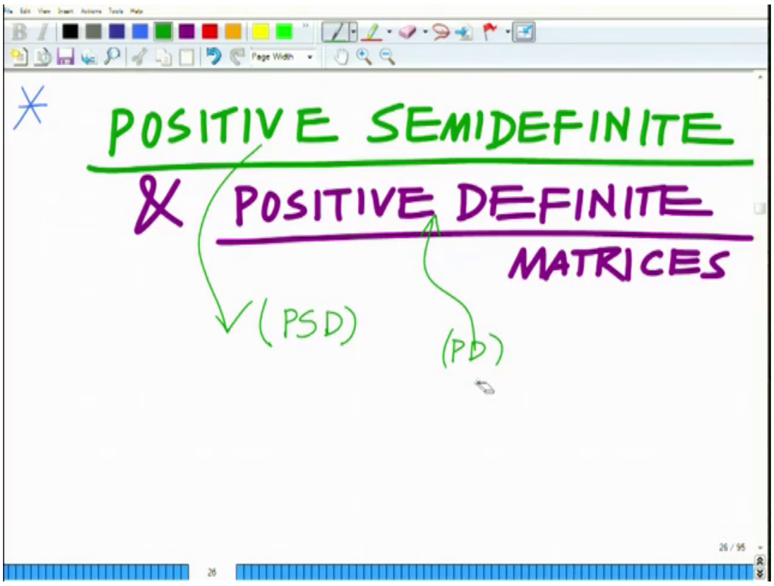
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

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

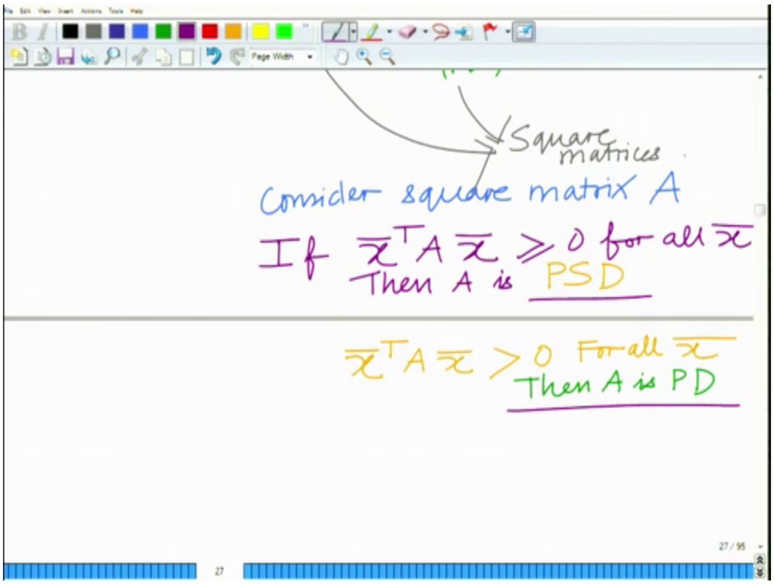
Hello, welcome to another module in this massive open online course. So, we are looking at the mathematical preliminaries for optimization all right. And we have looked at the Eigenvectors and Eigen values and this we will start looking at a different, different type of matrices known as positive semi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

(Refer Slide Time: 00:35)



So, we are going to look at the properties of and definition of positive semi definite and as well as positive definite, positive semi definite and positive definite matrices. And this is often abbreviate, abbreviated as PSD matrix and this is often abbreviated as a PD, positive definite matrix. A matrix can be positive semi definite matrices that is there can be positive semi definite matrices and positive definite matrices. And of course, both of these are also defined, once again only for square matrices all right. Similar to the concept of Eigen values and Eigenvectors, these are defined for square matrices ok.

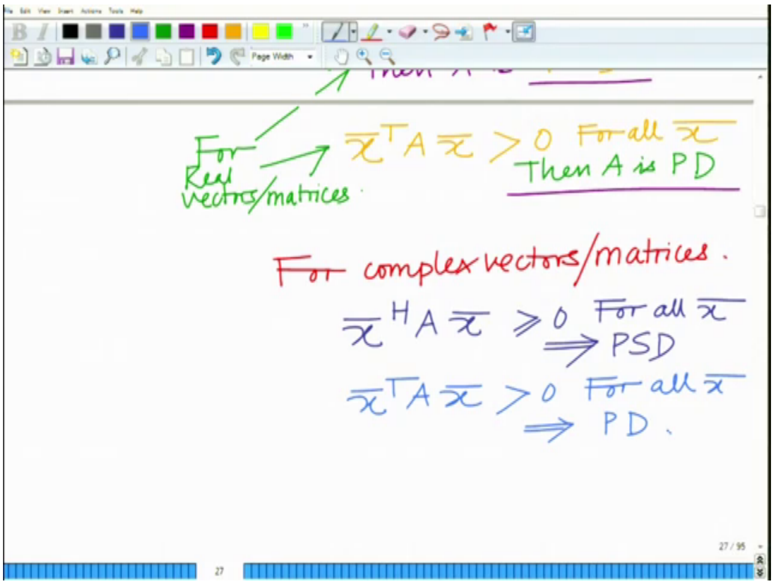
(Refer Slide Time: 01:51)



So, also defined only for square ok; now, if now, matrix A consider a matrix A; consider as square matrix A. Now, if for the real case, if x bar transpose A x bar greater than or equal to 0 for all x bar then A is a positive semi definite matrix ok. So, x bar transpose A x bar is greater than or equal to 0. Now, if x bar transpose A x bar is strictly greater than 0 for all x bar, then A is a positive definite matrix. So, this positive semi definite and positive definite matrix, if x bar transpose A x bar is greater than or equal to 0 right, for all x bar, all vectors x bar then it is positive semi definite, if it is strictly greater than 0 then it is positive definite.

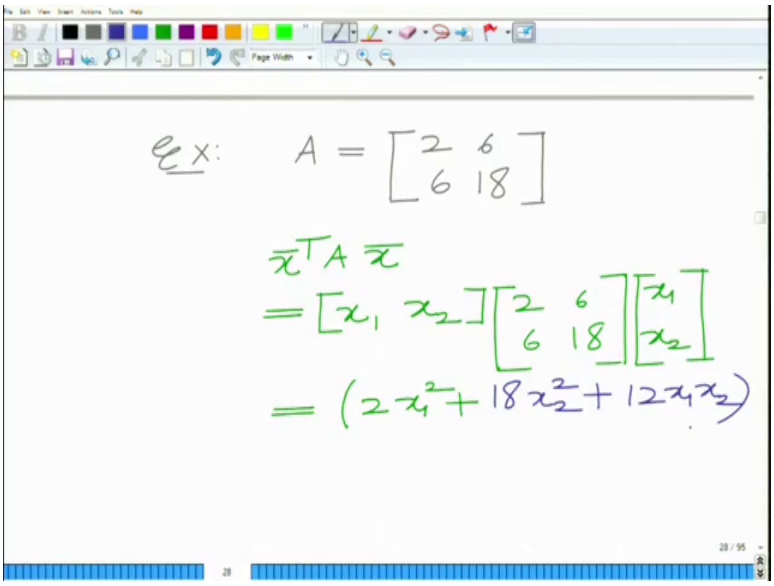
Now, these are for real vectors and real matrices, now, for complex vectors and matrices. Now, this definition is for.

(Refer Slide Time: 03:45)



Now for complex, for complex vectors or matrices as we have seen before; we have to replace the transpose by the Hermitian. So, x bar Hermitian A x bar greater than or equal to 0, for all x bar implies positive semi definite and further x bar Hermitian, x bar strictly greater than 0 for all x bar implies the positive definite matrix. What this definition is for complex matrices, complex matrices and vectors.

