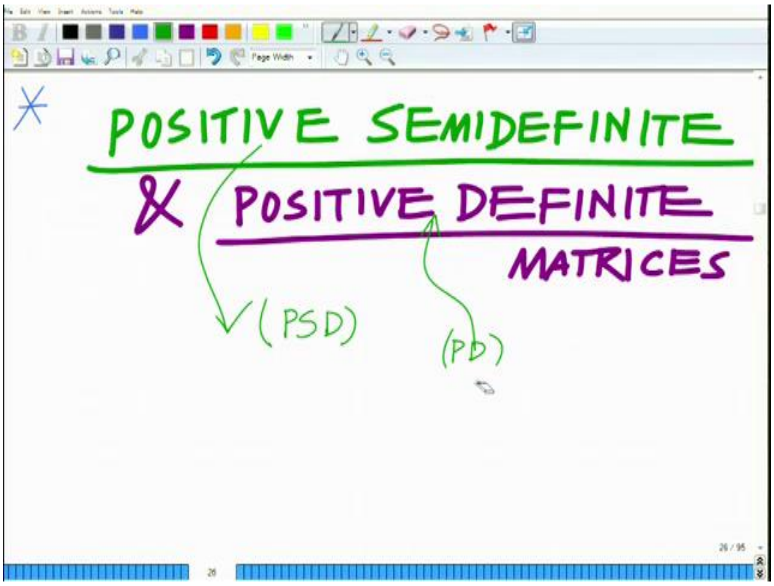
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 03
Positive Semi Definite (PSD) and Positive Definite (PD) Matrices and their Properties

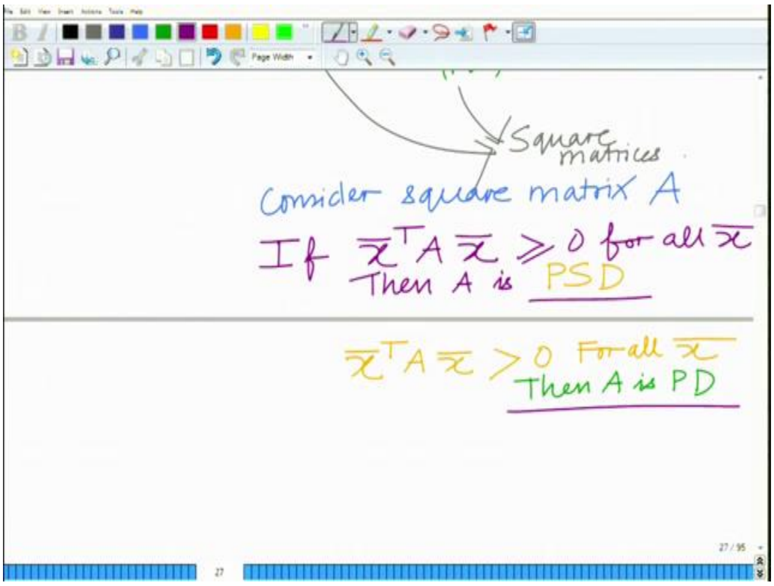
Hello, welcome to another module in this massive open online course. So, we are looking at the mathematical preliminaries for optimization. We have looked at the Eigenvectors and Eigen values and now we will start looking at a different type of matrices known as positive semi definite and positive definite matrices.

(Refer Slide Time: 00:35)



So, we are going to look at the definition and properties of Positive Semi Definite (PSD) matrix and Positive Definite (PD) matrix. A matrix can be Positive Semi Definite matrix and Positive Definite matrix only if it is a square matrix.

(Refer Slide Time: 01:51)

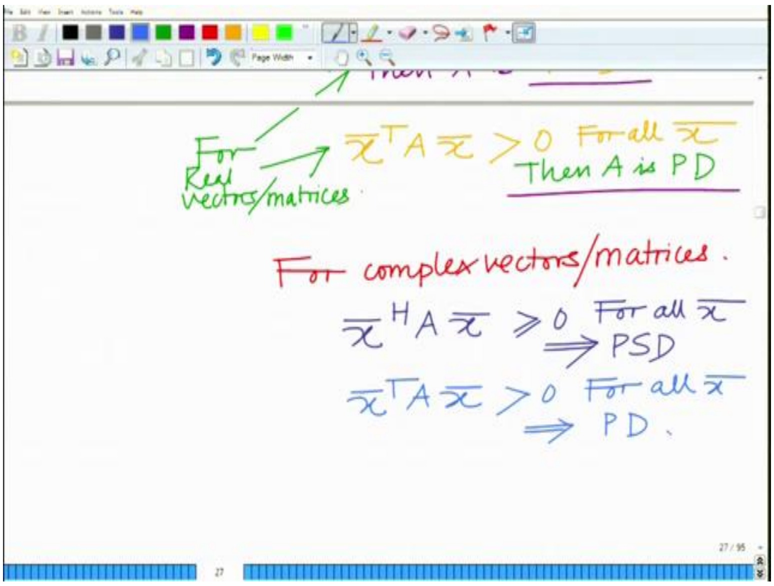


Consider a square matrix A. Now; for all real vectors $\bar{x}$;

if $\bar{x}^T A\bar{x} \geq 0$; then A is a positive semi definite (PSD) matrix.

if $\bar{x}^T A\bar{x} > 0$; then A is a positive definite (PD) matrix .

(Refer Slide Time: 03:45)

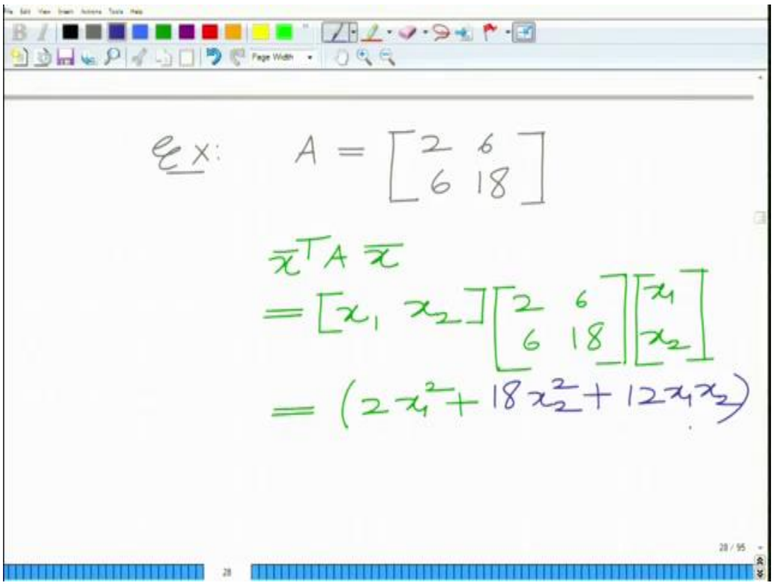


Now for complex vectors or matrices, we have to replace the transpose matrix by the Hermitian matrix. Therefore; Similarly, for all complex vectors $\bar{x}$;

if $\bar{x}^H A\bar{x} \geq 0$; then A is a positive semi definite (PSD) matrix.

if $\bar{x}^H A\bar{x} > 0$; then A is a positive definite (PD) matrix .

