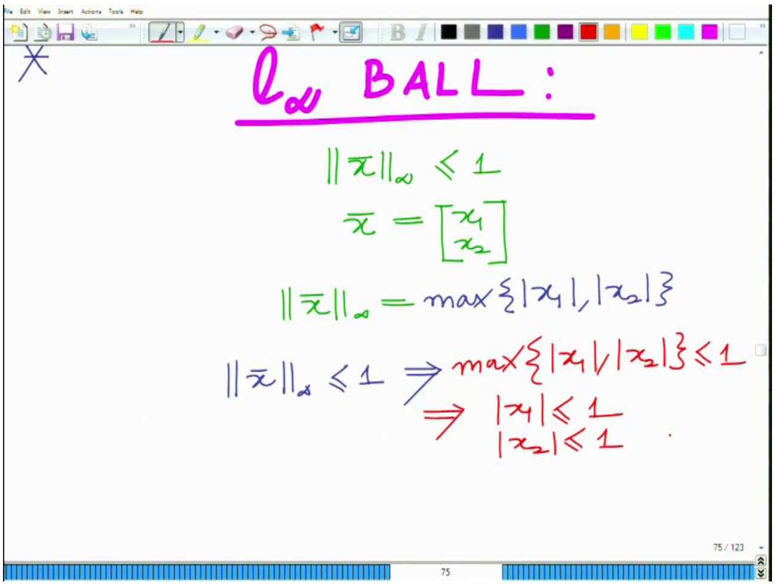
Applied Optimization for Wireless, Machine Learning, Big Data

Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 19
Norm balls and Matrix Properties: Trace, Determinant

Hello, welcome to another module in this massive open online course. Let us continue our discussion on the l_∞ balls.

(Refer Slide Time: 00:35)



The l_∞ norm ball is defined as follows.

$$\|\bar{x}\|_\infty \leq 1$$

Let us consider a 2-dimensional vector $\bar{x}$ as

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Then l_∞ norm is defined as the maximum of the absolute values of the components of that vector.

$$\|\bar{x}\|_\infty = \max\{|x_1|, |x_2|\}$$

Therefore l_∞ norm ball is defined as

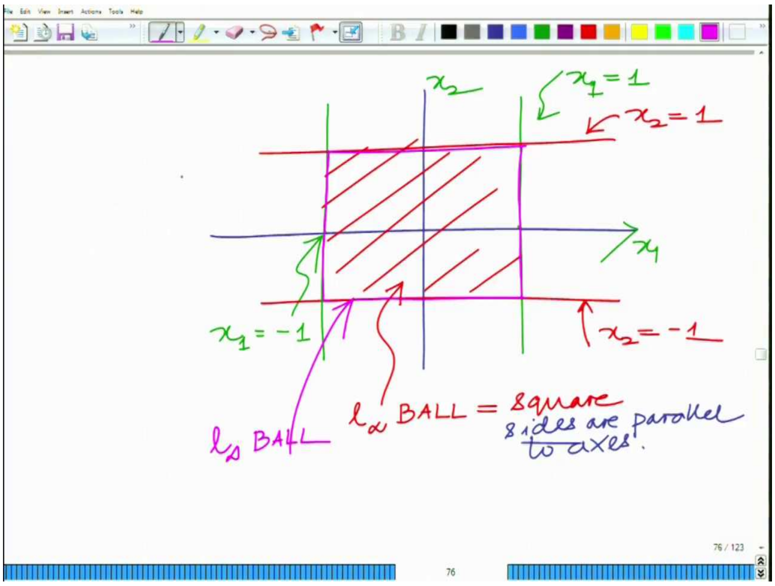
$$\|\bar{x}\|_\infty \leq 1$$
$$\max\{|x_1|, |x_2|\} \leq 1$$

This simply implies that

$|x_1| \leq 1$ and $|x_2| \leq 1$

Hence, l_∞ norm ball implies that each of the quantities of vector is less than or equal to 1.