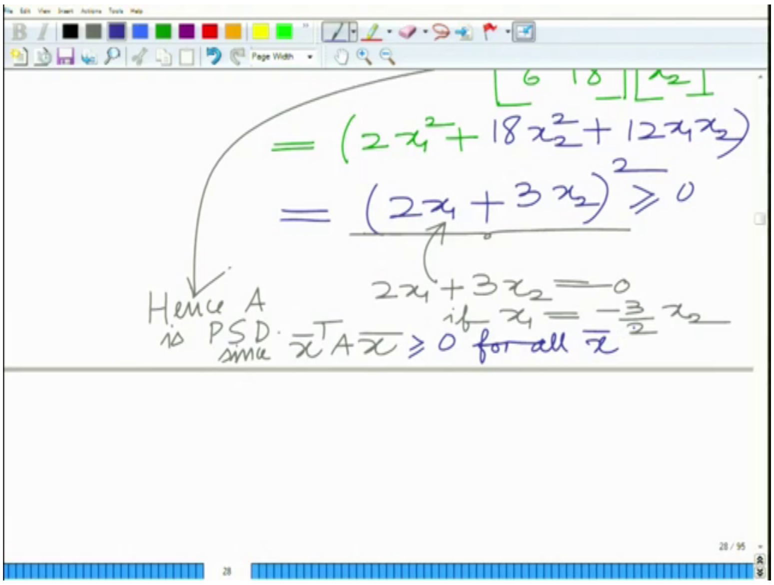
(Refer Slide Time: 04:59)



Let us take a simple example to understand this consider a square matrix consider, a square matrix A equals 2 6 comma 18. Let us consider the square matrix 2 cross 2 square matrix 2 6 6 18.

Now, let us look at x bar transpose, A x bar. This is a 2 cross 2 matrix. So, the vector x bar will be two dimensional x 1 x 2 times 2 6 6 18 into x 1 x 2 and this is equal to you can see this will be equal to well, this will be equal to twice x 1 square, when you multiply this out plus 18 x 2 square plus 12 times x 1 x 2. And this will be equal to you can easily check twice x 1 plus 3 times x 2 square, which is a perfect square and therefore, this is always greater than or equal to 0.

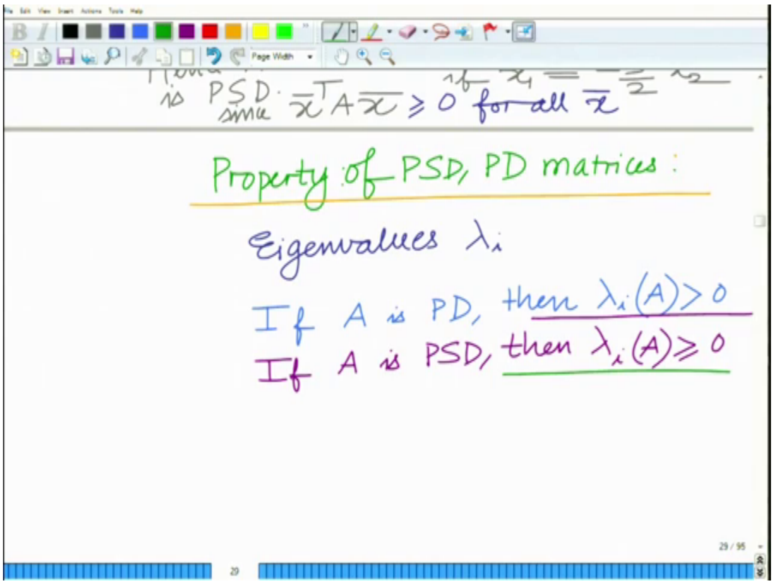
(Refer Slide Time: 06:26)



Hence, but and hence now, you can see this also, this not just this is not strictly greater than 0, because 2 x 1 plus 3 x 2 equal to 0, if x 1 equals minus 3 over 2 x 2 all right. So, this is only greater than or equal to 0, if 2 x 1 plus 3 equal to equal 0 then this will be 0 that is x bar transpose, x bar will be 0 otherwise, it is greater than.

So, therefore, in general it is greater than equal to 0. Hence, the matrix A is positive semi definite. Hence, this matrix hence, A is positive semi definite since x bar transpose A x bar is greater than or equal to 0 is greater than or equal to 0 for all x bar. Let us look at some, let us look at a property of this positive semi definite.

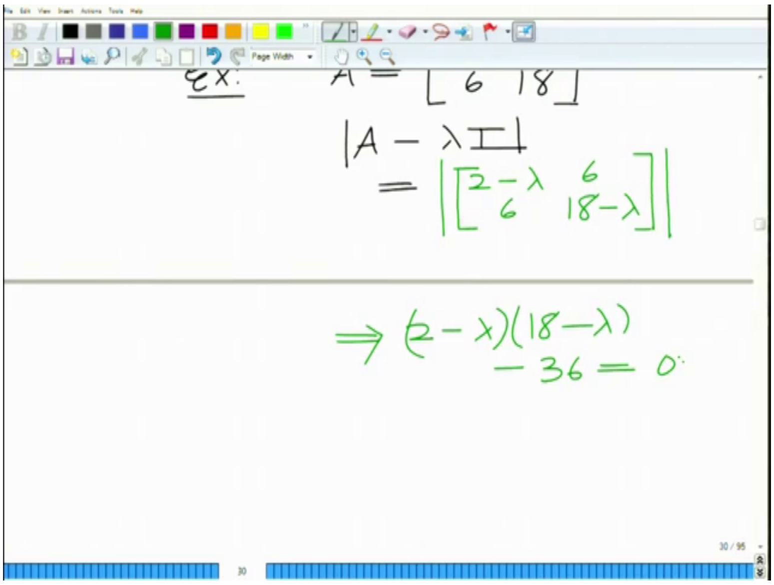
(Refer Slide Time: 07:58)



Let us look at a very interesting property, let us look at an interesting property of this. Now, consider the Eigen values since, these are square matrices Eigen values. So, the Eigen values if A is positive definite then the Eigen values lambda i of A denoting this by lambda i bracket A. These are strictly greater than all the Eigen values have to be greater than 0. On the other hand, if A is PSD then the Eigen values are greater than equal to 0 that is some of the Eigen values can be 0 and rest of them are greater than 0 ok.

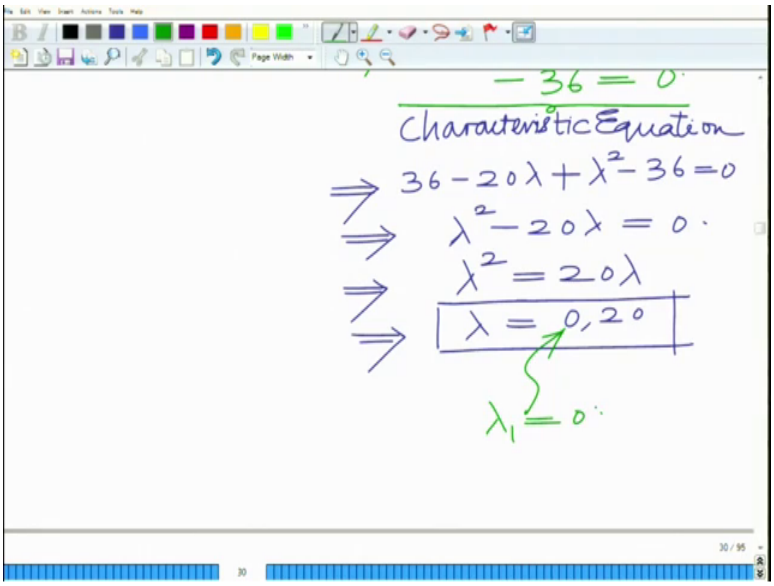
So, this is an interesting point for a PD matrix the Eigen values are strictly greater than 0 for a PSD matrix the Eigen values are greater than or equal to 0 again let us check verify this property on the previous example. So, let us look at A.