(Refer Slide Time: 04:59)



Let us take a simple example to understand this. Consider a square matrix A as follows.

$$A = \begin{pmatrix} 2 & 6 \\ 6 & 18 \end{pmatrix}$$

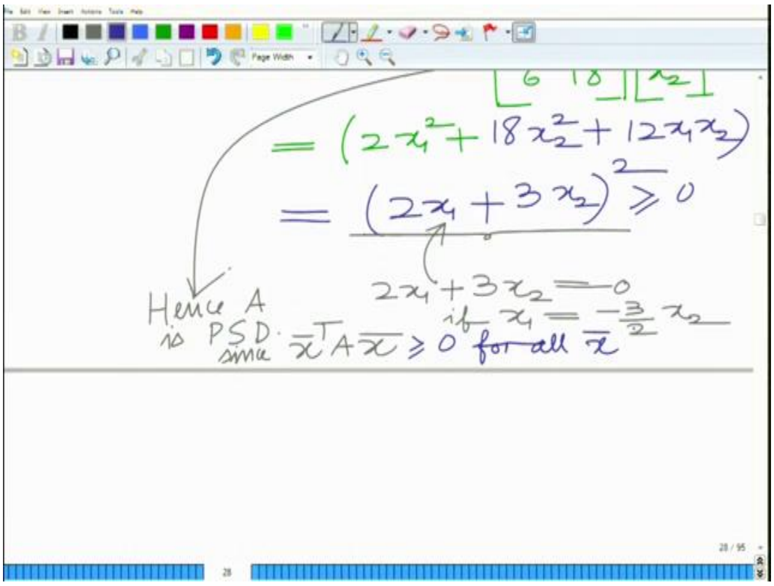
Now on calculating $\bar{x}^H A\bar{x}$; we get

$$\begin{aligned} \bar{x}^H A\bar{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{pmatrix} 2 & 6 \\ 6 & 18 \end{pmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= 2x_1^{\;2} + 18x_2^{\;2} + 12x_1x_2 \\ &= \left(2x_1 + 3x_2\right)^2 \end{aligned}$$

Now as $\bar{x}^H A\bar{x}$ is a square so it is always greater than or equal to 0. That is

$$\bar{x}^H A\bar{x} = \left(2x_1 + 3x_2\right)^2 \geq 0$$

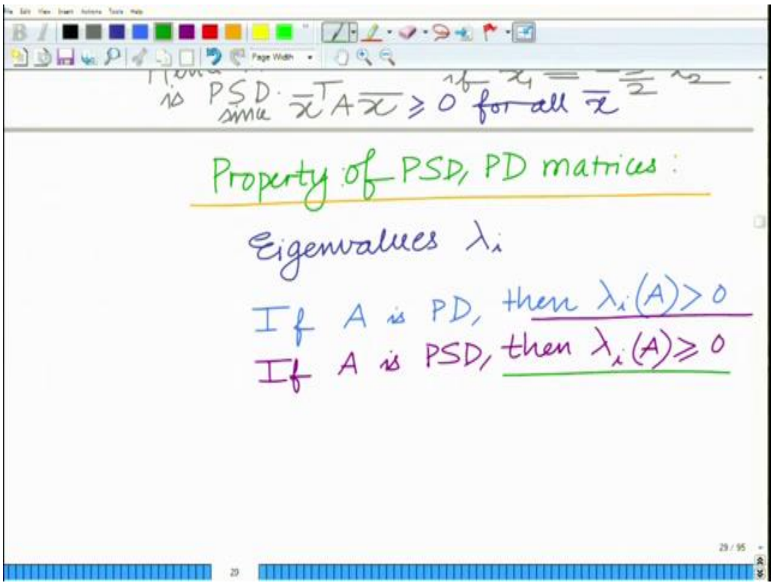
(Refer Slide Time: 06:26)



Hence, this is not strictly greater than 0, if $x_1 = -\frac{3}{2}x_2$. Therefore we can say that if $x_1 = -\frac{3}{2}x_2$; the matrix is a positive semi definite (PSD) matrix for all $\bar{x}$.

Now let us look at a property of this positive semi definite matrix.

(Refer Slide Time: 07:58)



Let us look at an interesting property of this. Now, consider the *i* number of Eigen values λ_i of a square matrix A. So if A is a positive definite matrix then all of its Eigen values

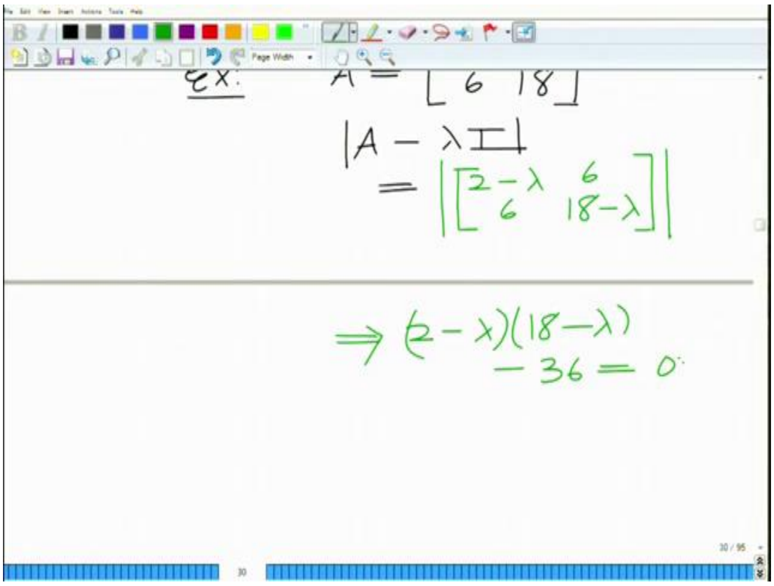
$\lambda_i(A)$ are strictly greater than 0. On the other hand, if A is positive semi definite matrix then all of its Eigen values $\lambda_i(A)$ are greater than or equal to 0; that is some of the Eigen values can be 0 and rest of them are greater than 0. So we can write it as follows.

If A is PD; Then $\lambda_i(A) > 0$.

If A is PSD; Then $\lambda_i(A) \geq 0$.

Now, let us check this property on the previous example.

