Optimal Control Guidance and Estimation

Prof. Radhakant Padhi

Department of Aerospace Engineering

Indian Institute of Science, Bangalore

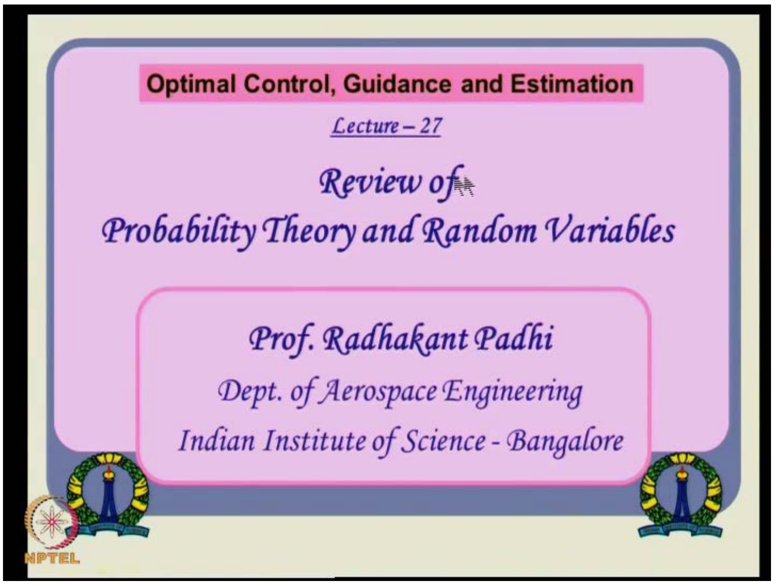
Module No. # 12

Lecture No. # 27

Review of Probability Theory and Random Variables

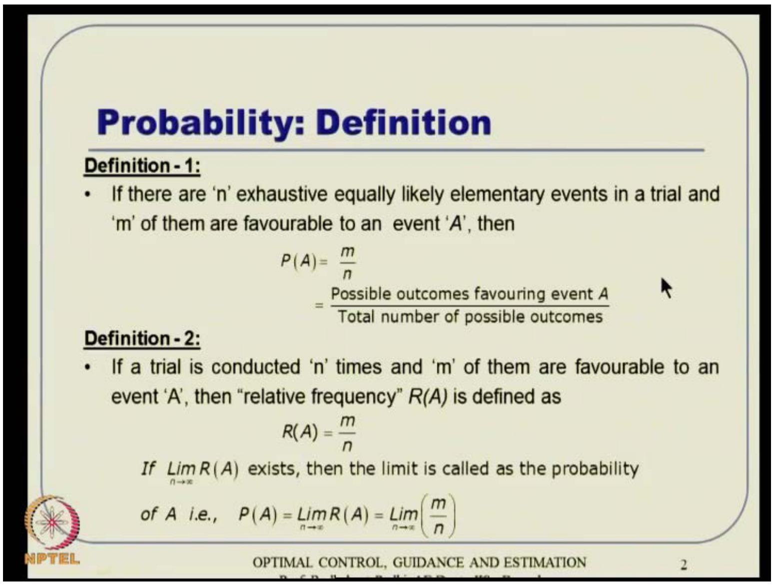
Hello, everybody, let us continue with our lecture series on this optimal control guidance and estimation, we have just started some estimation concepts last time, we actually studied this l Q observer followed by some overview of kalman filter implementation and things like that.

(Refer Slide Time: 00:19)



So, next couple of lectures, we will go through this derivation process of a kalman filter to understand it much better, and then you talk about some issues of implementation and things like that actually. Alright, so in this particular lecture, we will primarily concentrate on review of probability theory and random variable which is actually needed for derivation of kalman filter later.

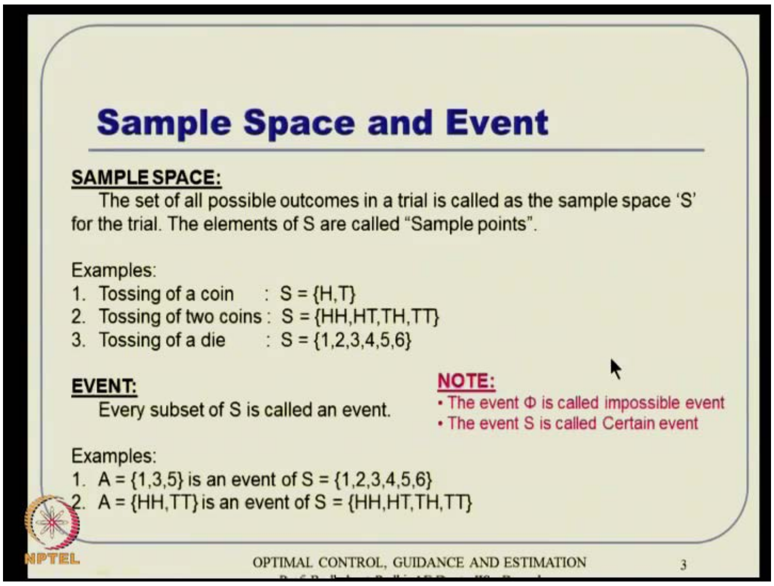
(Refer Slide Time: 01:00)



Allright, these are some of the very basic definitions of probability, there are a two different ways of looking at it, so first definition is something like this, if there are n exhaustive equally likely elementary events in a trial, and m of them are favorable to an event a, then probability of a is defined as n over m. Alright, so we have n exhaustive elementary events and m of them are favorable to event a, that is how with it goes, then probability of a is nothing but a m by n which means, possible outcomes favoring event a divided by total number of possible outcomes, it is very interactive actually.

So, definition two tells something very similar, but in a little bit different way, but else if a trial is conducted n times and m of them are favorable to an event a, then there something called relative frequency R A which is defined as m over n, it does not define A probability directly, it tells because sample space is in not large and all that actually, so we tell relative frequently turns out to be m by n, and in the limit where m tends to infinity, and if the limit exists then the limit is defined as probability actually. So, essentially, it tells you that the number of samples - I mean - should be large then only it can define something like probability basically.

(Refer Slide Time: 02:40)



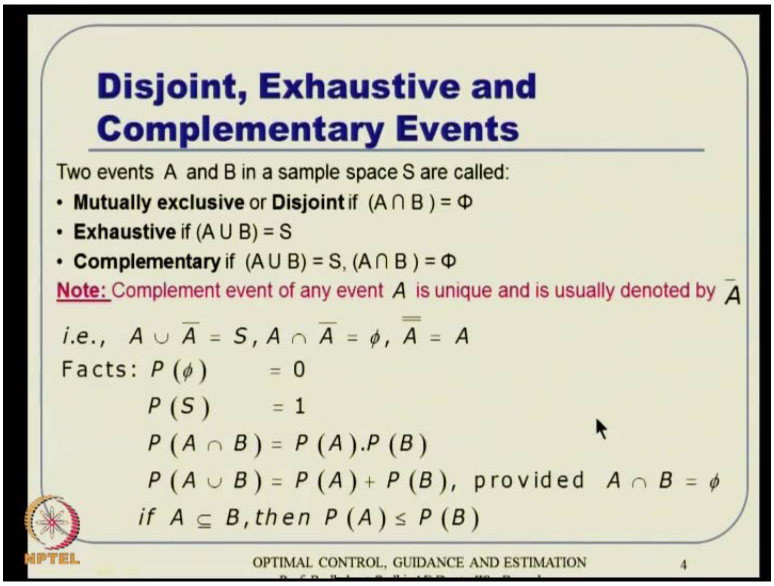
What is sample space and events and think like that, if you think about sample space this is defined as something like the set of all possible outcomes in a trial is called the sample space s, for it further trial and the elements are also called sample points.

So, examples of tossing a coin is something like this, you have this something like if you toss a coin one time - I mean - this is either head or tail, so you can think of a head be the sample point and tail is a sample point, whereas sample space it contains either all heads or all tails, so h and t - I mean - if you take that way that is the sample space. Simultaneously, if you toss two coins then obviously, there are four possibilities, so either you can get h h - head head, head and tail or tail and head or tail and tail, sort of thing something like that, so each of these combinations are nothing but sample points whereas a total set s is nothing but the sample space actually.

Similarly, if you toss a die, die has 6 sides, so we have each of these 1, 2, 3, 4, 5, 6 are sample points where as the total collection of all the sample points if you define that as a set that is nothing but the sample space. So, the event is defined as something like every subset of s is something we called an event. So, if you take any subset of s either single element, two elements thing like that, even the full set is nothing but as event actually. And there something like two special events, one is phi which is called impossible event - null set actually, and then the total set - I mean - this is entire everything is there in that, so the event s is called the certain event if you take the total thing it is guaranteed to have everything actually.