(Refer Slide Time: 09:36)



Again look at us look at the example A equals our matrix 2 6 6 18. Now, to calculate the Eigen values, one has to consider the characteristic polynomial determinant of a minus lambda i, which is the determinant of 2 minus lambda 6 6 18 minus lambda and this can be simplified. The determinant can be simplified as 2 minus lambda times 18 minus lambda minus 36 equal to 0 and this is the characteristic equation ok. Remember the characteristic polynomial and the characteristic equation for the matrix, this implies.

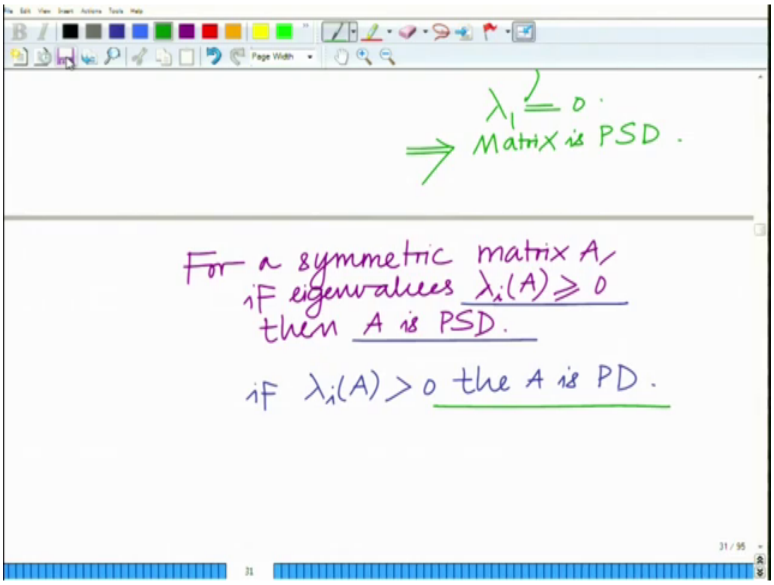
(Refer Slide Time: 10:55)



Now if you simplify this, this implies 36 minus 20 lambda plus lambda square minus 36 equal to 0. This implies lambda square minus 20 lambda equal to 0, which implies lambda square, which implies lambda square equals 20 lambda, which implies lambda equals 0 comma 20. These are the two Eigen values ok.

So, you can see. In fact, what you can see is very interestingly, that one of the Eigen values is in fact 0, correct. If you call this lambda 1, lambda 1 is in fact, 0 1 of the Eigen values is 0, this implies.

(Refer Slide Time: 12:13)

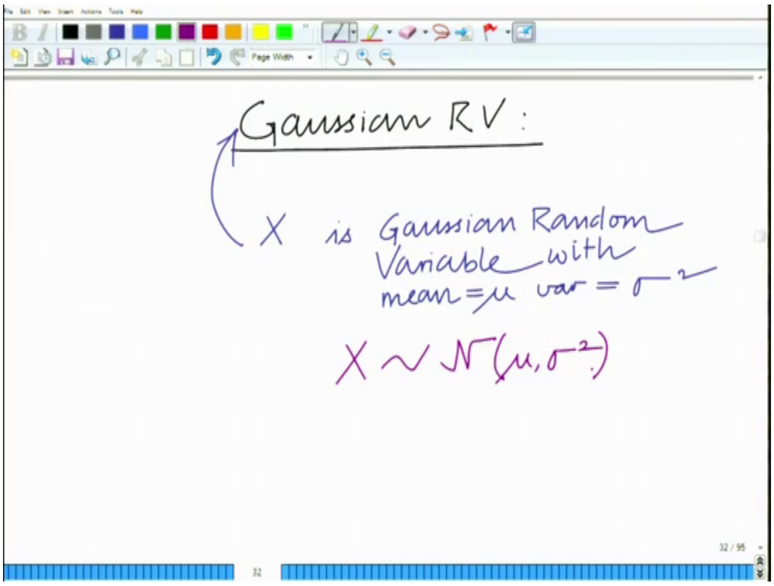


In fact, you can also see and we also checked that the matrix is PSD, positive semi definite. Now, for a symmetric matrix, if the Eigen values are greater than equal to 0 matrix, one can also conclude that, that is not only the forward property. But also the reverse that is right, that if for a symmetry that is for a symmetric matrix. If the Eigen values are greater than equal to 0, the matrix is positive semi definite, if the Eigen values are greater than 0 the matrix is positive definite for symmetric matrix A.

So, the interesting property for a symmetric matrix A, if Eigen values lambda i of a greater than equal to 0 then A is positive semi, if lambda i greater than equal to 0 then A is positive semi definite, if lambda i A is greater than 0 then A is a positive, then A is a positive definite matrix.

Let us now, continue looking at another important concept that is of the Gaussian random variable, which we are also going to use frequently in our framework of optimization.

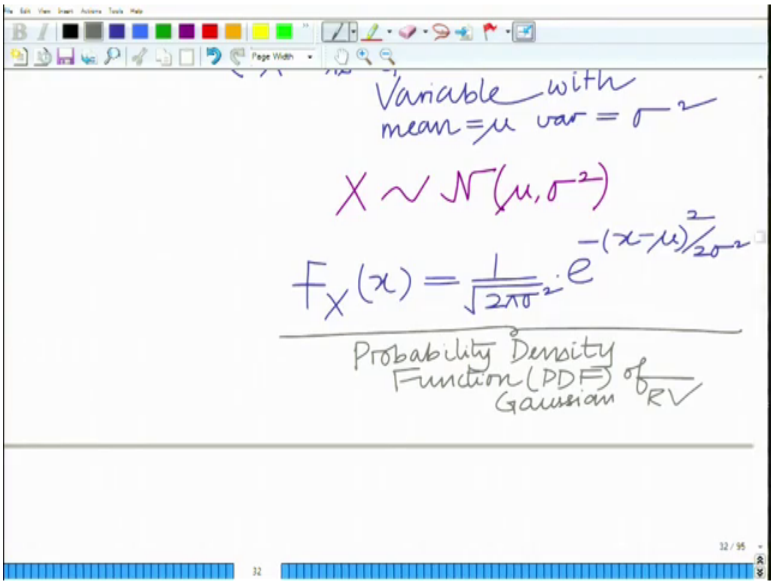
(Refer Slide Time: 14:02)



So, we want to look at the basic concepts of Gaussian Random Variables, which are central to our discussion on optimization. So, what is a Gaussian random variable? Well X is a Gaussian Random Variable, X is Gaussian Random Variable, variable with mean equal to mu and variance equal to sigma square that is, it is denoted by this notation Gaussian.

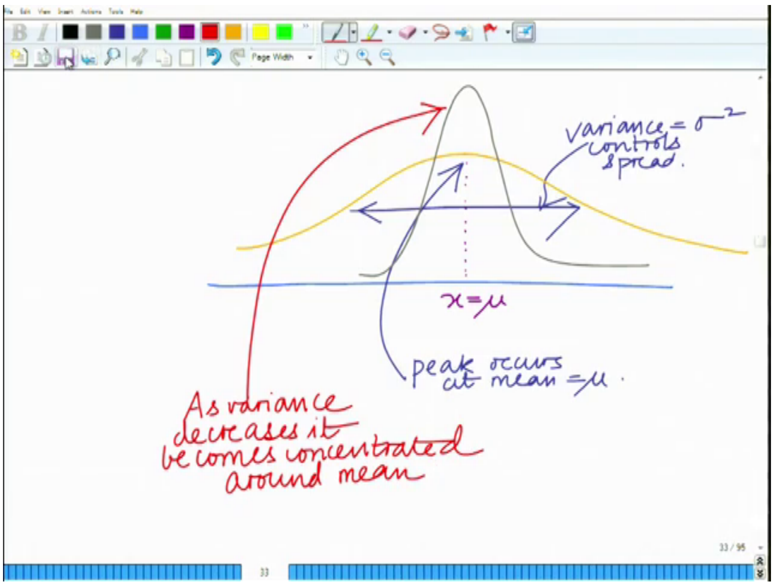
This is also known as a Normal Random Variable is denoted by this notation N mean script N mean mu variance sigma square ok.