(Refer Slide Time: 09:36)



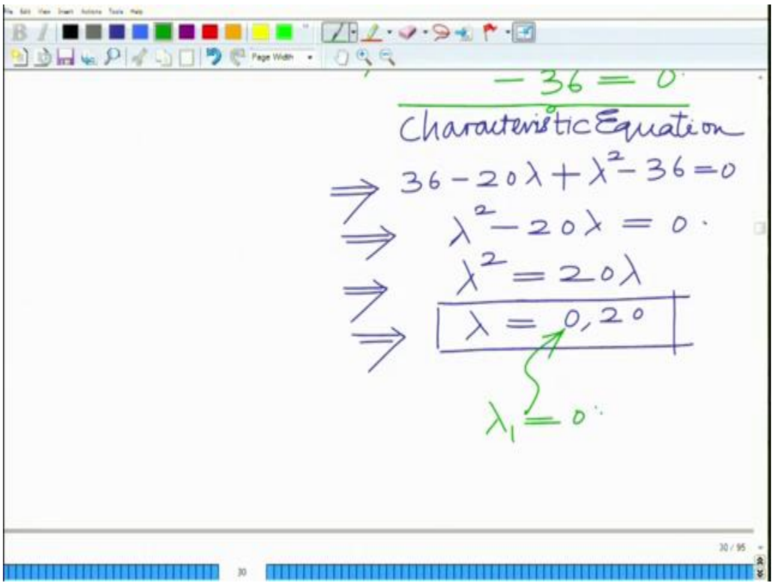
So, let us take matrix A.

$$A = \begin{pmatrix} 2 & 6 \\ 6 & 18 \end{pmatrix}$$

Now, to calculate the Eigen values, consider the characteristic polynomial.

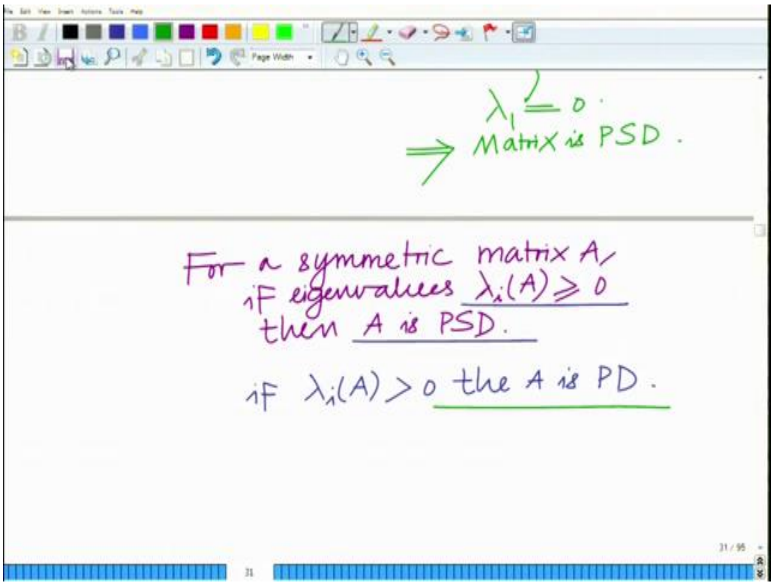
$$|A - \lambda I| = 0$$
$$\left| \begin{pmatrix} 2-\lambda & 6 \\ 6 & 18-\lambda \end{pmatrix} \right| = 0$$
$$(2-\lambda)(18-\lambda) - 36 = 0$$
$$\lambda^2 - 20\lambda = 0$$
$$\lambda = 0, 20$$

(Refer Slide Time: 10:55)



So the two Eigen values are 0 and 20. Here we can see that one of the Eigen values is 0. This implies that this matrix A is a positive semi definite matrix.

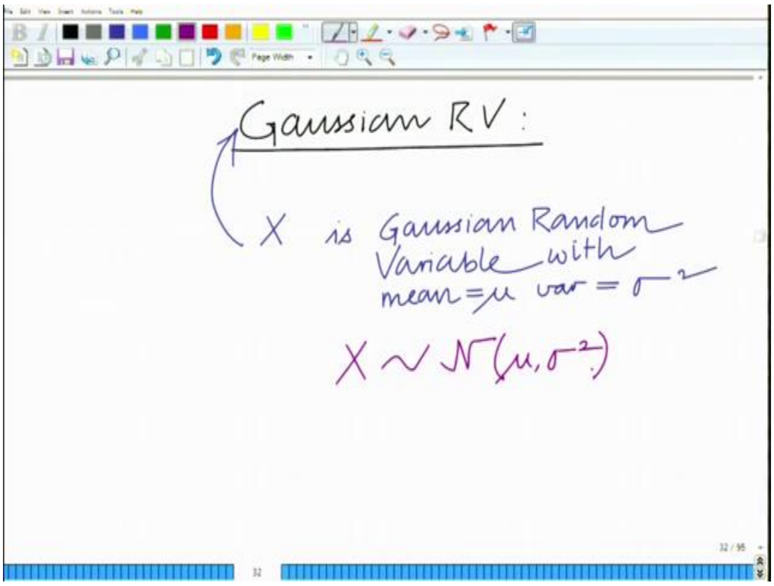
(Refer Slide Time: 12:13)



The reverse of this is also true which means that, for a symmetric matrix, if the Eigen values of a matrix are greater than or equal to 0, the matrix is a positive semi definite matrix and if all the Eigen values of a matrix are greater than 0 then the matrix is positive definite matrix.

Now lets look at another important concept that is the Gaussian random variable, which we are also going to use frequently in our framework of optimization.

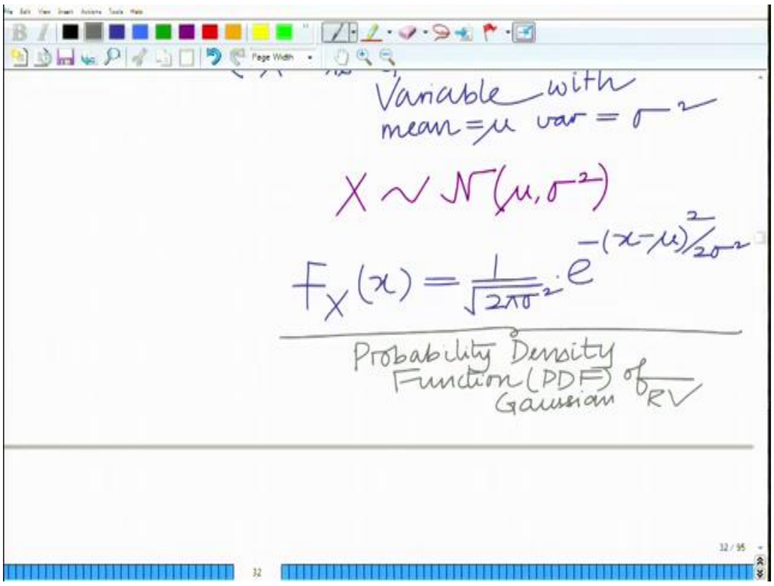
(Refer Slide Time: 14:02)