So, for example, if you take 1, 3, 5 out of this set, and then this is an event and, suppose, this we just collect h h and tail tail out of that and that is an event.

(Refer Slide Time: 04:43)



Wherever other concepts something like disjoint set - I mean - disjoint event, exhaustive event, complementary event, and things like that, so this is very similar or very close to set theory actually. If you know the that way, so two events A and B in a sample space s are called mutually exhaustive or disjoint events, if there is nothing in common between them; that means, a intersection B is a null set.

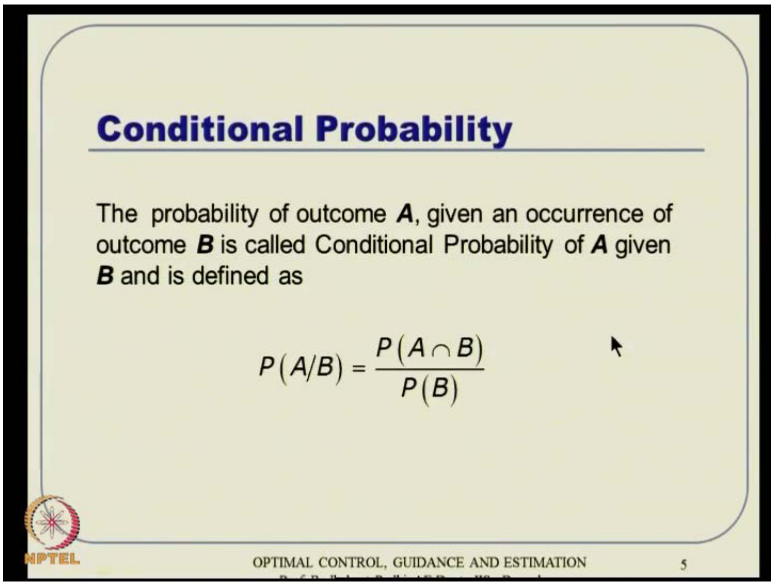
And if there exhaustive - I mean - if these 2 A and B are exhaustive, A union of turns out to be the sample space, everything is contained in that if you take union of that. And think like complementary, is something very interesting if you have A union B is s, but a intersection b is phi; that means, the complement each of them very well and together there kind of contains all sample space, and there is nothing common in between them actually, so that is a kind of thing we are looking for here.

And there is something called complement event also and that is defined as for any event a there is a complement event a bar such that a union a bar is s actually, and also a intersection a bar is phi, and obviously, if you take compliment of compliment that A double bar, obviously, it terms out to be basically.

And some facts are something like this, you have this probability of phi is obviously, nothing but this is 0, probability of the sample space is 1, and probability of A intersection B is defined as it turns out to be probability of A into probability of B, and probability of A union B is summation of them, provided there is nothing common between them. And if it turns out to be a subset of B then probability of A is going to be less than equal to probability of b actually.

They are some of the results that sometimes commonly in probability theory basically, remember, all of that you may not need while deriving kalman filter, but just sort kind of precautional is will also help us and understand it slightly better.

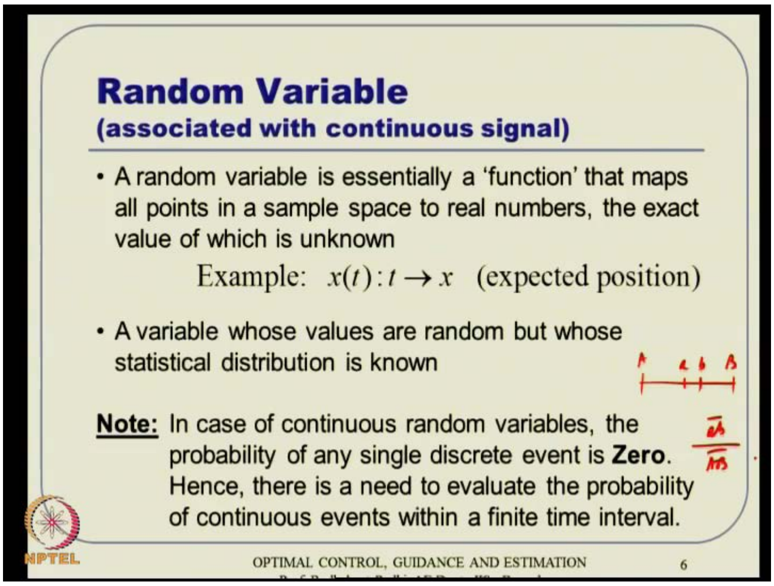
(Refer Slide Time: 06:58)



Is something called conditional probability and then conditional probability is defined something like this the probability of outcome given an occurrence of outcome B is something called as conditional probability of given b, and this is defined as a probability of A given B is probability of A intersection B divided by probability of B basically. There are examples, and then if you really if you are interested, we can see a couple of probability theory books or even nice mathematics books like classic and all that, so you can get a lot of ideas an including examples and all that.

And using somewhat there concepts just for your information there are nice advanced filtering theory - filtering techniques - have been proposed recently, so these are not very useless that way; in other words, even through these are not directly useful in kalman filter for say, if you know this then you can understand about filter techniques also better basically then.

(Refer Slide Time: 08:02)



These are something like discrete events and all that, so what we are really interested in systems theory is, typically, something that we associated with continuous signal and know the system - I mean - when we talk about x dot equal to some f of x and think like that then that is typically x of t is a continues variable and thing like that, so those things are randomly varying actually.

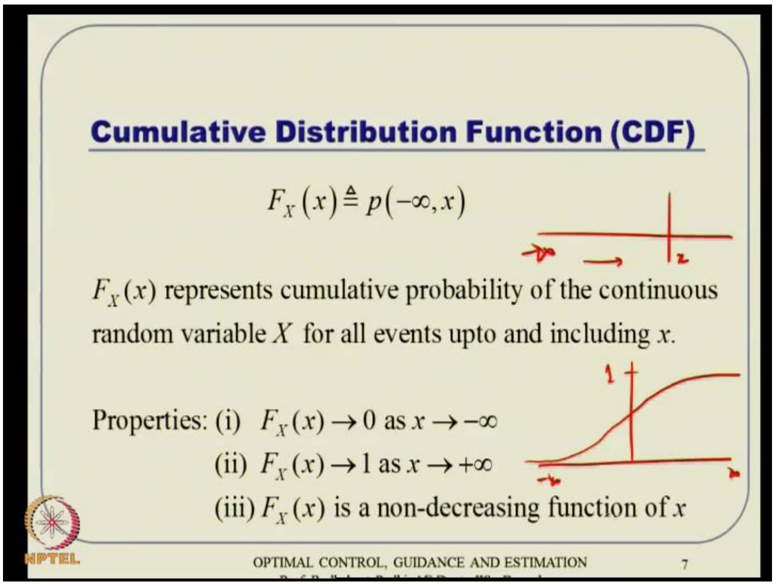
So, in that sense, you can define something called random variable; a random variable is essentially a function, remember this is actually a function that maps all points in a sample space to real number, but the exact value of the real number is not known actually. So, for example, if you think about x of t in case of given at t there is a value, but the exact value - the numerical value - is uncertain, this is not known really so that is called a random variable.

In other words, a variable whole values are random, but whose statistical distribution is known basically, also remember that is - I mean - any random variable cannot be purely random in that sense, we have to, we will talk about some sort of a statistical distribution necessary with that, we will talk about in a while, so this is called probability density function and think like that, so those things are known but the exact value that comes out in each experiment is typically are known.

So, in case of continues random variable, the probability of any discrete event turns out to be 0; obviously, because it is very intuitive to see that, it is not difficult at all. Suppose, you have got a continuous number sense actually, let us say, you you got we are talking about a number between these two, so obviously, if we talk about real numbers there are the infinite number in between, so if you take just one point out this then; obviously, the probability turns to be one over infinity, so that is not defined, so it, sorry, that is 0 because 1 over infinity turns out to be 0 actually.

So, what you have to how to go and about that, so we can talk about a probability of a single discrete event here, so then it turns out that we have to evaluate the probability of events within a finite time interval, let us say we talk about this interval now. Now, theoretically speaking these - I mean - there are a infinity numbers in between these two.

(Refer Slide Time: 10:41)



Alright, there are infinity numbers between the total - I mean - total interval also basically, so it turns out to be; now these sum is infinity over this interval and infinity over that interval.