(Refer Slide Time: 15:11)



And the probability density function of this every random variable has a probability density function that is given as 1 over square root of 2 pi sigma square for the Gaussian Random Variable e raised to minus x minus mu whole square by 2 sigma square. What is this? This is basically your PD F or the Probability Density Function. PD F of the Gaussian Random Variable with mean mu and variance equal to sigma square ok. So, this is the probability density function and you might also recall that.

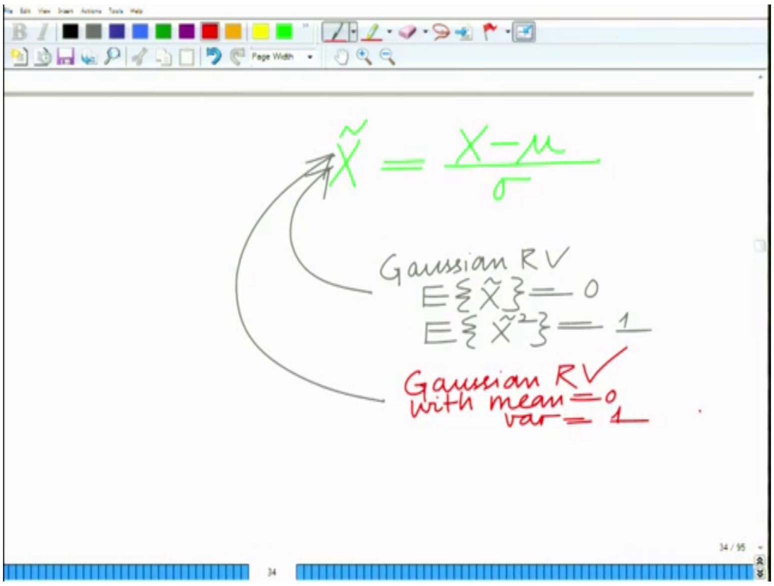
(Refer Slide Time: 16:11)



Many of you might be familiar that the shape of this probability density function is given by this bell shaped curve with the peak occurring, that is the peak of this occurs at x equal to Mu. That is the mean and the spread of this is controlled by the variance equal to sigma square controls the; so, the peak occurs at the mean.

So, the peak shifts in the Gaussian probability density function and it is symmetric about the mean. It is a bell shaped curve and the spread of this is controlled by the variance. For instance, of the variance decreases then the spread decreases the Gaussian probability density function becomes more and more, peak. Here, that is more and more concentrated around the mean ok. So, as the variance decreases, it becomes more and more concentrated around the mean.

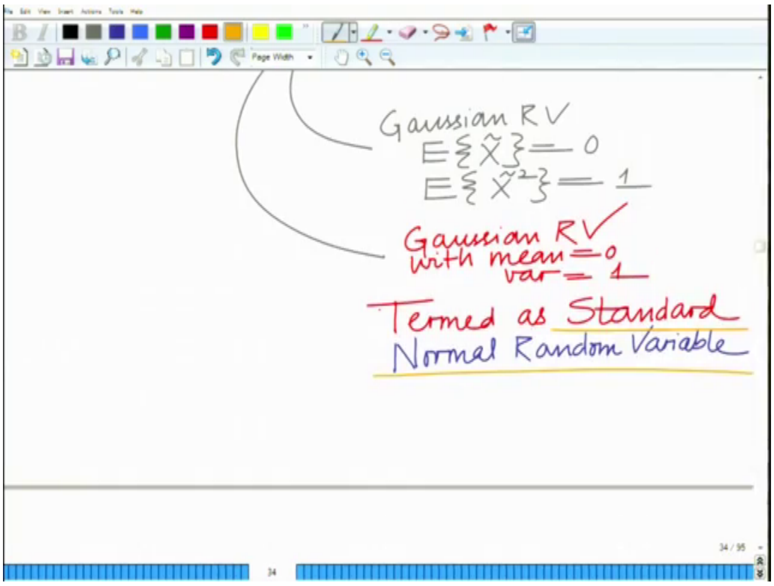
(Refer Slide Time: 18:13)



As the variance decreases, it becomes concentrated around the mean and now, one can define a new random variable X tilde by subtracting the mean and dividing by the square root of the variance or sigma, which is the standard deviation ok. So, now, you can see this X tilde is also which is derived by subtracting the mean and dividing by the standard deviation sigma. Now, this X tilde is also a Gaussian R V and there is something interesting about X tilde, the mean of X tilde will be 0 that is expected value of is 0.

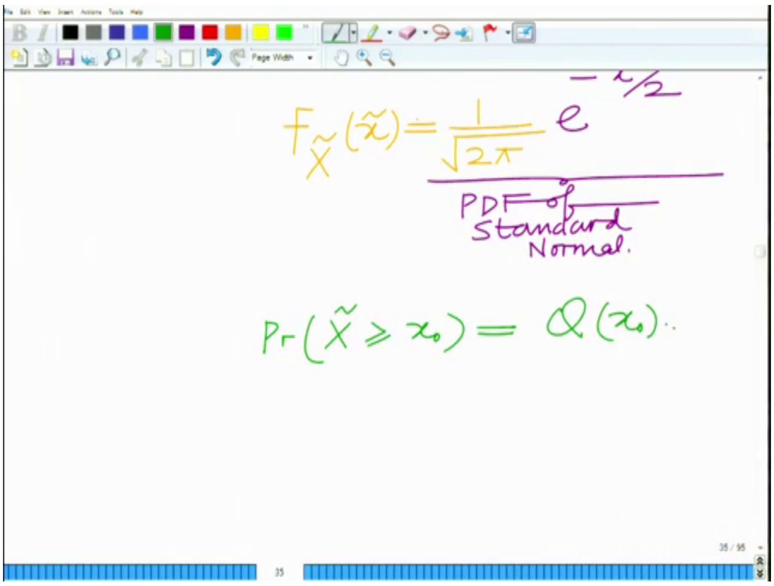
And the variance that is expected value of X tilde minus mu or X tilde square in this case as mean is 0 is equal to 1. So, mean equal to 0 and X tilde is Gaussian R V with mean 0 and variance unity ok. So, this is a Gaussian R V with mean equal to 0 and variance equal to 1, such a Gaussian R V with mean equal to 0 variance equal to 1.

(Refer Slide Time: 19:22)



This is termed as the standard normal random variable, standard normal the system, as the standard normal random variable ok. And now, we define the standard normal random variable is used to define what is known as the Q function.