So, look at the basic concepts of Gaussian Random Variables. Let X is a Gaussian Random Variable with mean equal to μ and variance equal to σ^2. This is also known as a Normal Random Variable; is denoted by

$$X \sim \mathrm{N}\left(\mu, \sigma^2\right)$$

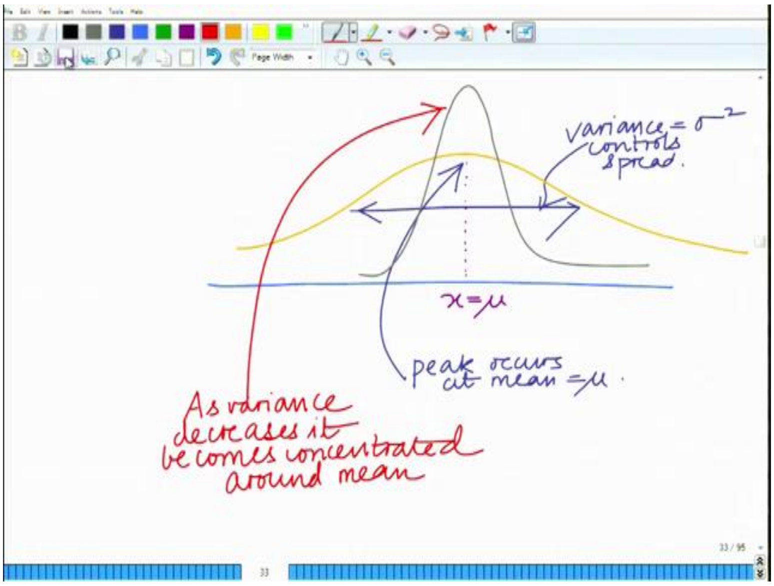
(Refer Slide Time: 15:11)



Every random variable has a probability density function (PDF) denoted by $F_X(x)$. The PDF of a Gaussian RV is given as

$$F_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

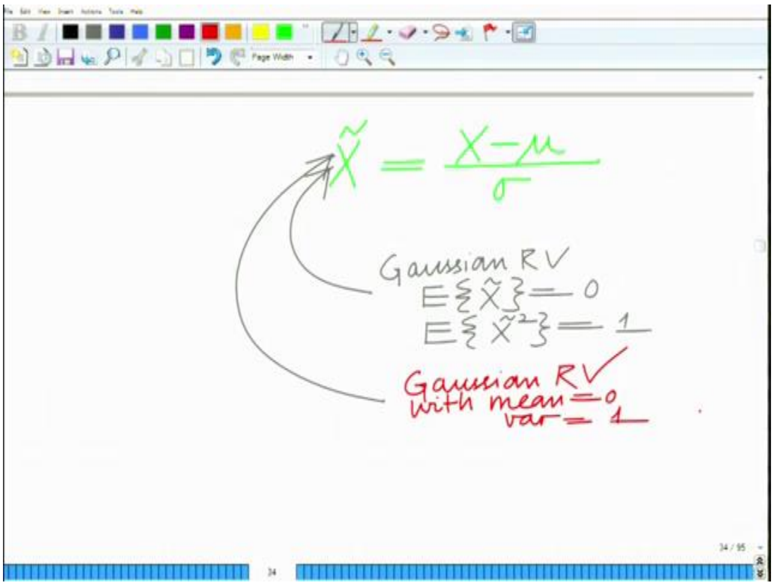
(Refer Slide Time: 16:11)



The probability density function of Gaussian RV is bell shaped curve with the peak occurring at mean $x = \mu$ and the spread is controlled by the variance σ^2.

So, the peak shifts in the Gaussian probability density function but the curve is always symmetric about the mean. For instance, if the variance decreases, this means the spread of the curve decreases and the Gaussian probability density function is more concentrated around the mean.

(Refer Slide Time: 18:13)



Now, one can define a new random variable $\tilde{X}$ such as

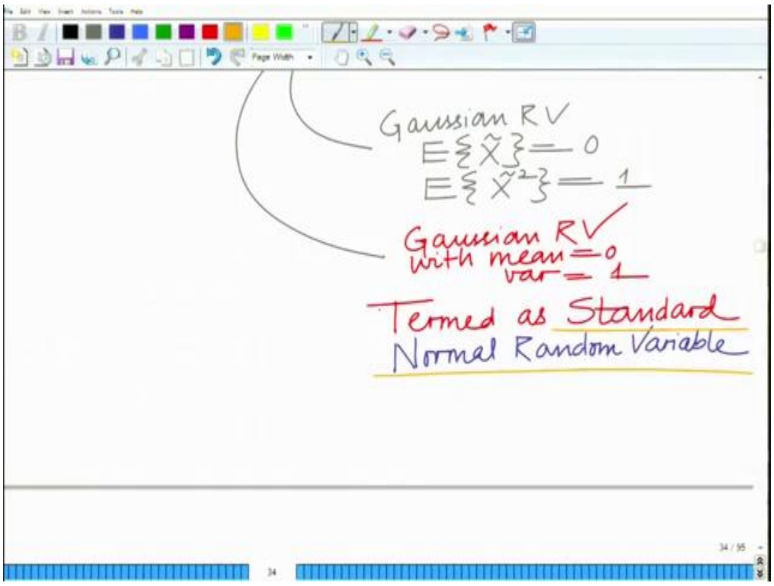
$$\tilde{X} = \frac{X - \mu}{\sigma}$$

Now, this $\tilde{X}$ is also a Gaussian RV and the mean of $\tilde{X}$ will be 0 that is the expected value of $\tilde{X}$ is 0. Also, the variance; that is the expected value of $\tilde{X}^2$ is equal to 1.

$$E\left\{\tilde{X}\right\} = 0$$
$$E\left\{\tilde{X}^2\right\} = 1$$

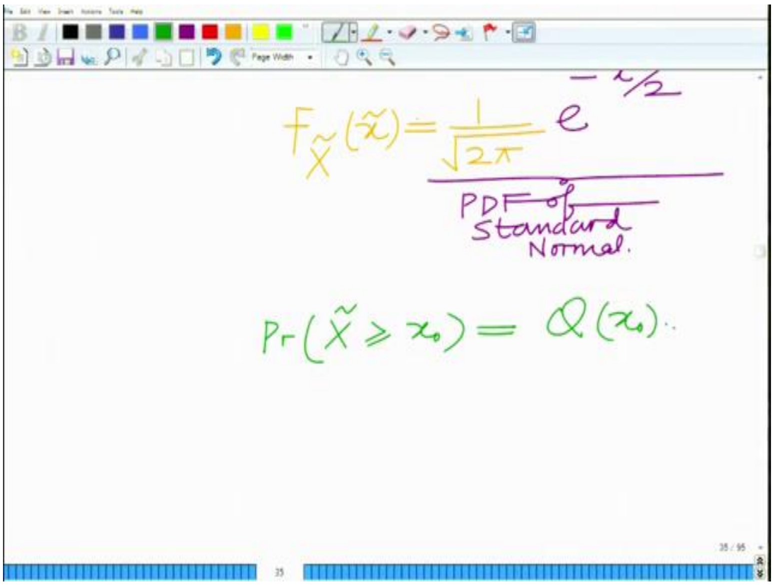
So, this is a Gaussian RV with mean equal to 0 and variance equal to 1.

(Refer Slide Time: 19:22)



This is termed as the standard normal random variable. And the standard normal random variable is used to define the Q function.

