). Multiplying with $\mathbf{u}^T$ from left and with $\mathbf{u}$ from right we get

$$\mathbf{u}^T \left(\mathbf{S} + \mathbf{S}^T \right) \mathbf{u} = 0$$
$$\implies \mathbf{u}^T\mathbf{S}\mathbf{u} + \mathbf{u}^T\mathbf{S}^T\mathbf{u} = 0 \tag{12}$$

Since $\mathbf{u}^T\mathbf{S}\mathbf{u}$ is a 1 dimensional matrix, it is equal to its transpose

$$\mathbf{u}^T\mathbf{S}\mathbf{u} = \mathbf{u}^T\mathbf{S}^T\mathbf{u}$$

But from equation (12)

$$\mathbf{u}^T\mathbf{S}\mathbf{u} + \mathbf{u}^T\mathbf{S}\mathbf{u} = 0$$
$$\implies \mathbf{u}^T\mathbf{S}\mathbf{u} = 0$$

$$\boxed{\therefore \text{ if } \mathbf{S} \text{ is a skew-symmetric matrix then } \mathbf{u}^T\mathbf{S}\mathbf{u} = 0}$$

Problem 4

Given that $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{A} \in \mathbb{R}^{m \times n}$

(a) **Prove** $||\mathbf{x}||_\infty \leq ||\mathbf{x}||_2$ Let x_k be the maximum element in $\mathbf{x}$. Then

$$||x||_\infty = |x_k| = \sqrt{x_k^2} \leq \sqrt{x_1^2 + ... + x_{k-1}^2 + x_k^2 + x_{k+1}^2 + ... + x_m} = ||x||_2$$

$$\boxed{||x||_\infty \leq ||x||_2}$$

The equality holds when all elements other than the maximum element are 0.

(b) **Prove** $||\mathbf{x}||_2 \leq \sqrt{m}||\mathbf{x}||_\infty$ Let x_k be the maximum element in $\mathbf{x}$. Then

$$||x||_2 = \sqrt{x_1^2 + ... + x_{k-1}^2 + x_k^2 + x_{k+1}^2 + ... + x_m} \leq \sqrt{x_k^2 + ... + x_k^2 + ... + x_k^2}$$

$$\implies ||x||_2 \leq \sqrt{mx_k^2} = \sqrt{m}|x_k| = \sqrt{m}||x||_\infty$$

$$\implies \boxed{||x||_2 \leq \sqrt{m}||x||_\infty}$$

The equality holds when all elements are equal.

(c) **Prove** $||\mathbf{A}||_\infty \leq \sqrt{n}||\mathbf{A}||_2$ We know that

$$||\mathbf{A}||_2 = \max_{x, ||x|| \neq 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \tag{13}$$

From (a) and (b) we know that,

$$||\mathbf{Ax}||_2 \geq ||\mathbf{Ax}||_\infty$$
$$||\mathbf{x}||_2 \leq \sqrt{n}||\mathbf{x}||_\infty, \text{ as } \mathbf{x} \text{ is } n \times 1 \text{ vector}$$

By substituting this in (14), we get

$$||\mathbf{A}||_2 = \max_{x, ||x|| \neq 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \geq \max_{x, ||x|| \neq 0} \frac{||\mathbf{Ax}||_\infty}{\sqrt{n}||\mathbf{x}||_\infty} = \frac{1}{\sqrt{n}}||\mathbf{A}||_\infty$$

$$\implies \boxed{||\mathbf{A}||_\infty \leq \sqrt{n}||\mathbf{A}||_2}$$

(d) **Prove** $||\mathbf{A}||_2 \leq \sqrt{m}||\mathbf{A}||_\infty$ We know that

$$||\mathbf{A}||_2 = \max_{x, ||x|| \neq 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \tag{14}$$

From (a) and (b) we know that,

$$||\mathbf{Ax}||_2 \leq \sqrt{m}||\mathbf{Ax}||_\infty, \text{ as } \mathbf{Ax} \text{ is } m \times 1 \text{ vector}$$
$$||\mathbf{x}||_2 \geq ||\mathbf{x}||_\infty$$

By substituting this in (14), we get

$$||\mathbf{A}||_2 = \max_{x, ||x|| \neq 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \leq \max_{x, ||x|| \neq 0} \frac{\sqrt{m}||\mathbf{Ax}||_\infty}{||\mathbf{x}||_\infty} = \sqrt{m}||\mathbf{A}||_\infty$$
$$\implies \boxed{||\mathbf{A}||_2 \leq \sqrt{m}||\mathbf{A}||_\infty}$$

(e) **Prove** $||\mathbf{A}||_F = \sqrt{tr(\mathbf{A}^T\mathbf{A})}$

We know that for a matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$,

$$||\mathbf{A}_F|| = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2} \tag{15}$$

And,

$$\mathbf{A}^T\mathbf{A} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

$$\implies tr(\mathbf{A}^T\mathbf{A}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$$

$$\implies \sqrt{tr(\mathbf{A}^T\mathbf{A})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$$

$$\therefore \boxed{||\mathbf{A}||_F = \sqrt{tr(\mathbf{A}^T\mathbf{A})}}$$

(f) Prove $\frac{1}{\sqrt{m}}||\mathbf{A}||_1 \le ||\mathbf{A}||_2 \le \sqrt{n}||\mathbf{A}||_1$