(Refer Slide Time: 02:26)



This further can be simplified as

$$-1 \leq x_1 \leq 1, \text{ and}$$
$$-1 \leq x_2 \leq 1$$

Hence, consider the above simplified form of l_∞ norm ball for its graphical representation of l_∞ norm ball. So above simplification of l_∞ norm ball defines four half spaces which are as follows.

1. $x_1 \geq -1$,
2. $x_1 \leq 1$,
3. $x_2 \geq -1$,
4. $x_2 \leq 1$

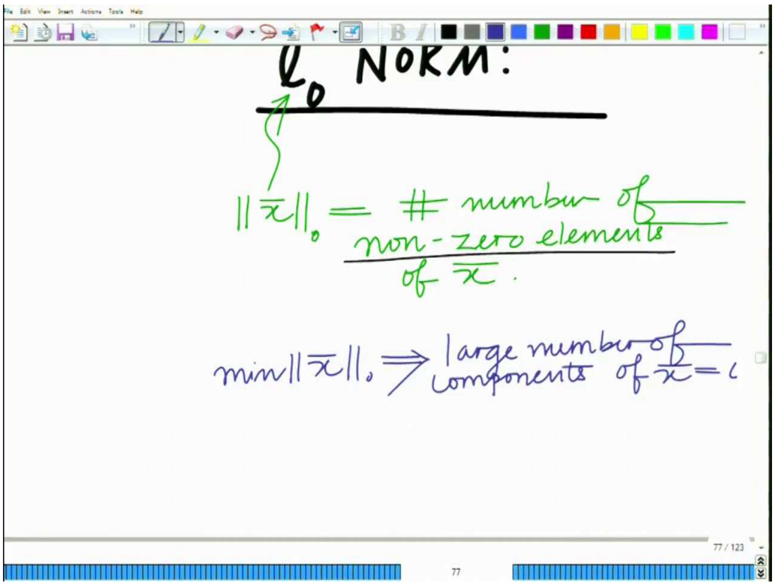
Where x_1 and x_2 are coordinate axes. So the four hyperplanes corresponding to these half spaces are

1. $x_1 = -1$,
2. $x_1 = 1$,
3. $x_2 = -1$,
4. $x_2 = 1$

When these hyperplanes are placed on the x_1 x_2 coordinate system and corresponding half spaces are mentioned then the intersection of these half spaces forms a square such that its sides are parallel to the axes.

Remember, the l_1 norm ball has tilted square shape which has the diagonals along the axes. Similar to this, the l_∞ norm ball also has square shape but it has the sides parallel to the axes. So, both the norm balls, l_1 and l_∞; have square shape but their orientation is different. This is interesting because balls are thought as of circles and spheres, but in case of l_1 and l_∞ norm balls, these are square.

(Refer Slide Time: 05:14)



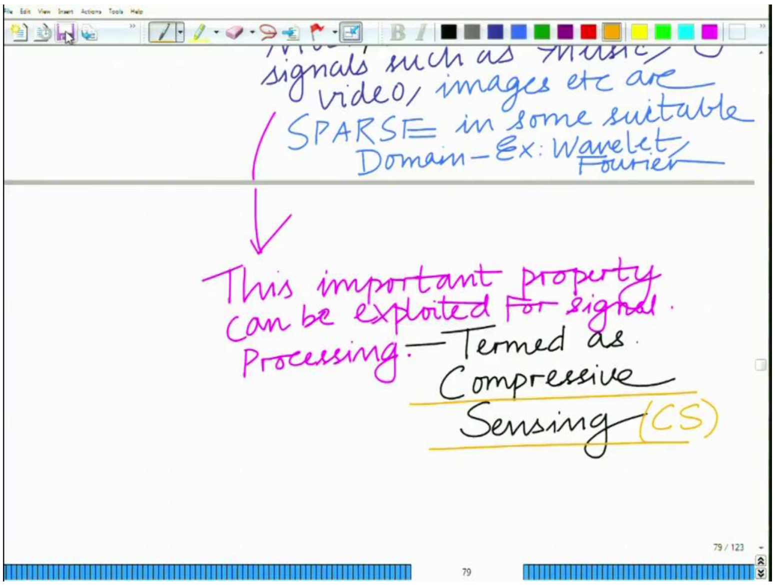
Now; let us discuss about l_0 norm. The l_0 norm of vector $\bar{x}$ i.e. $\|\bar{x}\|_0$; is equal to the number of non-zero elements of $\bar{x}$. So, if one minimizes the l_0 norm of vector $\bar{x}$ then it results in a large number of components of $\bar{x}$ will be zero and such a vector is commonly known as a Sparse Vector. A sparse vector basically denotes a vector in which there are only very few non-zero components, and a large number of components are zeros. A simple example of sparse vector is as follows.