(Refer Slide Time: 20:04)



Hence, on substituting $\mu = 0$ and $\sigma^2 = 1$ in the earlier expression; the probability density function of the standard normal random variable is

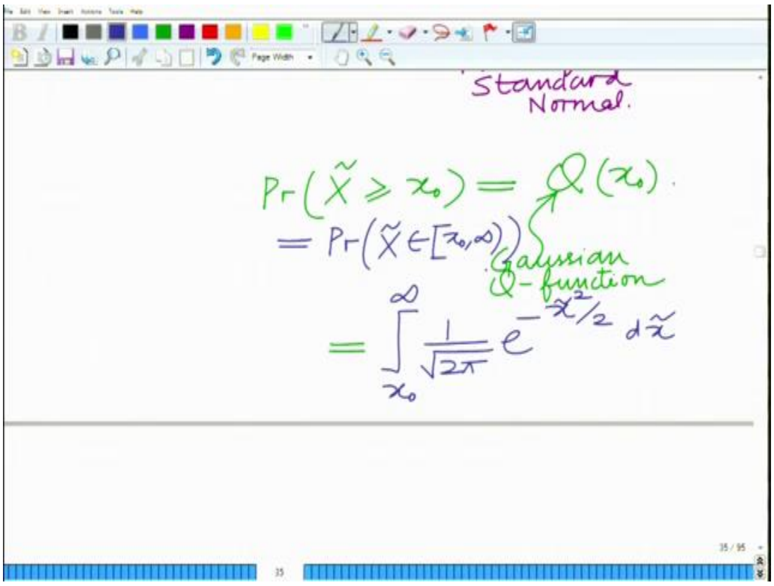
$$F_{\tilde{X}}(\tilde{x}) = \frac{1}{\sqrt{2\pi}} e^{\frac{-\tilde{x}^2}{2}}$$

Now the Q function of the PDF of the Standard Normal is the probability that Standard Normal Gaussian Random Variable $\tilde{X}$ is greater than or equal to a quantity x_0.

$$\text{Prob}\left(\tilde{X} \geq x_0\right) = \text{Q}\left(x_0\right)$$

This is also termed as the Gaussian Q function.

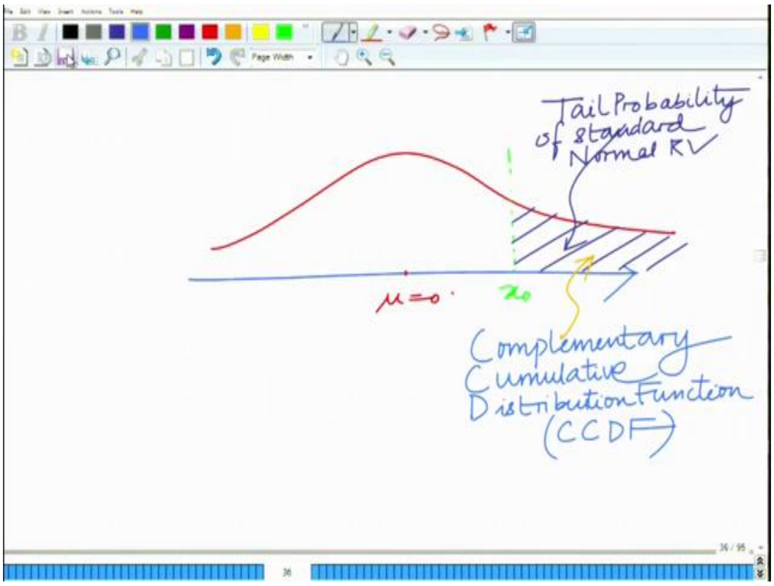
(Refer Slide Time: 21:16)



This is given by

$$\begin{aligned} \text{Q}\left(x_0\right) &= \text{Prob}\left(\tilde{X} \geq x_0\right) \\ &= \text{Prob}\left(\tilde{X} \in \left[x_0, \infty\right)\right) \\ &= \int_{x_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-\tilde{x}^2}{2}} d\tilde{x} \end{aligned}$$

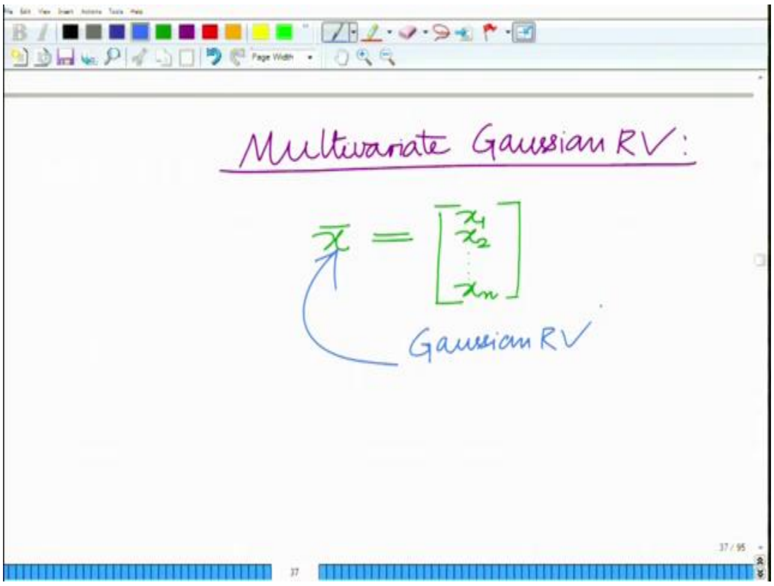
(Refer Slide Time: 22:55)



So, it denotes the probability that the area under the PDF greater than or equal to x_0. This is also known as the tail probability of the Standard Normal Random Variable and the Complementary Cumulative Density Function (CCDF). The Cumulative Density Function gives the probability that the random variable takes values less than x_0; the complement of that or 1 minus the CDF gives the probability that it is greater than or equal to x_0.

Now, let us come to the Multivariate Gaussian Random Variable or Multivariate Gaussian Random Vector. It is a Gaussian Random Vector with multiple components, each of them individually Gaussian and all of them being jointly Gaussian.

(Refer Slide Time: 25:01)

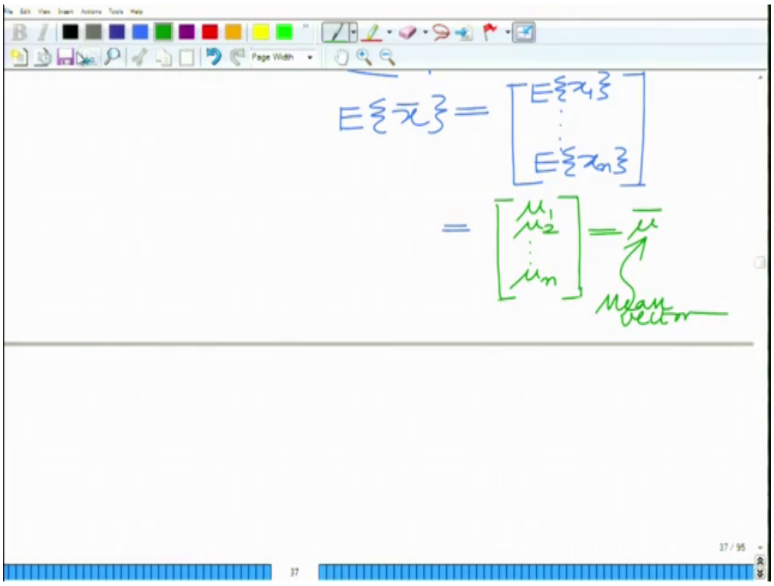


Now, a Multivariate Gaussian RV is given by

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

This is a Gaussian random vector having n components.