So, if you denote it something like, let say, you denote it something like a b and then its small a small b, so what you do is - I mean - this is like probability turns out to be length of a b divided by length of A B probably. Then, this is something like infinity over infinity and obviously, there is a finite number associated after that actually. If you give a particular function like that you can use l hospital rule and you will get for that actually. Intuitively it is obvious; if you take some finite time interval then you can define a finite number associated with that and that turns out to be the probability of that actually.

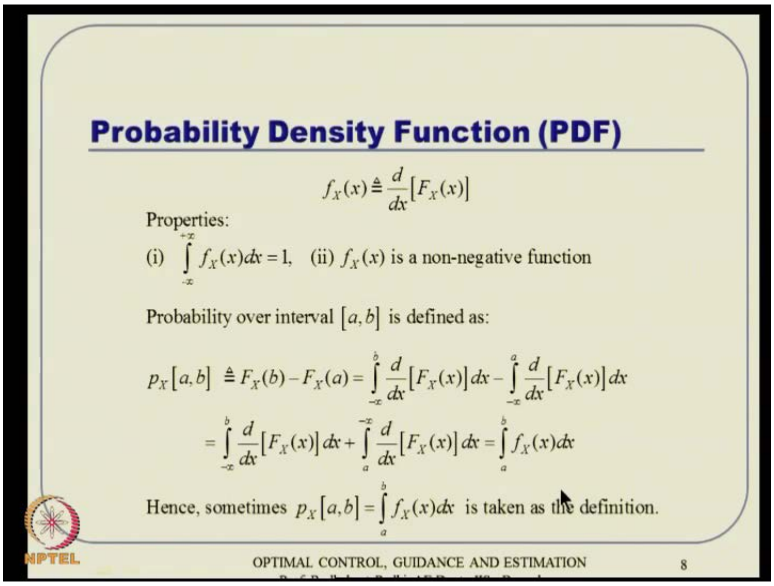
But how do you define all these - I mean - this distribution function and think like that, so there are concepts like cumulative distribution function, but cumulative distribution function is defined is something like this, f f capital X within bracket small x, that is defined as probability between the interval minus infinity to X.

So, it defines, it represents a cumulative probability of the continuous random signals X actually, for all events up to and including X, obviously right, this is, suppose, you start from of minus infinity and come on the way up to X, then whatever interval we are talking here, that will define what you want actually, s represents a cumulative probability of continuous random variable X for all events of 2 an including x basically up to this.

So, properties of this capital F x of x, it turns out to be 0 as x goes to minus infinity when x starts moving towards minus infinity, obviously, it turns the width becomes narrower and narrower, ultimately it turns out to be probable - this cumulative distribution function turns out to be 0. What about X goes to plus infinity, then it contains everything, and hence F of x goes to one, so how does it vary actually, if you see that if you want to see that how does it vary actually, then it turns out that if I plot something like minus infinity to plus infinity somewhere, and I talk about some P d f value of something like 1, then you starts from infinity, and then I start from 0 and then go towards that actually.

So obviously, F of x is a non decreasing function of X, it keeps on a continuously increases - I mean – increasing, start with 0 slowly starts building up keep on increasing, increasing, increasing, and somewhere it stabilizes at one actually, so that is the concept of distribution function - cumulative distribution function. Now, if you see interesting property of this distribution function that the derivative turns out to be 0 here, and turns out to be 0 here. So, obviously, derivative starts increasing and then decreasing and then it become again 0 and all the time it kind of remains positive, so you may see the slope actually. Slop is all time is positive initially t 0 final t 0 it somewhere in between it is maximum actually.

(Refer Slide Time: 14:37)



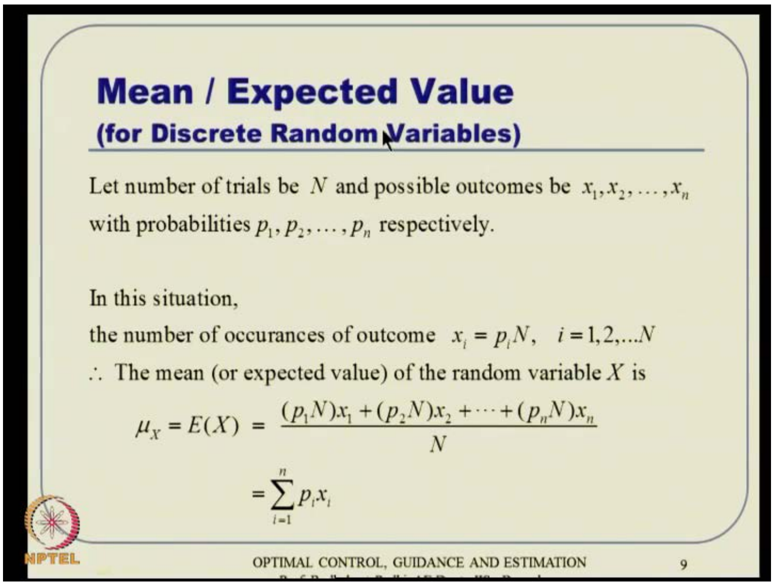
So, derivative contains lot of information and that is how this derivative is defined as probability density function, and this is what is - I mean - this is what will come across in lot of probability - I mean - lot of these random variables concepts and all that actually. So, the probability density function now is defined as derivative of probability distribution function actually, sorry, cumulative distribution function alright.

Now, this is defined a small f, so small f x of small x is d by d x of f x x actually, so what are the properties now, obviously, integration of f x from minus infinity to plus infinity is one and obviously, they told here, it is always a non negative function actually. Now, what you the way to defining this, it turns out that you can evaluate probability between as between some segment a b, once you know this actually how do that, so this is p x of a b is nothing but F x of b minus of x of a, so s minus infinity to b; obviously, by definition and this F x is nothing but integral of, sorry, this P x of a b is F x b minus F x a which is nothing but a integral of minus infinity to b this one, and integral of this one actually.

So, this turns out to be because this F of x capital f of x is integral of this small f of x, if you take the reverse, say, thing actually, so then if you combine that, if you work on that actually, what happens is minus infinity to b and change the interval - I mean - change the limits, so it talks about a to minus infinity, then it becomes positive actually.

So, now, what you can see is, I can think about now minus infinity plus this minus infinity goes, I can combine this two, and it turns out to be integral a to b, and d by n of f of x is f x actually, so I can put it that way. So, what sometimes people tell, I do not know to do all that, I just take this as a definition actually. So, p x of p x in the interval a b is integral of F x x d x actually, so once you know this small f of x, it is lot of information, because you can simply evaluate probability directly and many other things also possible.

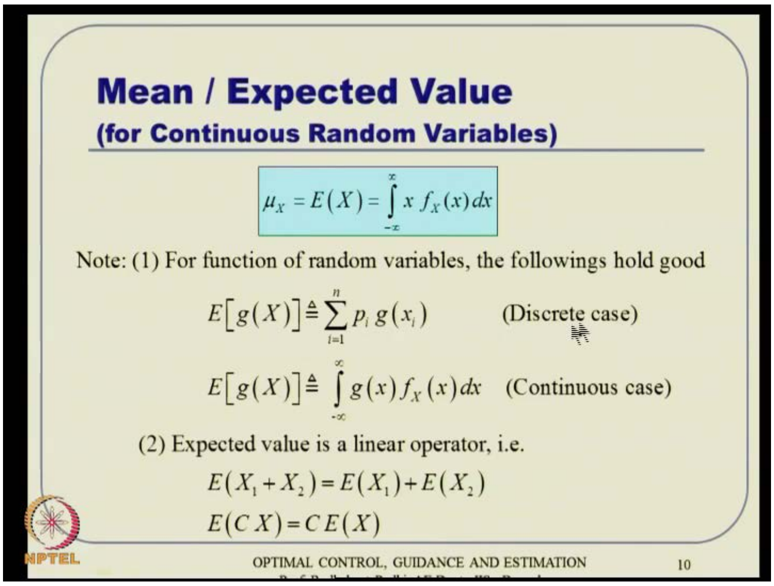
(Refer Slide Time: 17:21)



Now, something mean, an expected or expected value, and especially if it is a decretive event, rather easy to see, if you tell mean is nothing but, if I take n trials, the possible outcome or something like X 1 X 2 up to x n and each of that has probably it is P 1 P 2 P n, then what happens is the number of occurrence of outcome X 1 is nothing but X i in general is nothing but P i out of here into n, so this X i is, I can compute this way.

So, the mean tends out to be something like this P 1 of P 1 n into probability of that which is X 1 itself, and then it is P 2 of P 2 n to an n 2 and all the way up to n and then divide it by the total number of things actually, and n is the possible outcomes, so divided by total things actually. So, then it tends out that if n cancels out every where then it tends out that is nothing but the summation of p i x i basically, so this is rather easy to see but then what about continuous variable and all.