$$\bar{x} = \begin{bmatrix} 0 \\ 0 \\ x_1 \\ 0 \\ x_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

In real world, such vectors have found its usage in case of sparsely populated area which means that there are very few people (or users) in the area.

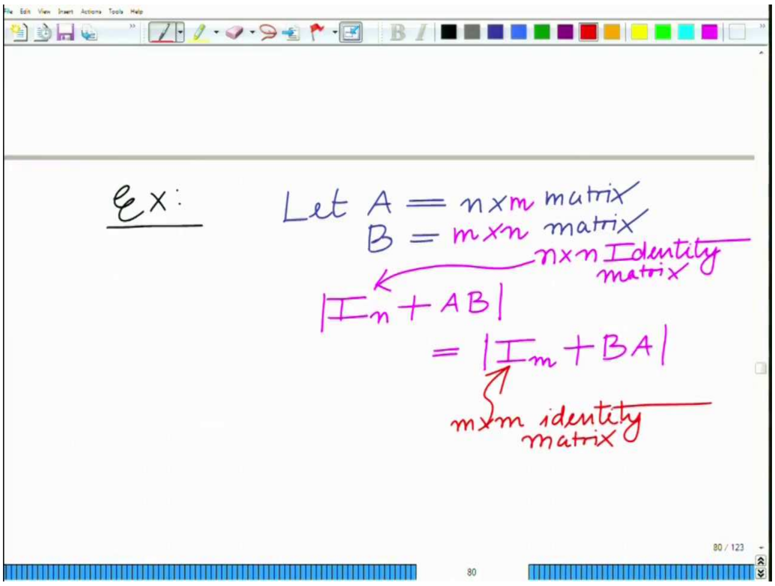
Most of the naturally occurring signals such as music, video or images are sparse under some suitable domain such as Fourier transform or wavelet. So they can be used as a sparse signal vector for signal processing to improve the performance of the signal.

(Refer Slide Time: 09:53)



This important property of such signals can be exploited for signal processing and this is termed as Compressive Sensing (CS). This is relatively a new field which recently has gained a lot of popularity.

(Refer Slide Time: 12:03)