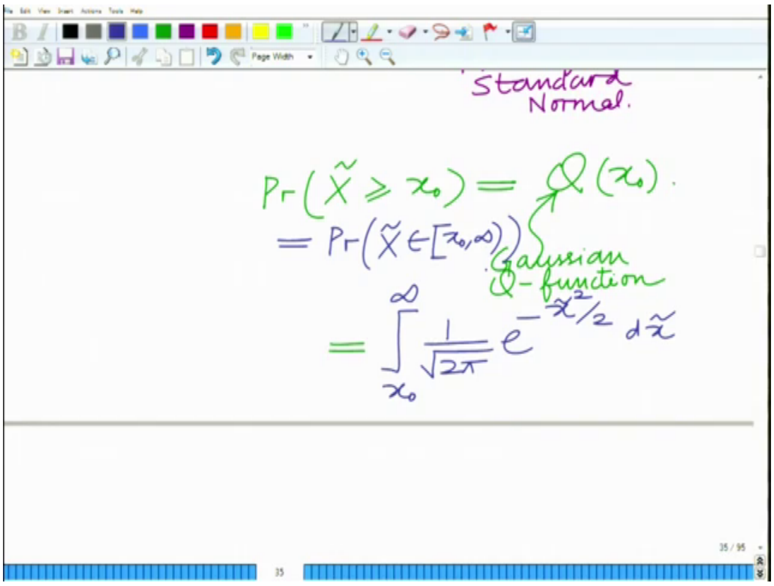
(Refer Slide Time: 20:04)



So, this now the probability density function of the standard normal random variable that is also simple to design, because it has mean mu equal to 0 and variance 1. So, you substitute mu equal to 0 and sigma equal to 1, in the earlier expression for the probability density function of the Gaussian Random Variable and what you get is the PD F of this Standard Gaussian Random Variable given as 1 over square root of 2 pi.

So, sigma square equal to 1 1 over square root of 2 pi e raised to minus X tilde square, remember mu is 0. So, X tilde minus mu is simply X tilde, X tilde squared divided by 2 again. Once again sigma square equal to 1, this is the PD F of the Standard Normal and one can define the Q function of the PD F of the Standard Normal, that is the probability that this Standard Normal Gaussian Random Variable X tilde is greater than or equal to a quantity x naught is denoted by Q of x naught. This is also termed as a Q function or the Gaussian Q function.

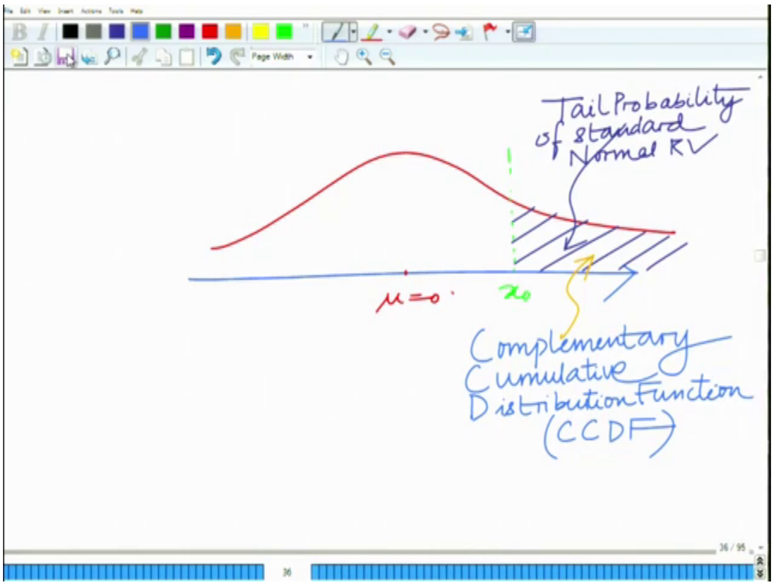
(Refer Slide Time: 21:16)



This is the probability that the standard normal variable X tilde is greater than or equal to x naught, which is given by the integral X tilde x x naught to infinity. The integral of the probability density function of the Standard Normal Variable 1 or square root of 2 pi e raised to minus x tilde square divided by 2 times d of x tilde that is the probability that this, this also equal to the probability that X tilde belongs to the interval x naught comma infinity.

That is basically, what we are asking is the question, what is the probability that X tilde the Standard Normal Random variable takes a value greater than or equal to x naught or basically, that it lies in the interval x naught to infinity and that is remember the probability that the random variable lies in a particular interval is given by the integral of the probability density function of the random variable over that interval. So therefore, it is given by integral x naught. This probability is given by the integral x naught to infinity 1 or square root of 2 pi e raised to minus X tilde square divided by 2 d x tilde and what it denotes as?

(Refer Slide Time: 22:55)

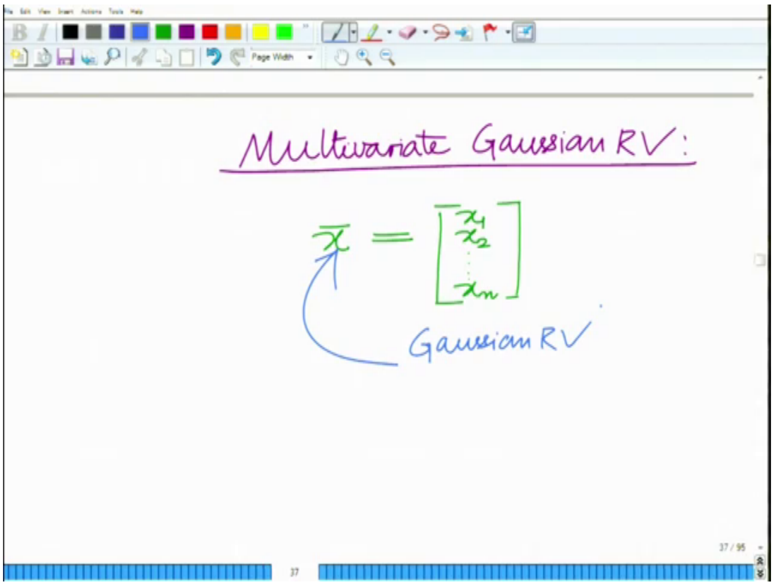


I have already told you it denotes the probability that this Gaussian Random variable with mean 0 and variance equal to 1 is greater than this quantity x naught. So, it denotes the probability that or the area under the PD F greater than or equal to x naught. This is also known as the tail probability of the Standard Normal Random Variable is also known as the tail probability of the Standard Normal.

There is probability under the tail starting from x naught, this is also known as the Complementary Cumulative Density Function. The Cumulative Density Function gives the probability that the random variable takes values less than x naught the complement of that or 1 minus the C D F gives the probability, that it is greater than or equal to x naught. This is therefore, known as the Complementary Cumulative, also known as the Complementary Cumulative Distribution Function or the C C D F of the Standard Normal Random Variable.

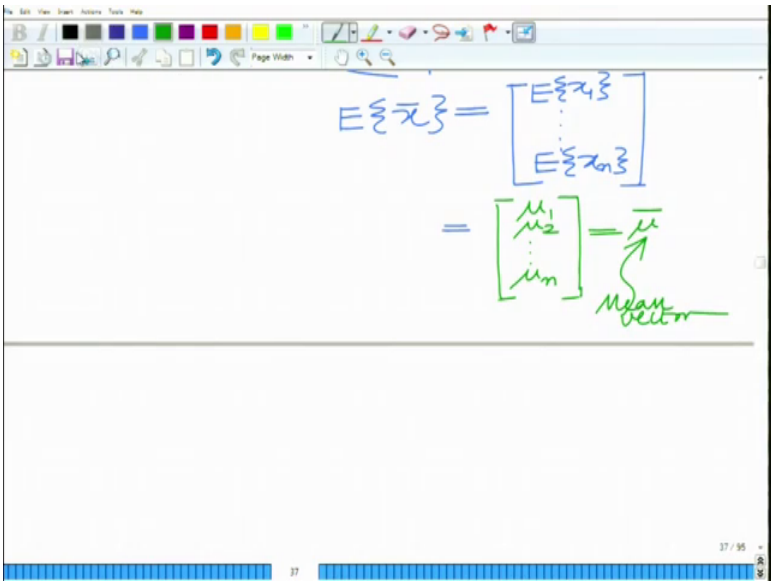
Now, let us come to a multi normal Multivariate Gaussian Random Variable or a Gaussian Random Vector. A Gaussian Random Vector with multiple components, each of them individually Gaussian and all of them being jointly Gaussian ok. So, we will talk about a Multivariate, Multivariate Gaussian R V.