(Refer Slide Time: 18:52)



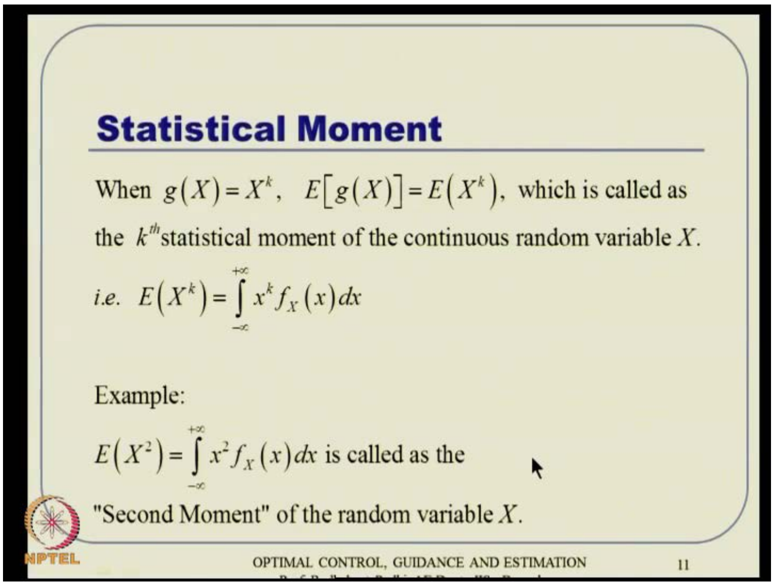
So, continuous variable is defined this way you talk about expected value, now here the expected value of x is integral of minus infinity to plus infinity x F x d x sort of thing, and what happens is for function of random variables, now suppose, you have a random variable but what about some of function of random variable.

Now, if you take a random variable X which is in discrete domain, it can take only discrete numbers and all that, so and you talk about of a function of that, then you talk about expected value of that then turns out to be like this, summation from i to i equal to 1 to n p i g of x i, but if it is a continuous case; that means, X is a continuous random variable then expected value is defined something like this actually.

And the great property of this expected value turns out that the expected value is a linear operator, so once something is a linear operator expected value is a linear operator, so one something is a linear operator its it is (()) we can do many operations rather variously, we will see that some of these things was - I mean - this particular expected value and all will be heavily used in kalman filtering also, and there we will use this linear operators behavior very frequently.

So, what do you mean by is an operator being linear, it satisfies the principle of the proposition; that means, if two signals X 1 plus X 2, and take expected value of the that then it is nothing but the expected value of X 1 plus expected value of X 2. Similarly, if you multiply this random variable with some constant c, then expected value of that turns out to be c times expected value of the values random variable itself, these two will be very useful later.

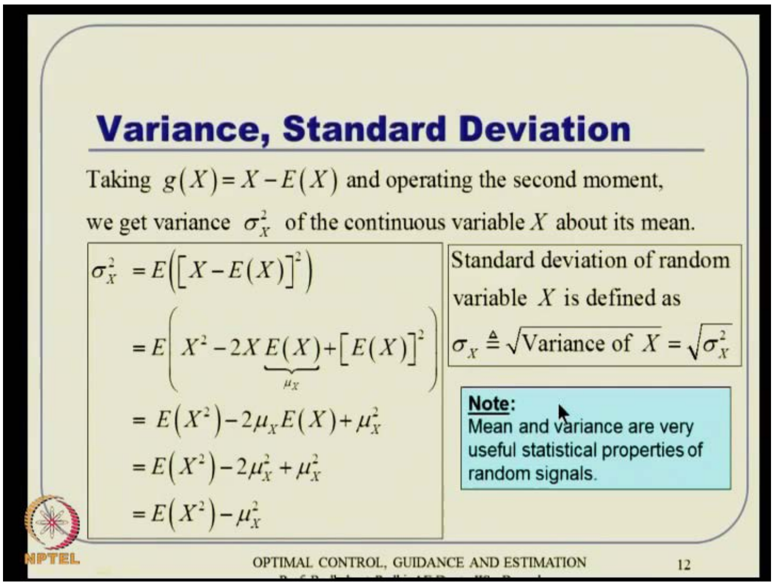
(Refer Slide Time: 20:52)



Then, a it is something called statistical moment and because this g is in general can be any function, so we are not interested in any function process about this is a typical function with X to the power k basically. So, when you take g of X is X to the power k what happens there in that particular case, obviously, is expected value of X to the power k by definition is g of X multiplied by p d f of that actually, so g of X multiplicities by this f of X whatever you see here.

And for example, if you talk about expected value of X square and then you have to do is integral of minus infinity to plus infinity, then k is 2 here and that is what you put here X square then evaluate this integral and this is typically called a second moment of the random variable, if you put X x actually not x square that is called first moment, so in general it is K th statistical moment of the continuous random variable X that is how it is defined.

(Refer Slide Time: 21:58)



Now, why it is useful, the second moment and things like that, because this variation concepts actually, so anytime we talk about a signal that is randomly varying the two things that is very critical, one is what is the mean valve, about which the vary - I mean - variation of auto reaction happens actually, and then after the mean value what is the spread of the valves actually, that is speared of the valves is given by the second moment and that is typically called variation and square root of that will turn out to be standard deviation.

So, what is variance? Variance is, what is the aspect random signal minus the expected valve; that means, value sort of thing, so you have some randomly variables, some randomly varying signal - a random variable X, all that you are telling is seeing the difference between the valves from its mean valve basically. And then taken that as a function this g of X, so then if you do that then you operate in the second moment and think like that, then you get this sigma square in other words, sigma X square is nothing but this particular derivation value what you get there whole square.

Now, essentially, this is the member, finally, the values is not known, what is the number, and this is also a number is a expected value, is a deterministic mean value sort of thing as a number, so when you have this sort of a operation going on x square, you can actually use this a minus b whole square formula, then it turns out to be something like x square minus two x expected value of x plus - I mean - expected value of whole square basically.

But remember, expected value of X is nothing but mu of x that is by definition actually, so what happened to this, this is nothing but mu x square by definition, this is mu x and that is mu X square, but expected value of X square we do not know - I mean - we have to just keep it.

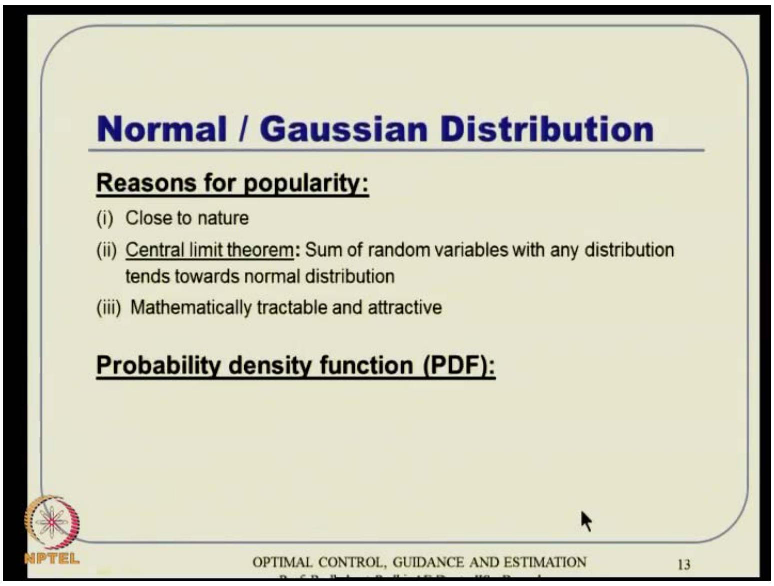
So, but remember now, we will use this linear property of the expected value as operator, so expected operator is a linear operator, so then what happens we can expand this bracket, and tell this is nothing but expected value of x square minus all the way the expected value can go - I mean - this is nothing but mu x me about mu x is a number remember that, the 2 mu x will come out, then expected value of x will come here, then it is nothing but mu X square.

What is expected value of x again is nothing but mu x, so this turns out to be 2 mu x square and this is plus mu x square, so it will ultimately turns out to be minus mu x square. So, sigma x square in general is expected value of x square minus mu x whole square - minus mu x square. So, this is a nice property, because you do not have to do this all the time you can just see expected value of X square minus mu X square.

Alright, then by definition the standard deviation is nothing but variance of X a square root of the variance of X, so sigma X is nothing but square root of sigma X square that is what the definition terms actually here. And typically, it is a positive square root actually, so the point here is this mean and variance or the rather standard deviation and all that are very useful statistical properties of any random signals, so somebody tells it is a random signal, we cannot tell where the number will lie, but if carries lot of experiment, we will be typically knowing what is the average value and what is the standard deviation or variance.