(Refer Slide Time: 25:54)



So, the mean of this is going to be a vector that is

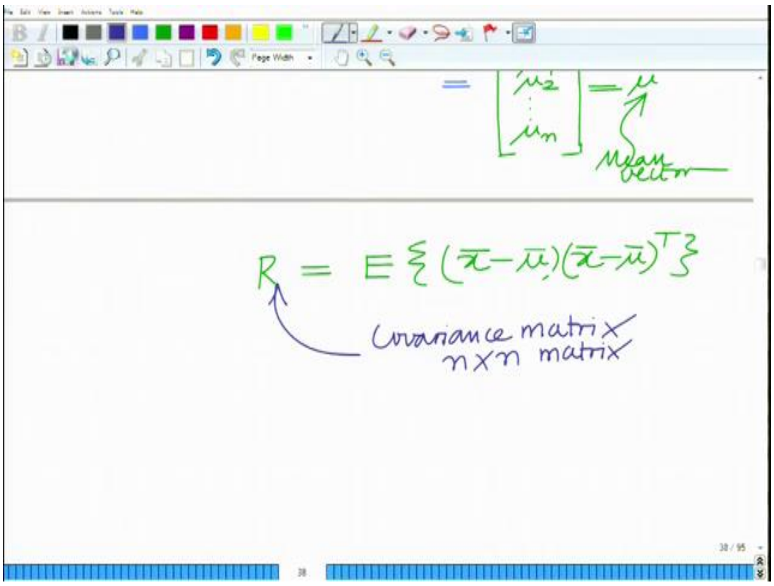
$$E\{\bar{x}\} = \begin{bmatrix} E\{x_1\} \\ E\{x_2\} \\ \vdots \\ E\{x_n\} \end{bmatrix}$$

$$= \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} = \bar{\mu}$$

So, this is basically the mean vector of the Multivariate Gaussian random variable.

Further instead of the variance, we will have the covariance matrix, which looks at the variance of each component and also the cross correlation.

(Refer Slide Time: 27:08)

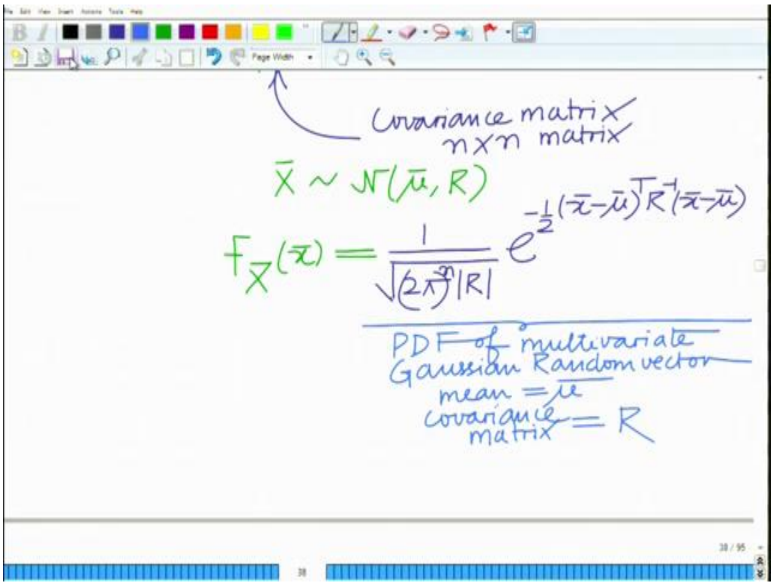


So the covariance matrix R is defined as follows.

$$R = E\left\{(\bar{x}-\bar{\mu})(\bar{x}-\bar{\mu})^T\right\}$$

This is the covariance matrix and is an $n \times n$ matrix.

(Refer Slide Time: 27:53)



So, the Multivariate Gaussian Random Variable is denoted as Gaussian Random Variable with mean vector $\bar{\mu}$ and covariance matrix R,

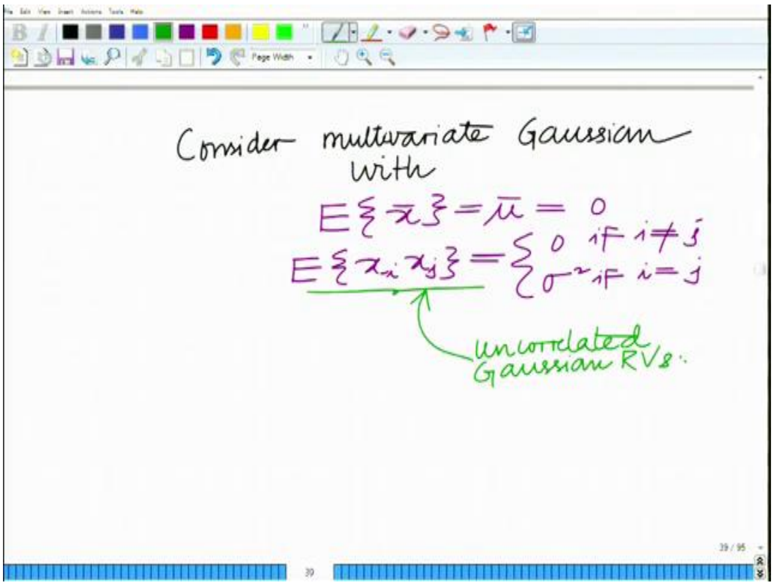
$$\bar{X} \sim \mathrm{N}\left(\bar{\mu}, R\right)$$

And the PDF of Multivariate Gaussian Vector is given as

$$F_{\tilde{X}}\left(\tilde{x}\right) = \frac{1}{\sqrt{\left(2\pi\right)^{n}\left|R\right|}} e^{\frac{-\left(\tilde{x}-\tilde{\mu}\right)^{T} R^{-1}\left(\tilde{x}-\tilde{\mu}\right)}{2}}$$

Let us look at an interesting special case of this Multivariate Gaussian Random Vector, which is when the different components of this Gaussian Random Vector are independent.