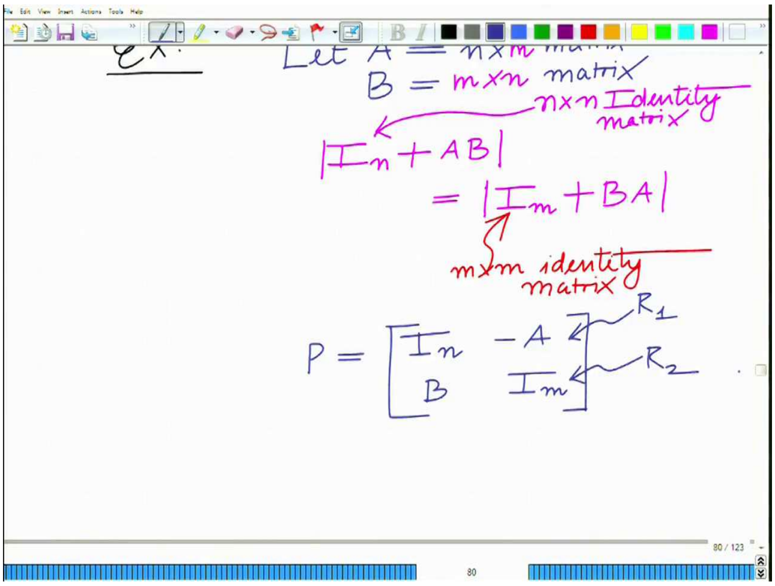


Let us have some simple examples related to matrices and their properties. So, let matrix A is $n \times m$ matrix and matrix B is $m \times n$ matrix. First property is,

$$|I_n + AB| = |I_m + BA|$$

Where I_n and I_m are $n \times n$ and $m \times m$ matrices respectively.

(Refer Slide Time: 13:56)



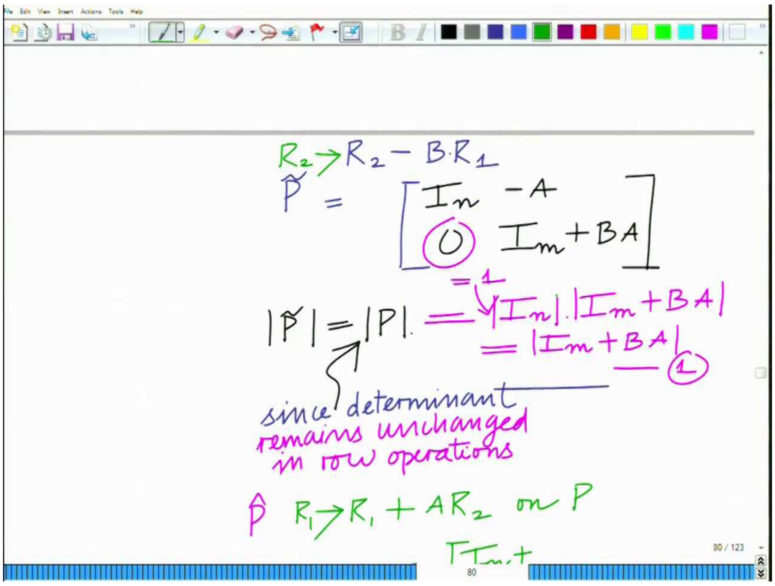
To verify this property, the solution is as follows.

Consider a matrix P given as

$$P = \begin{bmatrix} I_n & -A \\ B & I_m \end{bmatrix}$$

Let us perform block row operations. Here, the first row of above matrix P is block row 1 denoted by R_1 and similarly its second row is block row 2 denoted by R_2 .

(Refer Slide Time: 14:34)



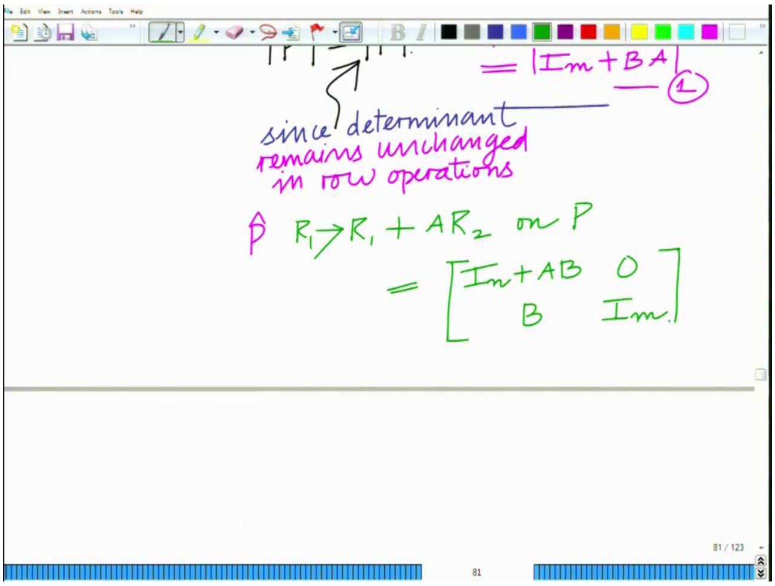
Now, perform $R_2 \to R_2 - BR_1$ on matrix P.