(Refer Slide Time: 25:01)



Now, a Multivariate Gaussian R V is given by x bar equals x 1 x 2 up to xn and this is a Gaussian random vector x bar. This is a Gaussian random vector x bar and x bar equals x 1 x 2 x n. So, this has n components we will denote the mean now each of the components is going to have a mean.

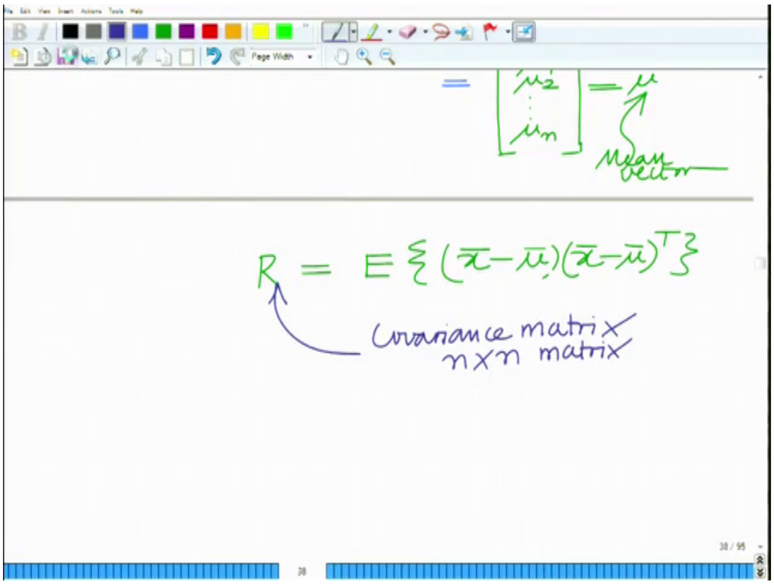
(Refer Slide Time: 25:54)



So, the mean is going to have a b a vector that is expected x bar equals the expected value of each of the components. So, this is going to be an n dimensional vector, which we will denote by this. Let the mean of the various components be mu 1 mu 2 up to mu n. So, this is basically your mean vector. So, the mean is going to be a vector.

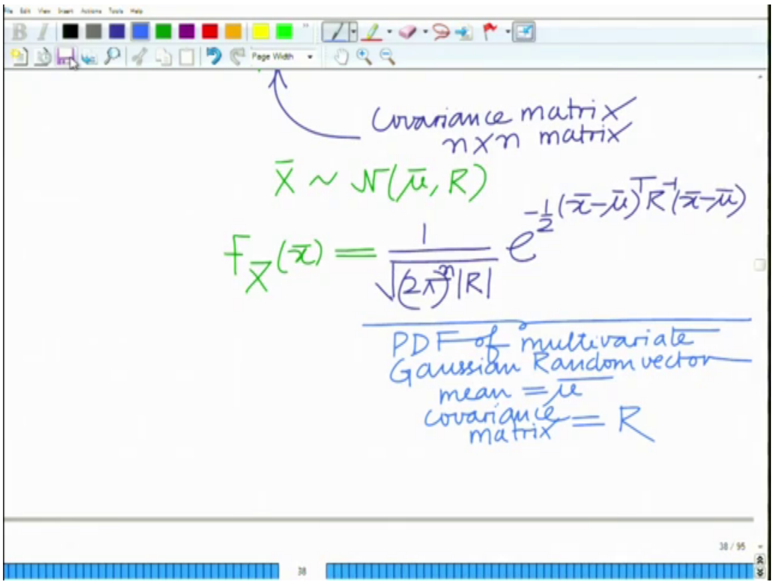
So, you can call this also as the mean vector of the Multivariate Gaussian random variable and further instead of the variance. We will have the covariance matrix, which looks at the variance of each component and also the cross. There also the covariance correct, the cross correlation corresponding to each of the elements of this random vector and the covariance matrix is defined as follows.

(Refer Slide Time: 27:08)



That is R equals expected value of x bar minus mu bar into x bar minus M u bar transpose and this is remember, this is termed as the covariance matrix. This is termed as a covariance matrix and this is an n cross n matrix, the system there is a covariance matrix and this is an n cross n matrix.

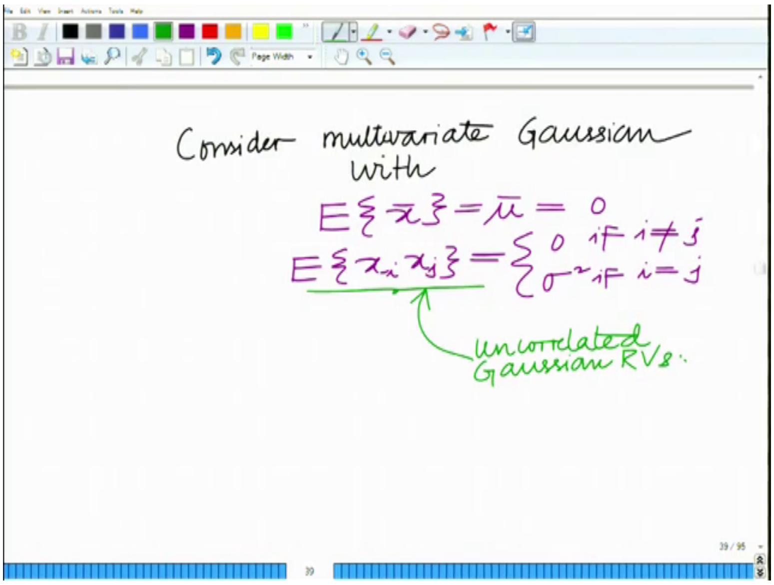
(Refer Slide Time: 27:53)



The probability density function of the Multivariate Gaussian Random Variable into and you can also denote this as n again. So, the Multivariate Gaussian Random Variable you can denote it as Gaussian Random Variable with mean vector mu bar, and covariance matrix R, and the probability density function is given as 1 over square root 2 pi raised to the power n raised to the power of n the determinant of the covariance matrix R times e raised to minus half x bar minus mu bar transpose R inverse x bar minus mu bar. And this is basically your PD of the Multivariate Gaussian Random Vector PD F of Multivariate Gaussian Vector with mean equals mu bar and variance equal to a covariance sorry, covariance matrix and covariance matrix equals R, all right and that is the expression for the PD F of the Multivariate Gaussian Random Vector.

Let us now, look at an example to understand this and let us look at an interesting special case of this Ultivariate Gaussian Random Vector, which is when the different components of this Gaussian Random Vector are independent.