(Refer Slide Time: 30:04)



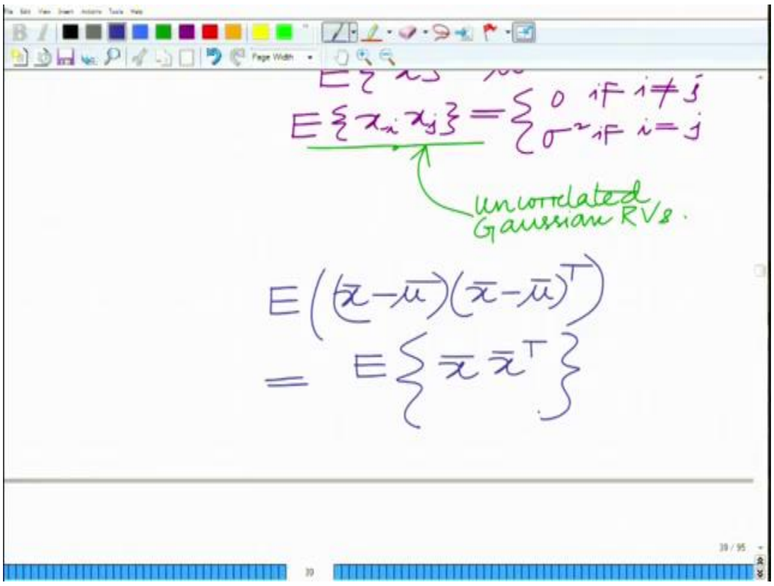
So, consider a Multivariate Gaussian with

$E\{\bar{x}\} = \bar{\mu} = 0$ and

$$E\{x_i x_j\} = \begin{cases} 0 & \text{if } i \neq j \\ \sigma^2 & \text{if } i = j \end{cases}$$

These are basically known as uncorrelated Gaussian random variables.

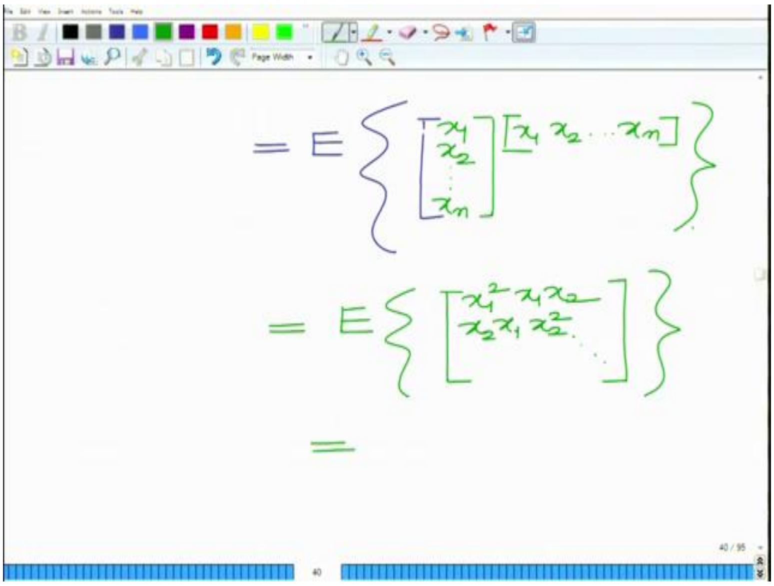
(Refer Slide Time: 31:50)



And now, if we compute the covariance matrix of this, that will be given as

$$R = E\left\{(\bar{x} - \bar{\mu})(\bar{x} - \bar{\mu})^T\right\}$$
$$= E\left\{\bar{x} \cdot \bar{x}^T\right\}$$

(Refer Slide Time: 32:17)

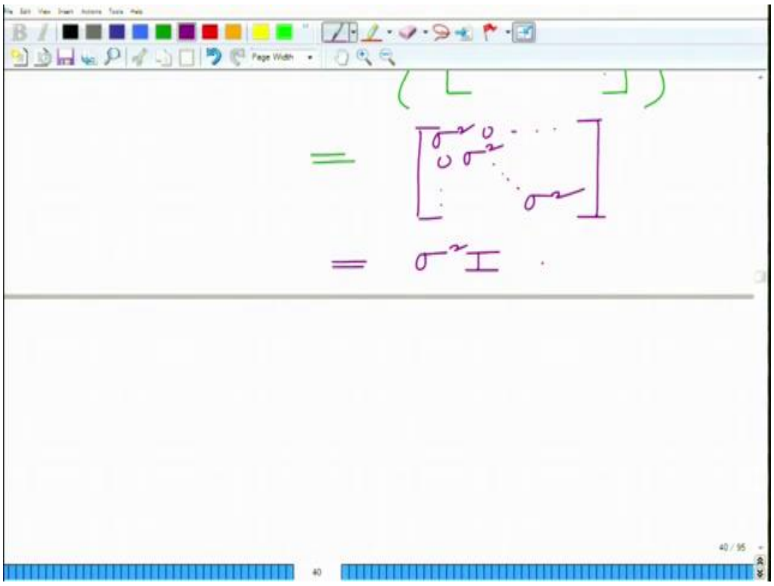


Therefore we can write

$$R = E\left\{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}\right\}$$
$$= E\left\{\begin{bmatrix} {x_1}^2 & x_1x_2 & \cdots \\ x_2x_1 & {x_2}^2 & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}\right\}$$

And now, if we look at this matrix, the expected value of each diagonal element is equals to σ^2 , and the expected values of the off diagonal entries are 0 as these are the uncorrelated random variables.

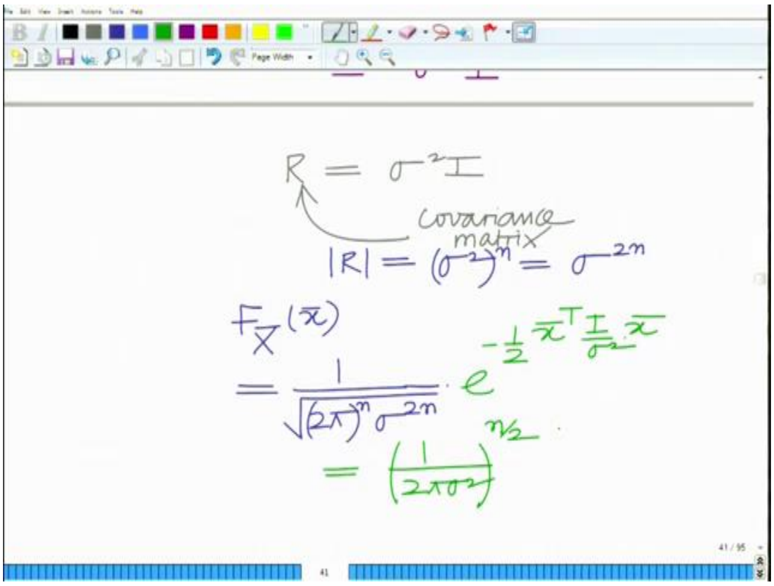
(Refer Slide Time: 33:29)



And therefore, the covariance matrix basically, you can see reduces to

$$R = \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$
$$= \sigma^2 I$$

(Refer Slide Time: 33:50)