$$\tilde{P} = \begin{bmatrix} I_n & -A \\ 0 & I_m + BA \end{bmatrix}$$

Now as the block row operations do not change the determinant value, therefore the determinant of matrix $\tilde{P}$ equal to determinant of P. Hence;

$$\begin{aligned} |P| &= |\tilde{P}| \\ &= |I_n|.|I_m + BA| \\ &= |I_m + BA| \end{aligned}$$

(Refer Slide Time: 16:32)



Now, perform $R_1 \to R_1 + AR_2$ on matrix P.

$$\hat{P} = \begin{bmatrix} I_n + AB & 0 \\ B & I_m \end{bmatrix}$$

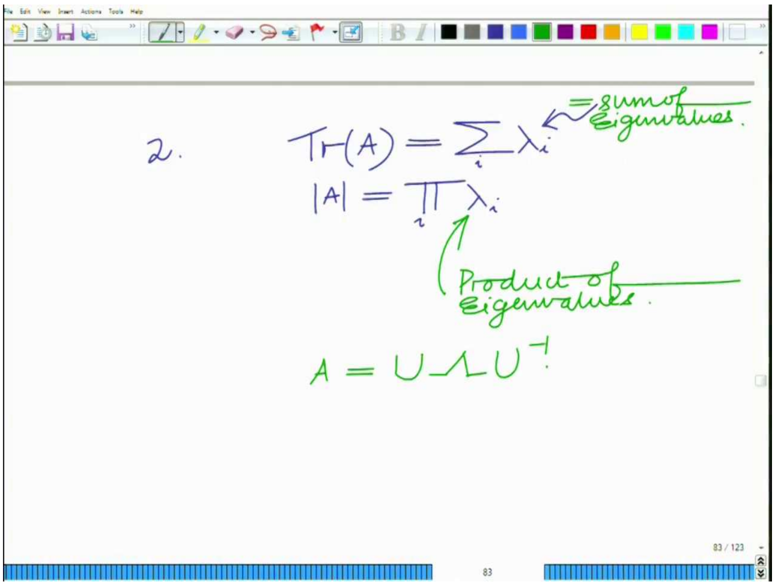
Once again, the determinant of matrix $\hat{P}$ equals to determinant of P. Hence;

$$\begin{aligned} |P| &= |\hat{P}| \\ &= |I_n + AB| . |I_m| \\ &= |I_n + AB| \end{aligned}$$

Hence from both the values of determinant of P, it will be concluded that

$$|I_n + AB| = |I_m + BA|$$

(Refer Slide Time: 19:56)



Another property of matrix is that the trace of a matrix is the sum of its eigenvalues. Hence for a matrix A with *i* eigenvalues λ_i;

$$Tr(A) = \sum_i \lambda_i$$

Further in this property, the determinant of a matrix is the product of its eigenvalues. So,