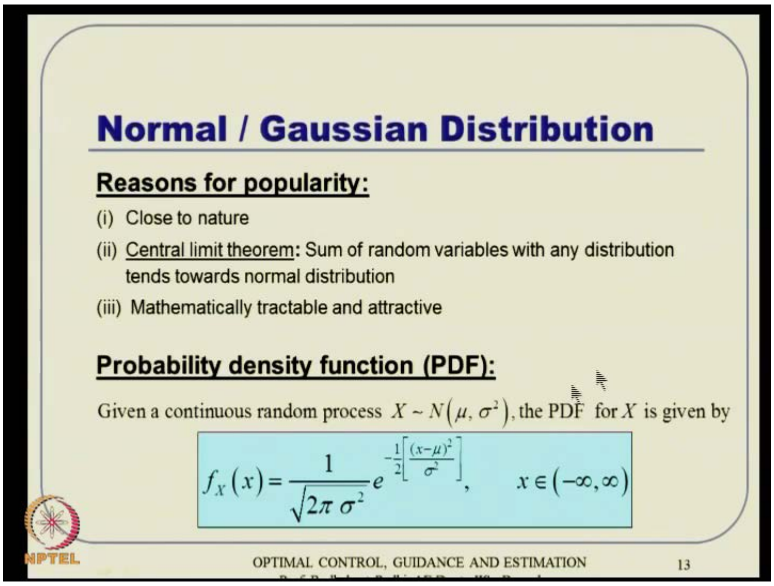
And in fact, the entire kalman filter turns out to be - I mean - making a kind of a track required between this mu expected value and variance actually, or other covariance what we will called that actually, because it turns out to be a vector signal, then see what we are talking here is scalar variables and all when it turns out to be a vector of random variables actually, then you talk about something called covariance, and then the entire filtering theory is expected value and covariance matrix that is what we will talk about actually, we will see that as we go along.

(Refer Slide Time: 26:36)



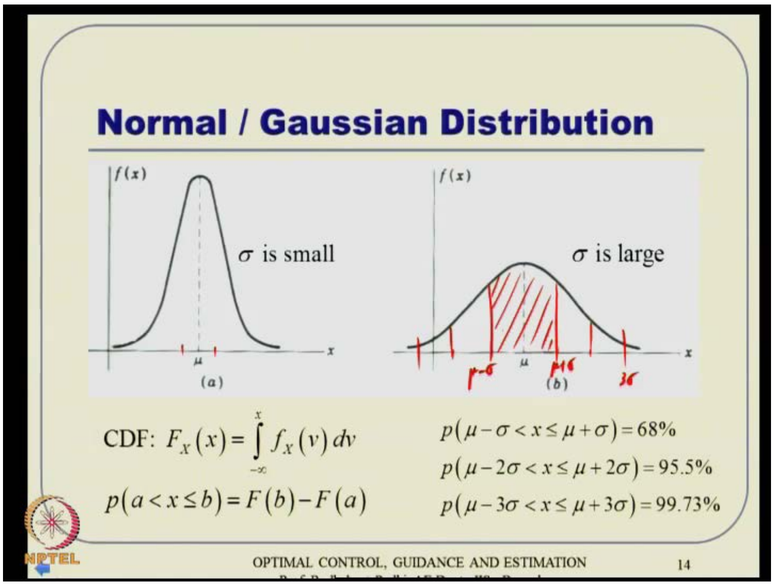
Then, this particular distribution is of very importance in any probability theory including kalman filtering, and the reasons for this popularity of Gaussian distribution is three things, first thing is it is close to the nature; that means by default many of the distribution turns out to be Gaussian actually, but on top of that this is a central limit theorem as well, which tells us sum of random variables with any distribution, you can start with any distribution if you keep on hiding them of, so some of random variables with any distribution ultimately tends towards normal distribution, so that is why this normal distribution or Gaussian distribution very popular actually. But that is not only thing is what is also good about it is mathematically tractable and attractive actually; that means, you can do lot of algebra very easily and what is the probability density function associated with that.

(Refer Slide Time: 27:42)



This is something like this, it is correct wised by mu and sigma square, and this probability density function turns out to be something like this, one over square root of 2 pi sigma square exponential minus of this, thing actually but x varies from minus infinity to plus infinity, take any value of X that is density function will give a number actually.

(Refer Slide Time: 28:08)



How does it look like? Well, if you plot it something like this, if sigma is small it will turn out to be something like this, and if sigma is large then its large distribution - I mean - it turn out to be like this actually.

So, what happens here is, it depends on, see, mu remains same probably both, both the plots, but if sigma is small, every most of the values are centered around mu basically, but if sigma is large then many of the numbers are deviated away from mu basically that is what it tells here, or del under the power probably will remain same, because ultimately if you take minus infinity to plus infinity integration both of that will give us 1.

Otherwise, if you take cumulative density function, this one what we discussed here, so a cumulative distribution function; cumulative distribution function is something like this, so ultimately probability between this interval a and b turn out to be like this, and it is very interesting property of that most of the time it is useful from two sigma onwards especially, and if you take something like minus sigma to plus sigma, in other words, you got a sigma value here.

Typically by this actually turns out to be something like three sigma value, so this one is two sigma, this one is sigma and, similarly, this is sigma this is two sigma, let us say, so if I take this and this one and then talk about probability between this interval, that is, minus sigma and well mu minus sigma to be exact this is mu plus sigma.

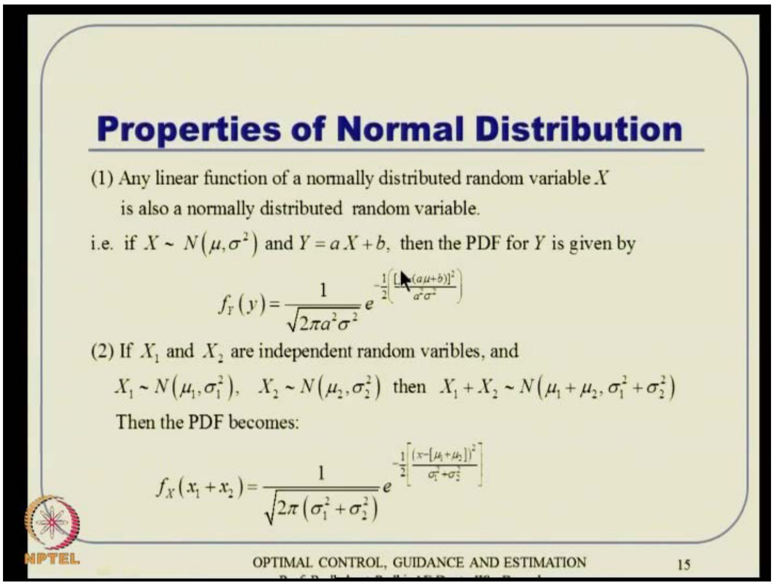
If I take the probability between these intervals turns out be a 68 percent; that means, if I take the area under this, and divide it by total area it turns out to be 68 percent. And then if I take two sigma; that means, between this interval now, it must write a it is 95 percent - 95.5 percent already, if I take three sigma it is 99.73 and for all practical purpose we typically stop at 3 sigma value the close minus 3 sigma value.

What is good? Remember, if you suppose you are estimating something, this is you do not know actually what is the value, but you are estimating something, you got this value mean and you tell all my signals are bounded - I mean - very close to the mu; that means, this is how it is like, sorry, they are ideally, this should not be like that, it tells my sigma is small actually, so that is typically very close to that, so most of the values will fall close to sigma really a close to mu, then it is a good estimate.

And I will have a lot of confidence in something where sigma is small, so if any sensor that you are using and then bring some experiment to characterize whether the sensor is good or bad, you not only see the mu, you should also see what is the value of sigma associated with mu. If sigma is small then the confidence on that value on that average value mu is much higher actually, but somebody tells us only mu and then remain silent on about sigma is not good also.

As if you carry out let us say - I mean - just to be simple two experiments and probably one value is minus point 0 1 and other value plus 0.01 then, the average value is 0 and another two numbers where the one is minus 1000 and another one is plus 1000, the mean is still 0. So, just talk about mean value does not gives us a complete picture, it always as the variation associated with the mean value that gives us a complete picture actually. Alright, so these are Gaussian distribution, and why it is mathematically tractable let us see actually.