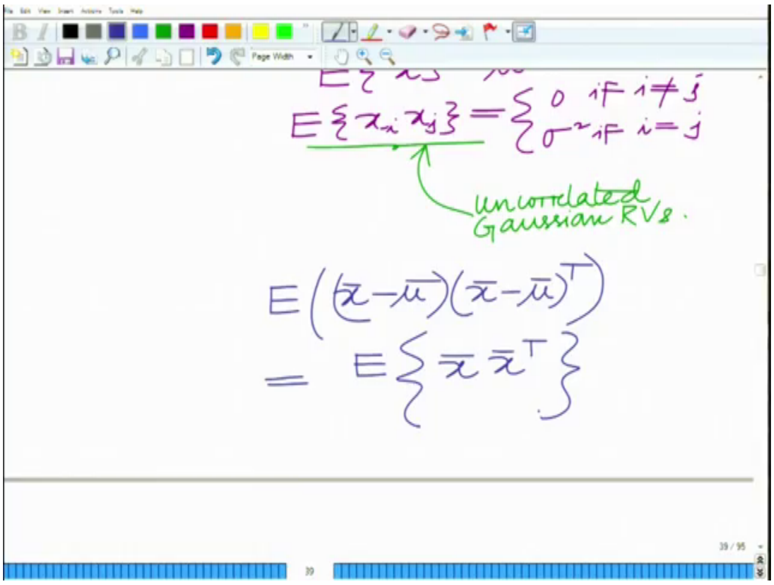
(Refer Slide Time: 30:04)



So, consider, consider a Multivariate Gaussian with expected x bar equals mu bar equals the 0 vector and expected value of x i into x j. The different components is equal to 0 if i not equal to j and sigma square if i equal to j alright. So, the, what we are seeing is the cross correlation expected value of x i x j, if i naught equal to j is 0 all right and all the variances of each of variance of each element is sigma square. So, these are basically what you can see is; basically these are known as uncorrelated random variables, because the correlation is 0.

So, these are basically uncorrelated, uncorrelated Gaussian Random Variables is the cross correlation, cross covariance is 0. These are uncorrelated Gaussian Random Variables ok.

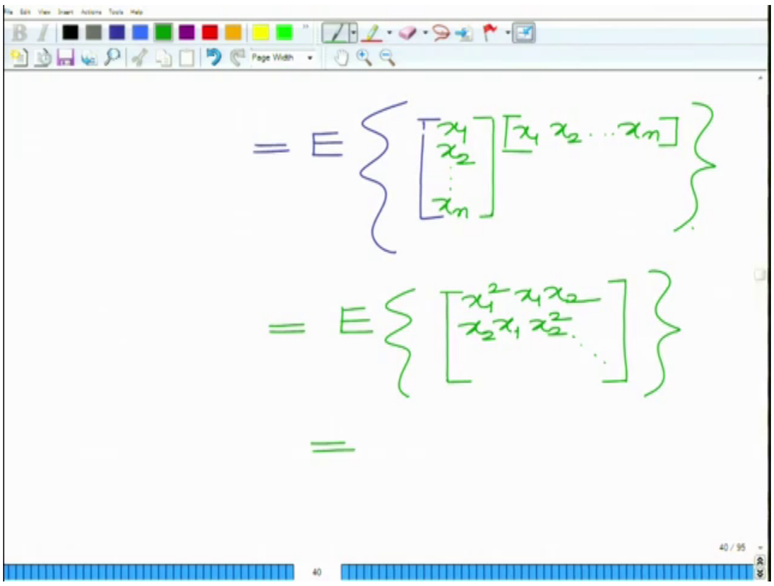
(Refer Slide Time: 31:50)



And now, if you compute the covariance of this, that will be given as of the covariance matrix of this, that will be given as well. We already seen that is x bar minus mu bar into x bar minus mu bar transpose.

Now, mu bar is 0. So, this will simply be expected value of x bar into x bar transpose which is now, let me write it in terms of it is vector.

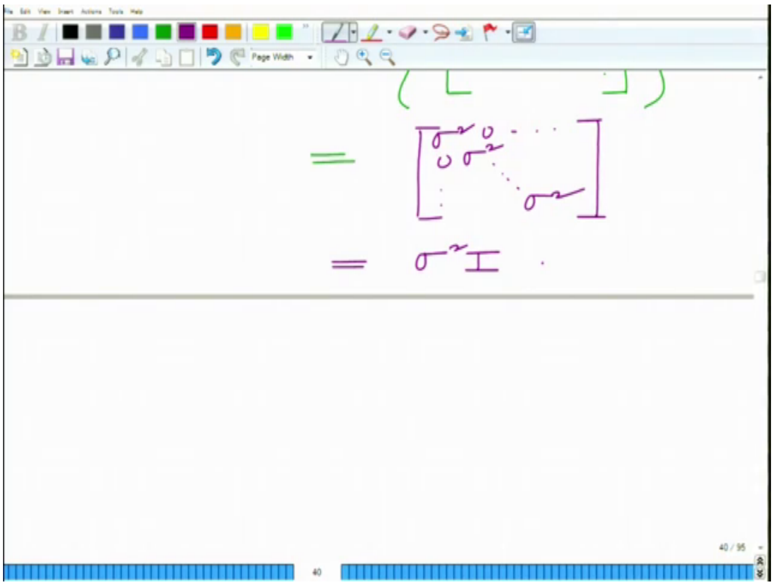
(Refer Slide Time: 32:17)



This is going to be expected value of x bar which is the vector x 1 x 2 x n into the transpose x 1, which is the row vector remember transpose of a column vector is a row vector and now, once you compute this what you will observe is if you compute this you will have entry such as x 1 square, the off diagonal entries will be x 1 x 2 x 2 x 1 x 2 square and so on.

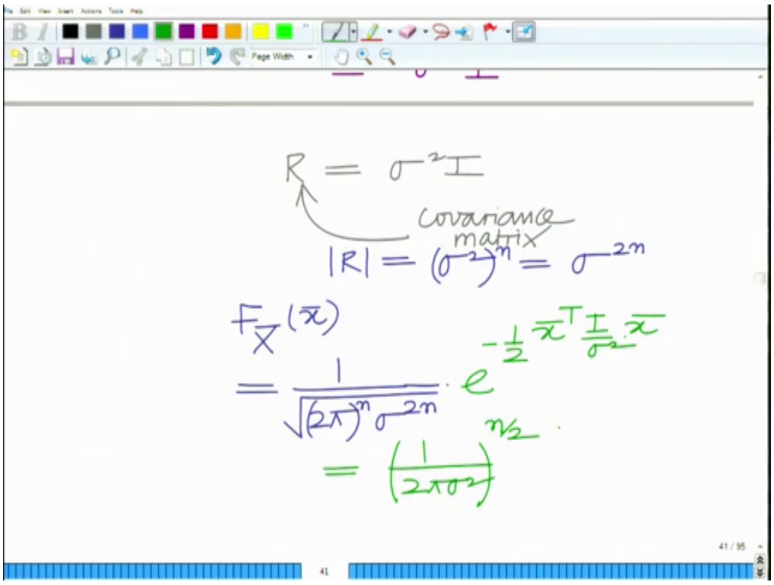
And now, if you look at this matrix expected value of each element on the diagonal expected value of x 1 square x 2 square x 3 square so on. Sigma square, when the expected values of the off diagonal entries x 1 x 2 x 2 x 1 and so on are 0, because these we have considered the random variables to be uncorrelated. And therefore, the covariance matrix basically, you can see reduces to sigma square.

(Refer Slide Time: 33:29)



All the diagonal elements are sigma square the off diagonal elements are 0. And therefore, this is also sigma square times identity.