$$|A| = \prod_i \lambda_i$$

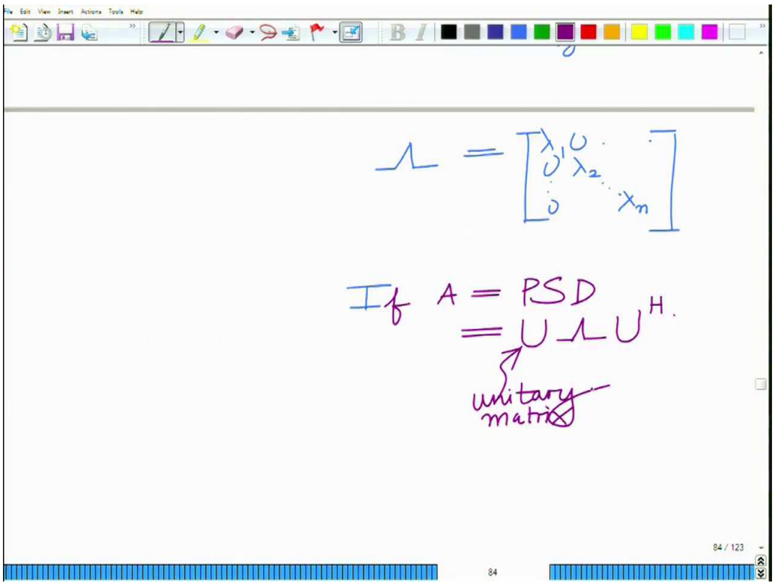
For verification of this property, consider a matrix A in the form of

$$A = U \Lambda U^{-1}$$

where U is the matrix of eigenvectors of A and Λ is the diagonal matrix of eigenvalues of A.

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

(Refer Slide Time: 21:33)

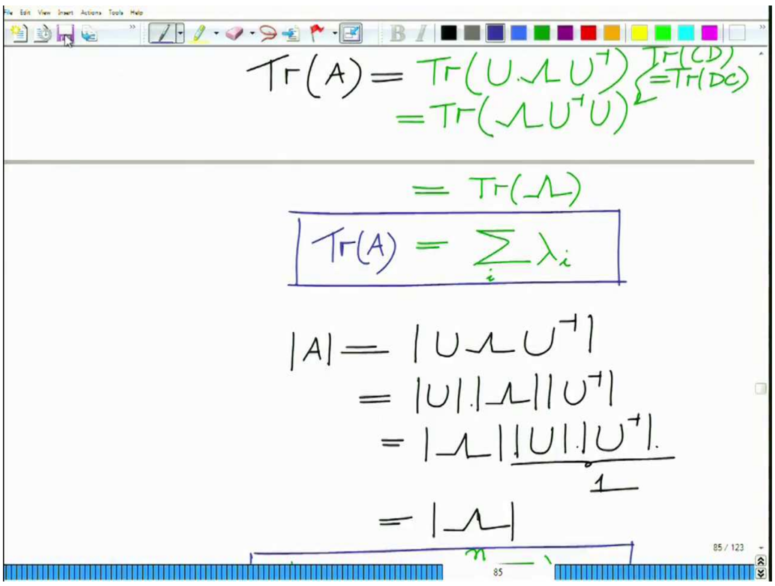


If A is a Positive Semi Definite matrix then it is known that

$$A = U \Lambda U^H$$

Here U is a unitary matrix so $U^{-1} = U^H$.

(Refer Slide Time: 22:39)



So the trace of matrix A is

$$Tr(A) = Tr\left(U \Lambda U^{-1}\right)$$

And for two marices C and D;