(Refer Slide Time: 32:40)

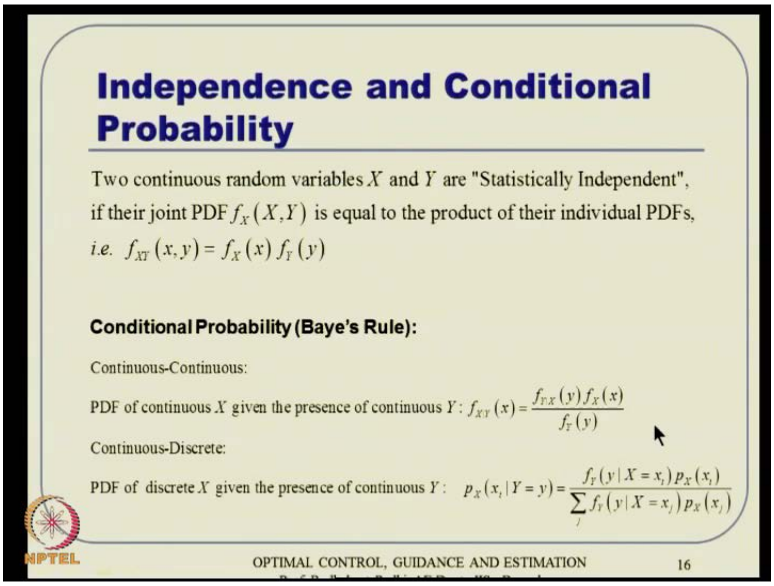


The properties of normal distribution something turns out to be like that, if I construct another random variable as a linear combination of this a linear function of this original random variable, something like y equal to a x plus b, then I can directly write the P d f of y basically.

So, p d f of y which turns out that the average value mu turns out to be a a mu plus b, hence the variance some sort to the sigma square, sorry, a sigma sort of thing, if you see if you compare this the variance turns out to be a sigma when the a square sigma square with the sigma square is original variance, so the variance of y turns out to be a square sigma square and average - I mean - that mu the is the original mean value for or expected value for x and the expected value for y will be exactly same linear function of mu; that means, if I know y is a X plus b, I can directly write what is the - I mean - mu of y and something like this and sigma of y will turn over to be a sigma basically.

That is one thing, other thing is if I take X 1 and X 2 are two independent random variables with characteristics being these things; that means, x 1 has a mean value mu 1 and then by the sigma one square similarly X 2 lines mean value mu 2 and variant sigma two square then X 1 plus X 2 will satisfy this; that means, its mean value will become mu 1 plus mu 2 and its variance will become sigma 1 square plus sigma 2 square. Then, the P d f became something like this F x of X 1 plus X 2 turns out to be something like this actually, so that is its easy if you know that is a normal distribution then many things can be done just by looking at some of these nice properties.

(Refer Slide Time: 34:50)

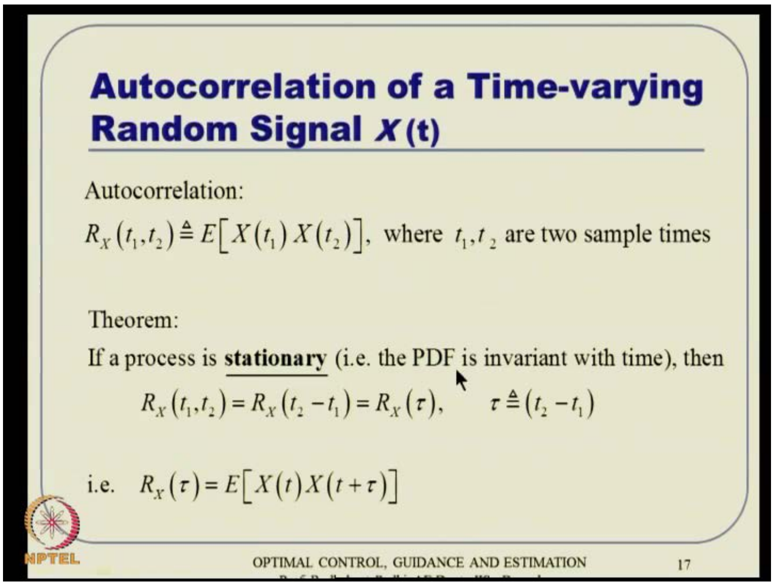


Then, these some concepts called conditional probability and think like that, so we have this two continuous random variables x and y are statistically independent, if there is joint P d f something called joint P d f now, where F x of X y is equal to the product of the individual p d f's; that means, if I talk about F x y of X y and turns out to be multiplication of both actually.

And this is what I am talking a little, little bit before this Baye's rule or region probability and think like that this is based on which people have proposed a very neat filtering ideas actually, and very recent literature if you the things are available Baye's and best belief or four filtering theory basically, so it talks about conditional probability.

And first thing is continuous-continuous; that means, if P d F of a continuous random variable X given the presence of a continuous random variable y is defined something like this, and that expression terms out to be like this. Similar, but if it is continuous discrete; that means, but if continues discrete; that means, the p d f of discrete x given the presence of continues y, it turns out be something like this little more complex.

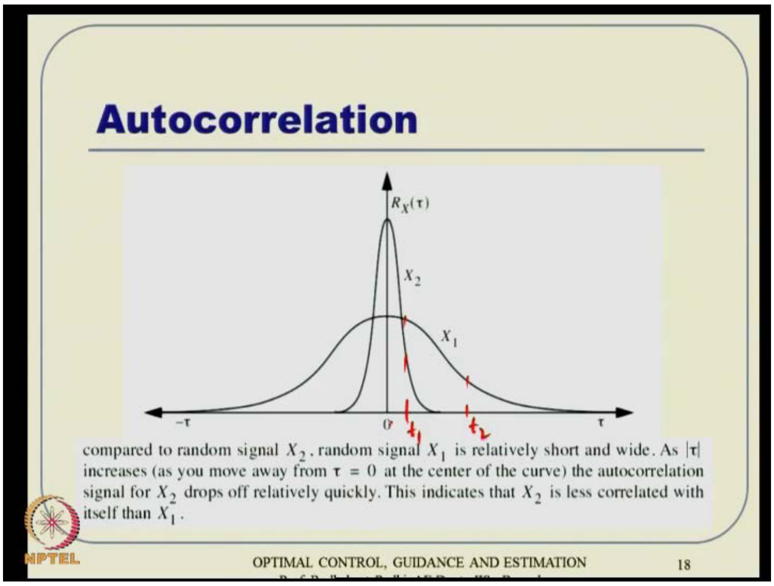
(Refer Slide Time: 36:12)



Here, we write need concept what you need very quickly this concept of something called auto correlation. Now, remember x x is a time varying function - I mean - time varying random signal, so X of t 1and x of t two will be different actually, so what about an expected value of x of t 1 into X of 2 if you take t 1 t 2 are two sample times then you construct it X of t 1 into x of t 2 and then take expected value of that what will turn out to be and whatever it turns out that is defined as auto correlation.

And the theorem tells us if the process is stationary; that means, by definition that pdf is in variant with time P d f does not un change actually, then this auto correction r x of t 1 t 2 is just of function between t 1 minus t 2 is just a function of t 1 minus t 2 basically, so r x of t 2 minus t 1 R x of t 1 t 2 is nothing but R x of t 2 minus t 1, the interval between them this tou. So, I can always write it as a function of tou, next we have to write t 1 t 2 and all this R x of tou that is expected value of X of t in to remember t 1 is t and t 2 t plus tou basically, so this is how it is.

(Refer Slide Time: 37:41)



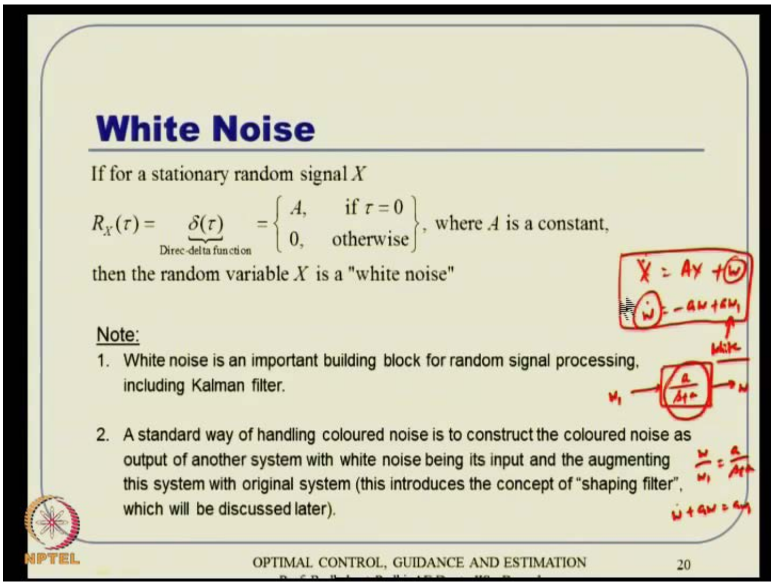
Now, this has a nice property, this R x gives us many interesting things, now let us see something like this, now we have let us a two signals X 1 and X 2, now we on to see which is more correlated with respect to itself - auto correction tells us that actually. If you see x 2 and x 1; x 1 turns out to be short, but it is wide actually, so if take a value t one here, let say, I put it out value t one here, and something like t 2 here, then there if it I take X 1 then there is a something which is non 0; that means, the multiplication of that will not turn out to be 0 basically.