Now, determinant of R is

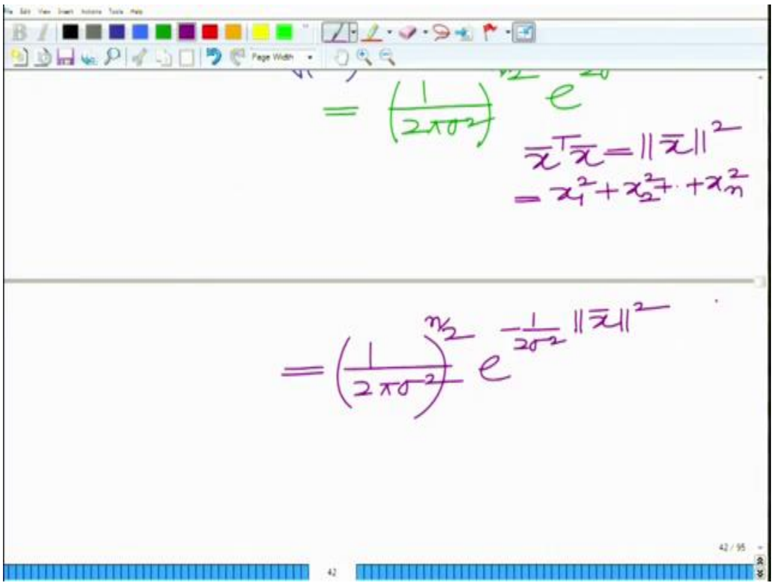
$$|R| = \left(\sigma^2\right)^n = \sigma^{2n}$$

And therefore, the probability density function is

$$F_{\bar{X}}\left(\bar{x}\right) = \frac{1}{\sqrt{\left(2\pi\right)^n \sigma^{2n}}} e^{\frac{-1}{2}\bar{x}^T \frac{I}{\sigma^2}\bar{x}}$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\|\bar{x}\|^2}$$

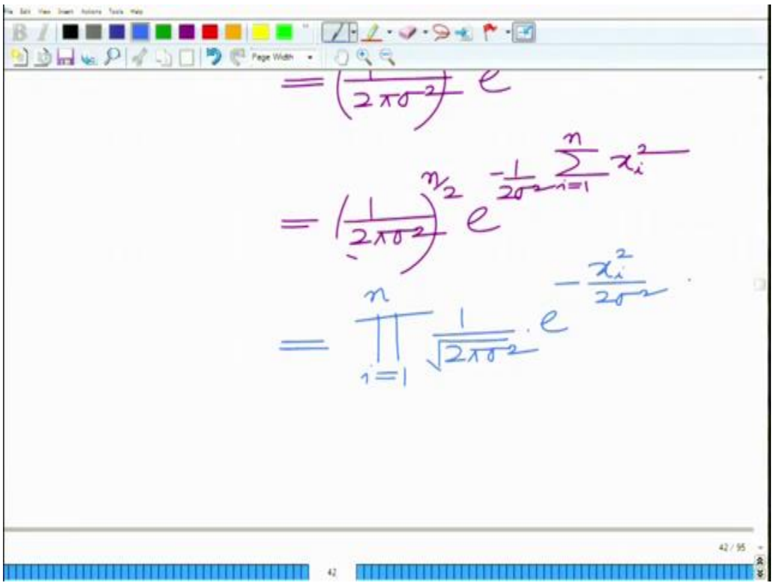
(Refer Slide Time: 35:19)



Which is because of the fact that

$$\bar{x}^T\bar{x} = x_1^{\ 2} + x_2^{\ 2} + \cdots + x_n^{\ 2}$$
$$= \|\bar{x}\|^2$$

(Refer Slide Time: 36:11)

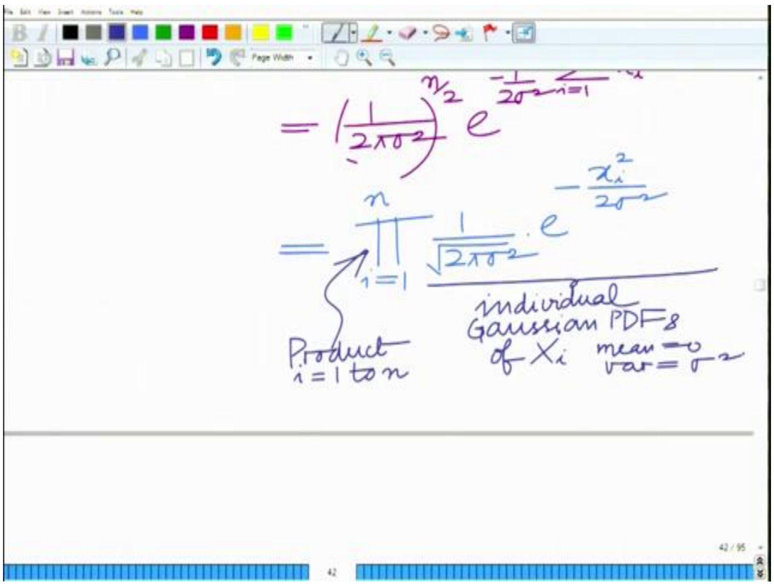


And further this is equal to

$$F_{\bar{X}}\left(\bar{x}\right)=\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2}$$

$$=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{x_i^2}{2\sigma^2}}$$

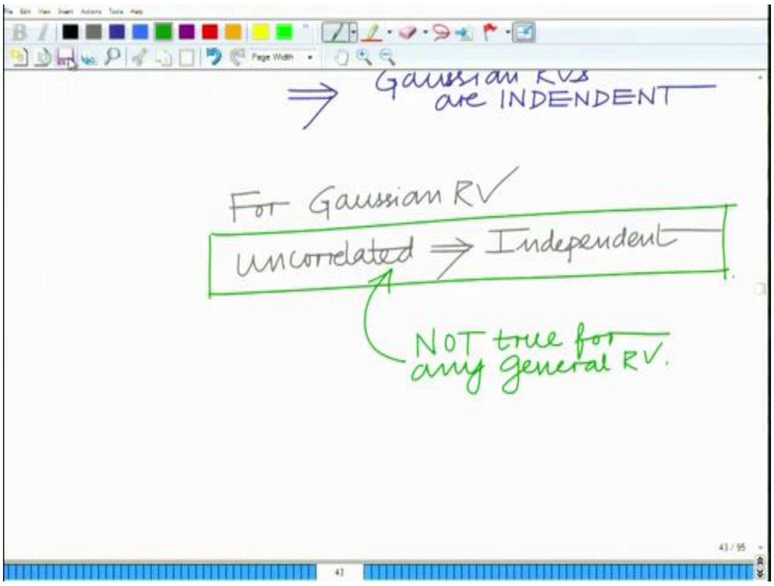
(Refer Slide Time: 37:06)



So, these are the individual Gaussian PDF's of the various random variables x_i with mean equal to 0 and variance equal to σ^2 and therefore, we can conclude that when this component Gaussian Random Variables are uncorrelated; the Multivariate Gaussian PDF equals the product of the individual PDF's.

This means that these random variables are uncorrelated as well as independent and this is a unique property of the Gaussian Random Variable. This is not true for any general random variable. It is an interesting property that is applicable only for the Gaussian Random Variables.

(Refer Slide Time: 38:15)



So, this implies that the uncorrelated Gaussian RVs are also independent. k. However, this is not true for any general random variable.

However it if in general it is only for a Gaussian Random Variable, it is true that if they are uncorrelated, they are also independent. This is not true for any general random variable all right. So, this small example illustrates this interesting property of the Multivariate Gaussian Random vector.

So, we will stop here and continue with other aspects in the subsequent modules.

Thank you.