$$Tr(CD) = Tr(DC)$$

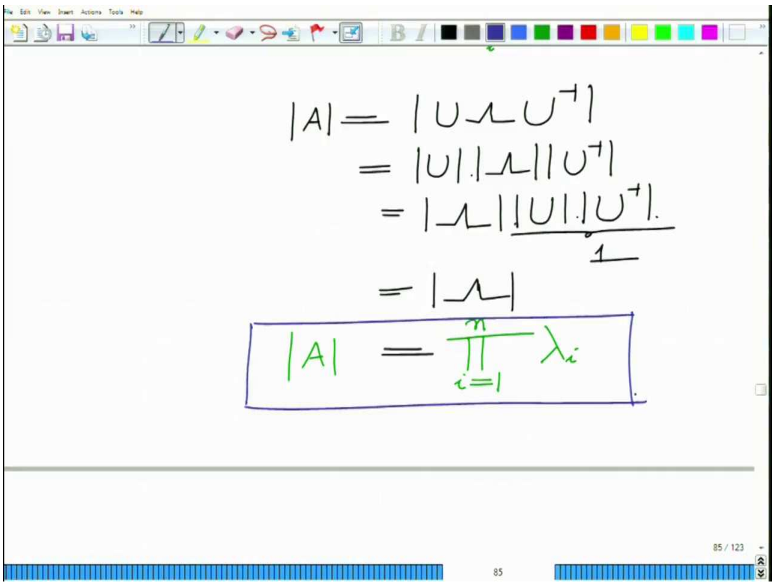
And also

$$CD = D^{-1}C$$

Therefore

$$\begin{aligned} Tr(A) &= Tr\left(\Lambda U^{-1} U\right) \\ &= Tr(\Lambda) \\ &= \sum_i \lambda_i \end{aligned}$$

(Refer Slide Time: 24:03)

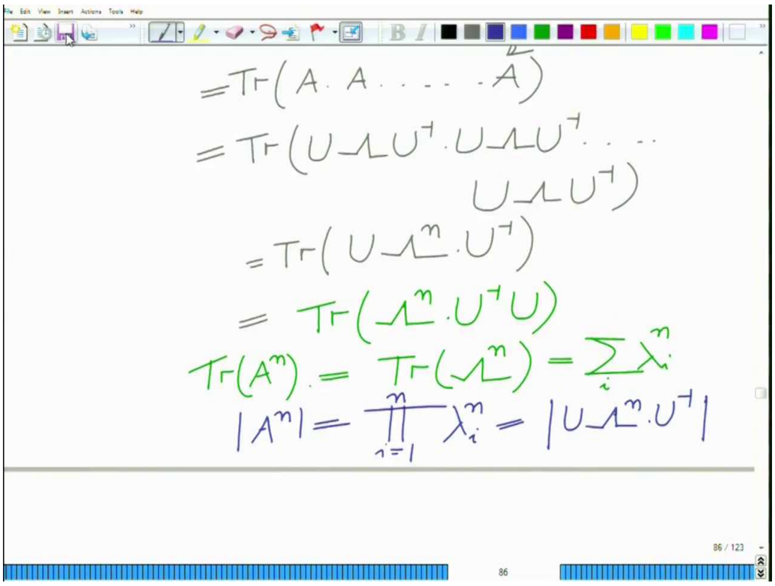


Similarly the determinant of A is

$$\begin{aligned} |A| &= \left|U \Lambda U^{-1}\right| \\ &= |U|.|\Lambda|.\left|U^{-1}\right| \\ &= |\Lambda|.|U|.\left|U^{-1}\right| \\ &= |\Lambda| \\ &= \prod_i \lambda_i \end{aligned}$$

This verifies the second property of matrix.

(Refer Slide Time: 25:32)



On expending this property further, one can see that

$$Tr\left(A^n\right) = \sum_i \lambda_i^n$$

This can be shown as

$$\begin{aligned} Tr\left(A^n\right) &= Tr\left(\underbrace{A.A.\cdots.A}_{n \text{ times}}\right) \\ &= Tr\left(\underbrace{U\Lambda U^{-1}.U\Lambda U^{-1}.\cdots.U\Lambda U^{-1}}_{n \text{ times}}\right) \\ &= Tr\left(U\Lambda^n U^{-1}\right) \\ &= Tr\left(\Lambda^n\right) \\ &= \sum_i \lambda_i^n \end{aligned}$$

Similarly,

$$\left|A^n\right| = \prod_{i=1}^{n} \lambda_i^n = \left|U\Lambda^n U^{-1}\right|$$

So, these are some interesting properties. Let us continue to this discussion by looking at other examples in the subsequent modules.