But about X 2, if I take that one when this is a there value of one there is nothing but that is 0 basically alright; that means, X 2 signal is less correlated to itself basically, because if I take difference between t 1 and t 2 higher, then it is quickly going to 0 actually, but if it is a distributed - I mean - if it is a wide - I mean - auto correlation function turns out to be wider then it turns - I mean - it is a something like more correlated to itself actually, so this means.

This picture tells us that X 2 is less correlated with itself then X 1 actually X 1 is more correlated to itself because, if I even if the difference between t 1 t 2 becomes wider and wider, the numbers will not go to 0 very quickly basically because, this - I mean - this little turn out some numbers positive numbers a non 0 basically thus the reason.

Now, what is the limit incase in the that is the case, now let us say this width turn out to be smaller and smaller ultimately the width turns out be 0 than what actually; that means, if I talk about the same time instant, then I will get some value for this x of t 1 into 2 X of t 1 as some value, but the moment X of t 2; t 2 is slightly different from X 1; that means, t 2 is something like t 1 delta t were delta t is very small value actually, then this number is not there; that means, one of that is 0 and hence everything is turns out to be 0.

(Refer Slide Time: 40:20)



So, in that particular case, this something defined like this and that is nothing but white noise, but what about this white noise many times and all that actually or may be will see that in our kalman filter derivation also basically, so all that it tells us for a stress, first of all the signal has to be stationary, the P d f should not change, and then it turns out R x of tou is nothing but just a direct delta function basically. So, that means, if tou is 0 the you know interval between two numbers t 1 t 2k, then it turns out to be some finite value a otherwise it is 0 basically, so note that white noise is an very important building block for random signal processing and including kalman filter.

Standard way of handling colored noise is to construct the colored noise as output of another system with white noise being it is input and augmenting this system with that original system, for example, if you have this a let us say, let me give an example here, let us say you have this x dot equal to a x plus something like w, but w is not white now then what you do, so what you tell is, I will put construct another some sort of transfer function or some function here, whatever it is then where I give something like w 1, I will give w.

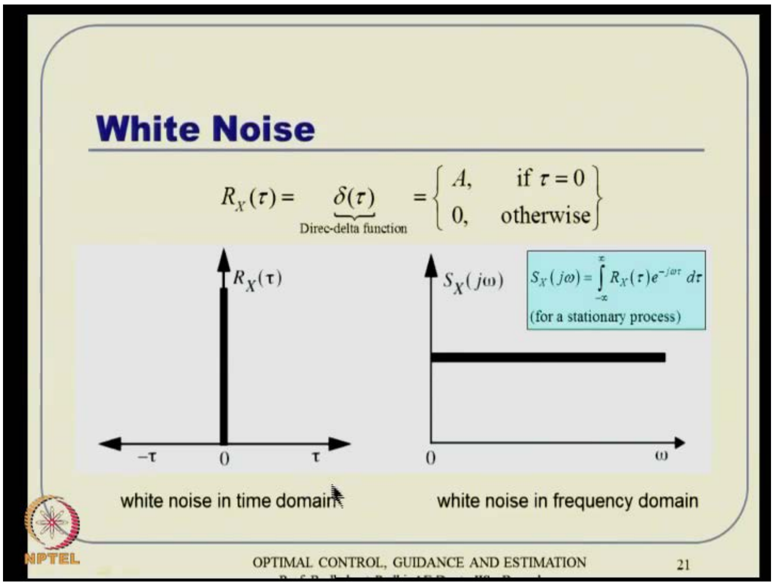
But this is a dynamic system remember that, so may be something like a first order system or second order, it depends on all that their noise behavior basically, but then what will do, suppose, it is a faster and then you put a w dot equation, and then tell which is something like, well, if you see this expression, it turns out to be, what actually, w by output by input is a by s plus a.

So, something like w dot plus a w is equal to a w 1 actually, so if you tell w dot is nothing but minus a w plus a w 1, Now, w become a state actually, if you see this thing is not a random noise any more, as a become a state, but this one happens to be random signal again, but this then modeling has to be done in such a way the w 1 turns out to be white actually.

So, that means, again you come back to the entire, if you see visualize the entire system w 1 has become the noise, now w 1 is white, so this is concept is called a shaping filter and typically this is not a very simple thing to get this transfer function what you talking here, but for important phenomenon these things are available as part of the modeling process basically.

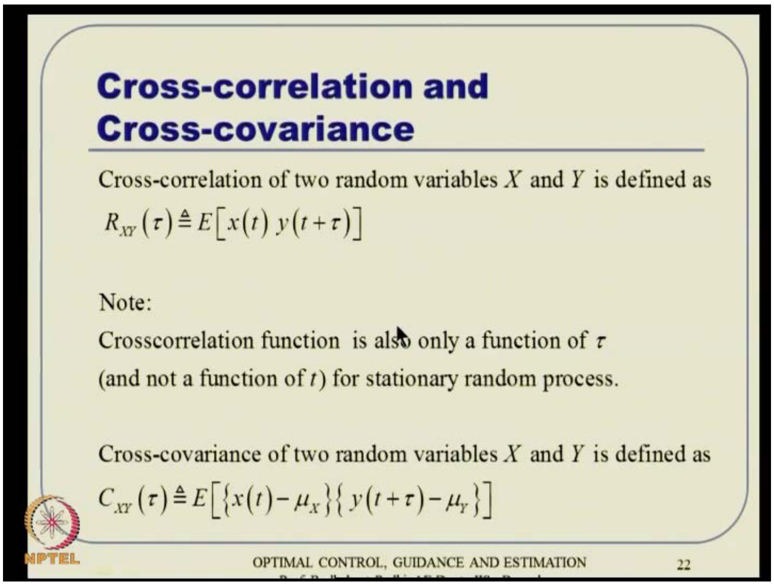
So, for example, if we talk about that let us talk about wind dust all that, there are guided model something like that way, so which will do exactly like this the guided model is all that all are it tells is, if I take a white noise and put it into in that the output of that function what I am having will gave you that particular noise which is physically happening basically, and that is called colored noise - something is non write is called colored noise actually; input will be white, but output will be colored, and then I can augment the original system that way were the input is still a white noise, and hence I can use kalman filter actually we will see an example in a subsequent class also basically that way.

(Refer Slide Time: 43:49)



Now, the same thing can be interrupted in an d different ways also, now if we talk about time domain and this R x r x of tou can be represented something like this exactly at 0 it is some value, that is not anything like this can also have some frequency domain interpretation as long as the process stationary, so you construct this for a transform sort of thing with then this is define something like this, and it turns out to be just a constant number, irrespective omega - I mean - if you plot it as a omega it turns out to be just constant; that means, it contains all frequencies spectrum actually, so there is an another interpretation and the frequency domain basically.

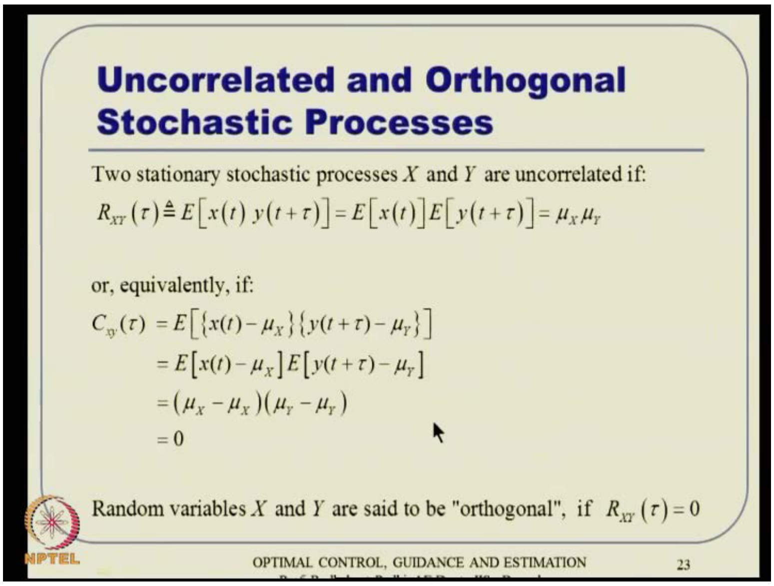
(Refer Slide Time: 44:36)



Then, there are concepts like something like cross correlative and cross covariance things like that. So, first thing is cross correlation, suppose, here we are talking auto correlation we talk about the same signal X and X, but suppose those signal is different we have something like x and y then what actually.