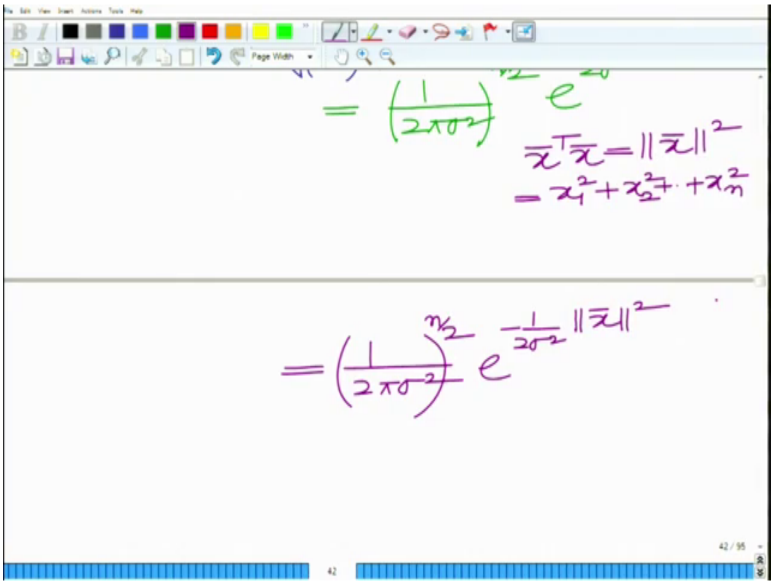
(Refer Slide Time: 33:50)



And now, once you compute; so, our covariance matrix equals sigma square times identity, this is remember, this is our covariance matrix. And therefore, the multivariate Gaussian Probability Density Function is given as 1 over square root. Now, if you look at the determinant of this determinant of R is sigma square times identity determinant of R is sigma square raise to the power of n, which is sigma to the power of 2 n and therefore, the probability density function is 1 over.

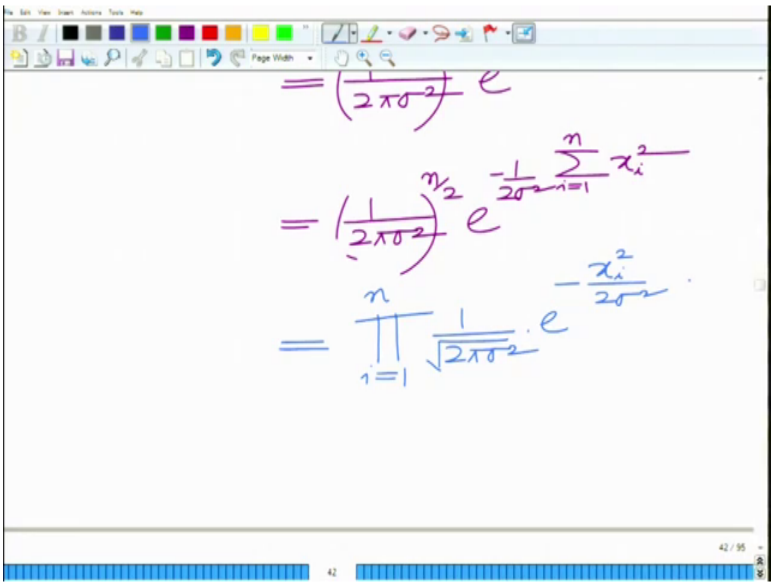
Let me just write it separately is 1 over square root of 2 pi raised to the power of n sigma raised to the power of 2 n that is the determinant times e raised to minus half x bar minus mu bar that is x bar transpose R inverse R sigma square identity R inverse is identity divided by sigma square times x bar.

(Refer Slide Time: 35:19)



Which is now, you can simplify this as 1 over 2 pi sigma square raise to the power of n by 2 times e raised to minus half or 1 over 2 sigma square x bar transpose x bar transpose identity x bar is x bar transpose x bar, but recall x bar transpose x bar recall x bar transpose x bar equals norm of x bar square, which is also equal to x 1 square plus x 2 square plus so on, up to x n square for a real vector x bar. And therefore, this is also equal to 1 over 2 pi sigma square to the power n by 2 times e raised to minus 1 over 2 sigma square norm x bar square.

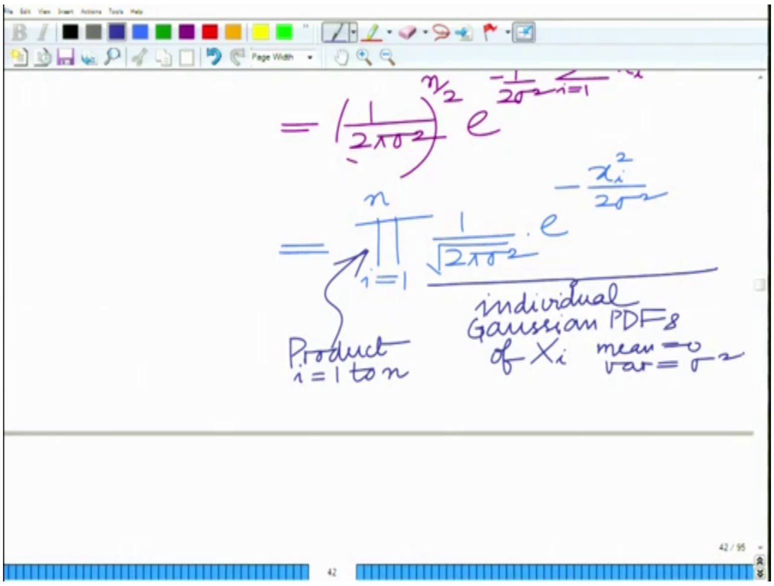
(Refer Slide Time: 36:11)



And this is also equal to 1 over 2 pi sigma square n by 2 e raise to minus 1 over 2 sigma square summation 1 equal to i to n x i square and interestingly.

Now, you can also write this as the product i equal to 1 to n 2 pi sigma square or 1 over you can also write it as 1 over square root of 2 pi sigma square e raised to minus x i square divided by 2 sigma square. And therefore, now, this is the product symbol similar to summation.

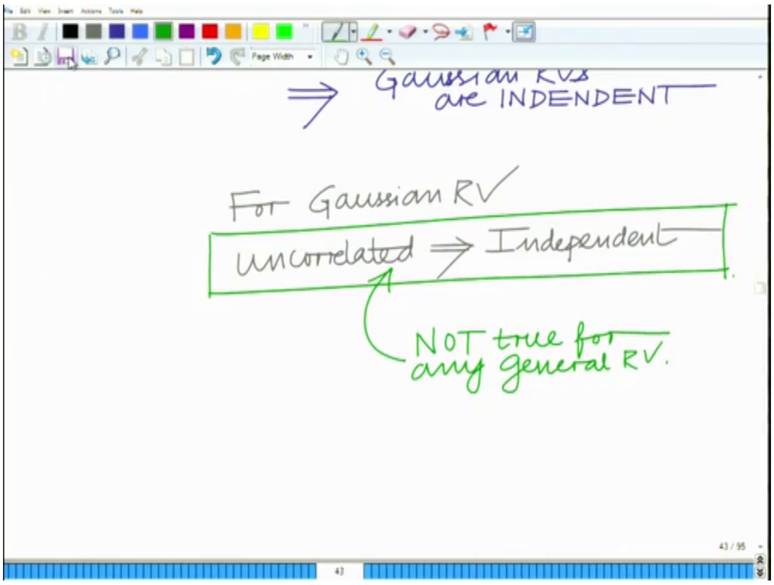
(Refer Slide Time: 37:06)



This is your product symbol, product i equal to 1 to n. And now, you see these are the individual PD F's. These are the individual Gaussian PD F's of the various random variables X i with mean equal to 0 and variance equal to sigma square and therefore, what you are seeing is that when these Gaussian, when this component Gaussian Random Variables are uncorrelated the joint Gaussian the Multivariate Gaussian PD F equals the product of the individual PD F's.

So, which means that these PD F's, these random variables are not only uncorrelated, but they are also independent and this is a unique property of the Gaussian Random Variable, if two Gaussian Random variables are uncorrelated, they are also independent. 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 this implies that the Gaussian R V's are also independent ok. So, for a Gaussian R V uncorrelated implies that they are independent. However, this is not true for any general random variable and this is the important property, not true for, but the other way round is always true right, for any random variable if they are independent. If two random variables are independent then they are going to be uncorrelated.

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, Multivariate Gaussian Random vector alright.

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

Thank you.