Let $\mathbf{x}$ be $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}$, then

$$\left(\sum_{i=1}^{m} x_i\right)^2 = \left(\sum_{i=1}^{m} x_i^2 + 2\sum_{i}^{m}\sum_{j}^{m} x_i x_j\right)$$

$$\ge \left(\sum_{i=1}^{m} x_i^2\right) = ||\mathbf{x}||_2^2 \tag{16}$$

$$\implies ||\mathbf{x}||_1^2 \ge ||\mathbf{x}||_2^2$$

$$\implies ||\mathbf{x}||_1 \ge ||\mathbf{x}||_2$$

We know that arithmetic mean $\ge$ geometric mean,

$$\implies x_i x_j \le (x_i^2 + x_j^2)/2$$

$$\implies 2x_i x_j \le x_i^2 + x_j^2$$

$$\implies 2\sum_{i}^{m}\sum_{j}^{m} x_i x_j \le (m-1)\sum_{i}^{m} x_i^2$$

$$\implies \left(\sum_{i=1}^{m} x_i\right)^2 \le \left(\sum_{i=1}^{m} x_i^2 + (m-1)\sum_{i}^{m} x_i^2\right) = m\left(\sum_{i=1}^{m} x_i\right)^2 \tag{17}$$

$$\implies ||\mathbf{x}||_1^2 \le m||\mathbf{x}||_2^2$$

$$\implies ||\mathbf{x}||_1 \le \sqrt{m}||\mathbf{x}||_2$$

Hence from (16) and (17) we get,

$$||\mathbf{Ax}||_2 \le ||\mathbf{Ax}||_1$$

$$||\mathbf{Ax}||_2 \ge \frac{1}{\sqrt{m}}||\mathbf{Ax}||_1, \text{ as } \mathbf{Ax} \text{ is } m \times 1 \text{ vector}$$

$$||\mathbf{x}||_2 \ge \frac{1}{\sqrt{n}}||\mathbf{x}||_1, \text{ as } \mathbf{x} \text{ is } n \times 1 \text{ vector}$$

$$||\mathbf{x}||_2 \le ||\mathbf{x}||_1$$

By substituting these in (14) (definition of induced matrix norm), we get

$$\implies \max_{x,||x||\ne 0} \frac{||\mathbf{Ax}||_1}{\sqrt{m}||\mathbf{x}||_1} \le \max_{x,||x||\ne 0} \frac{||\mathbf{Ax}||_2}{||\mathbf{x}||_2} \le \max_{x,||x||\ne 0} \frac{\sqrt{n}||\mathbf{Ax}||_1}{||\mathbf{x}||_1}$$

$$\implies \boxed{\frac{1}{\sqrt{m}}||\mathbf{A}||_1 \leq ||\mathbf{A}||_2 \leq \sqrt{n}||\mathbf{A}||_1}$$

(g) **Prove** $||\mathbf{A}||_2 \leq \sqrt{||\mathbf{A}||_1 ||\mathbf{A}||_\infty}$

Problem 5

Python Code:

```python
import numpy as np
from numpy.random import randn
from numpy.linalg import norm

def custom_norm(x, p):
    if p in [1, 2, np.inf]:
        return norm(x, p)
    return (sum([abs(elem)**p for elem in x])**(1/p))[0]

def induced_norm(A, p = 1):
    norm_Ax = 0
    for _ in range(1000):
        temp = randn(2, 1)
        x = temp/norm(temp)
        norm_Ax = max(norm_Ax, custom_norm(A.dot(x), p))
    return norm_Ax

A = randn(100, 2)
for p in [1, 2, 3, 4, 5, 6, np.inf]:
    norm_A = norm(A, p) if p in [1, 2, np.inf] else "-"
    norm_A_empirical = induced_norm(A, p)
    print("p: {0}, norm_A_empirical: {1}, norm_A: {2}".format(
        p,
        norm_A_empirical,
        norm_A,
    ))
```

Results:

```
p:   1, norm_A_empirical: 87.146911949156820, norm_A: 82.279901918493500
p:   2, norm_A_empirical: 11.268824198434068, norm_A: 11.268895430353254
p:   3, norm_A_empirical:  6.248385753051110, norm_A: _
p:   4, norm_A_empirical:  4.863832530448054, norm_A: _
p:   5, norm_A_empirical:  4.258207420557078, norm_A: _
p:   6, norm_A_empirical:  3.930596904232219, norm_A: _
p: inf, norm_A_empirical:  3.211614621668861, norm_A:  4.369508729553038
```

Comments: As we can see from the results of 1st, 2nd and ∞ norm that the norm calculated empirically using a random numbers is very close to the actual norm. As per the definition of induced norm, it should be the maximum stretch that a matrix can apply to any vector. And the empirical results show us exactly that. The reason for them to be slightly less than the actual norm is because of the number of iterations that we have considered. The larger the number of iterations, the larger is the chance to pick the vector which is stretched the most. But in any case the empirical norm will never be more than that of the actual norm and that is clearly observed from the results above.