In that case is no more auto correlation, but it is something like cross correlation, and this is very similar to that, but the fact is it is not X here actually, and again cross correlation function is also only a function of tou; that means, is total depend is not really a function of t, but the function of tou really, the difference between t 2 and tone sort of thing not a difference between t 1 and t 2 and if you have a cross correlation you also have a cross co variance, the cross covariance is, you find something like this the expected value of an x minus mu x and multiplied with y as s function t plus tou minus mu y actually.

(Refer Slide Time: 45:42)



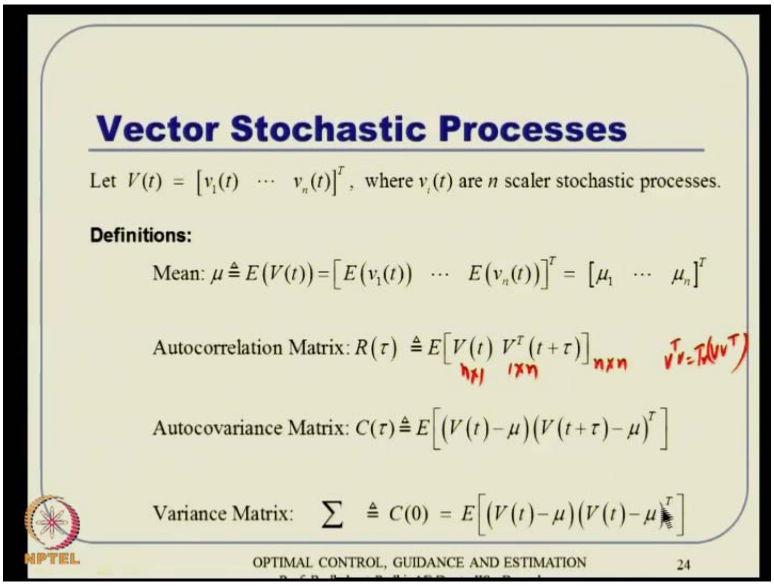
Then, something called stationary stochastic process and all, then or when if this, when the stationary process is x and y are un correlated when they are this cross cova - I mean - this function that we just talk about cross correlation function turns out to be just multiplication of this expected values of the two.

If you talk about alright the stochastic process x and y, and somebody tell are these are not correlated, these are uncorrelated, and this operation- I mean - this result can be used actually and in fact will see in kalman filter derivation many times will use this actually, because process noise to sensor noise is uncorrelated and an initial condition to something like sensor noise is uncorrelated like that actually will see many thing getting use there.

So, you can defined this way or equivalently you can define something in terms of cross covariance actually, instead of an cross correlation which is expected value of this - I mean - multiplications only you talk about the deviation of that from there all mean values and then take a multiplication of that and that turns out to be covariance. So, the cross covariance matrix of that can be written like this. And then if you see, then if you see this is nothing but this expected value of this in to expected value of that, because these are not correlated we just talked about something somewhere actually - well may be somewhere is there.

So, if you see thing expected value of these two, so this are uncorrelated, so when expected value, I can talk about that multiplication of these two then, I can use this expected value is a linear operator, I can take inside the mean of x - I mean - expected value x of t is mu of x, and expected value mu of x; mu of x is a number constant number; any constant number and the average or respective value of that is just a constant number, so if way that way. Then, expected value of that is again by definition mu f y minus mu of y, so it turns out to be 0 actually. And this property will be again will be varied actually and the infinite definition called orthogonality and random variables x and y are said to the orthogonal if this R x f tou turns out to be 0 basically, if I multiply this two turns out to be 0, then this signals are called orthogonal.

(Refer Slide Time: 48:38)



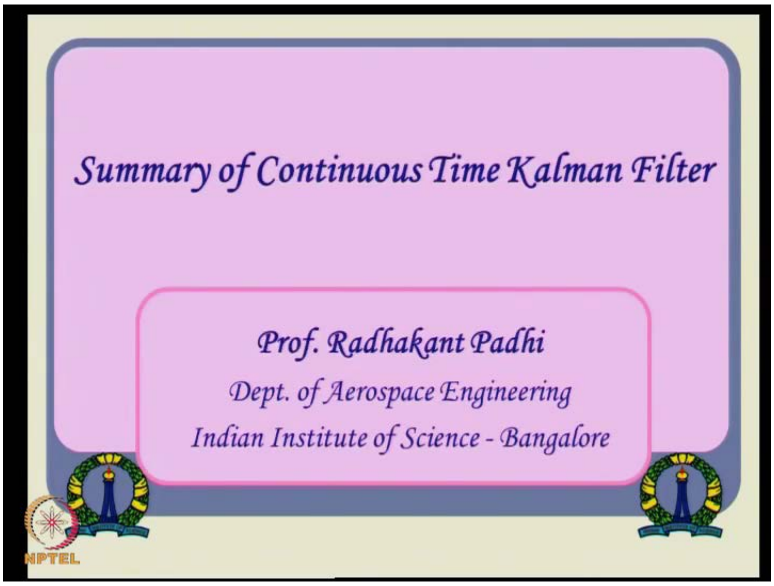
All these are in the form of some scalars, then what happens it term of vectors actually, so when you have a vectors of V which is un components but each of the component are something like stochastic process actually, scalar components stochastic process then what actually. Then the mean value by definition is nothing but the means of the everything they are once of definition actually, autocorrelation matrix, now it is V v transpose, and these are some something like outer products actually remember that V is that why V of n by 1 matrix, but this n by 1, then v transpose this is n by 1 and obviously, v transpose will be 1 by 1, so the total thing will be n by n; that means, use actually a matrix.

Similarly, auto covariance matrix, this can be define something like this V and V are typically have two products of the time not denote products actually, but also remember if you have some out of product wait let me see how to P of some outer product something like V v transpose and really want to the denote product, v transpose V then it turns out that is not trace out base actually. So, when we do out of product, it contains inner product information as well, just take the diagonal elements and submit out that will turn out to be inner product actually.

So, the covariance at - I mean - auto correlation matrix, auto variance matrix now and similarly the variance matrix, something this and these are something we will call about co variance matrix actually in thermal field. We talk about signal minus there average values, thus the variance from the mean value, and then operate it again with respect itself was in the sense of all to product actually then you will get that.

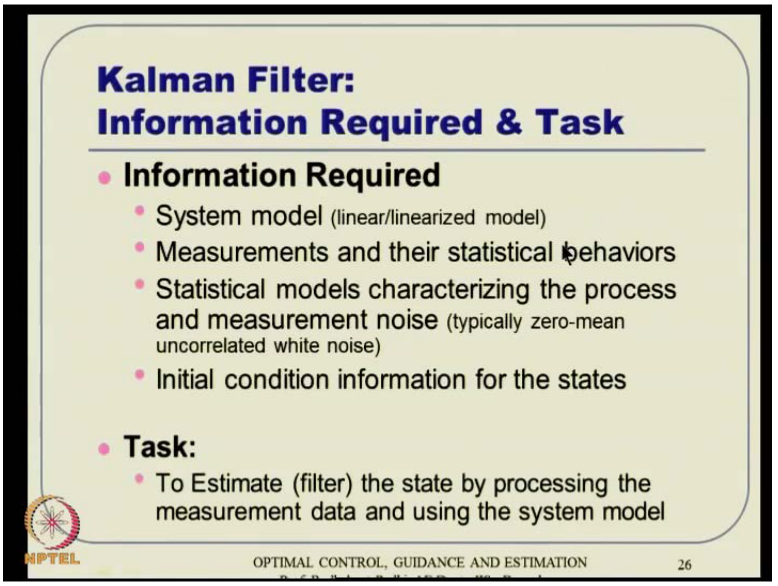
Alright, so this is some sort of a little bit over view of probability theory and stochastic process and some definition as we said to that and all that before stopping this lecture again whatever we discussed in the last lecture let us revisit little bit.

(Refer Slide Time: 51:07)



What we are interested in this particular lecture series is, gearing of towering this continue this kalman filter in general. And then looking back what you discussed last lecture we discussed something like continuous time kalman filter followed by E k. So, E k, let see how systematically the kalman filter in the linear domain can be derived and all that will do in the next class, but summary sort of thing, this is what it is, kalman filter what we required then information and what is the task.

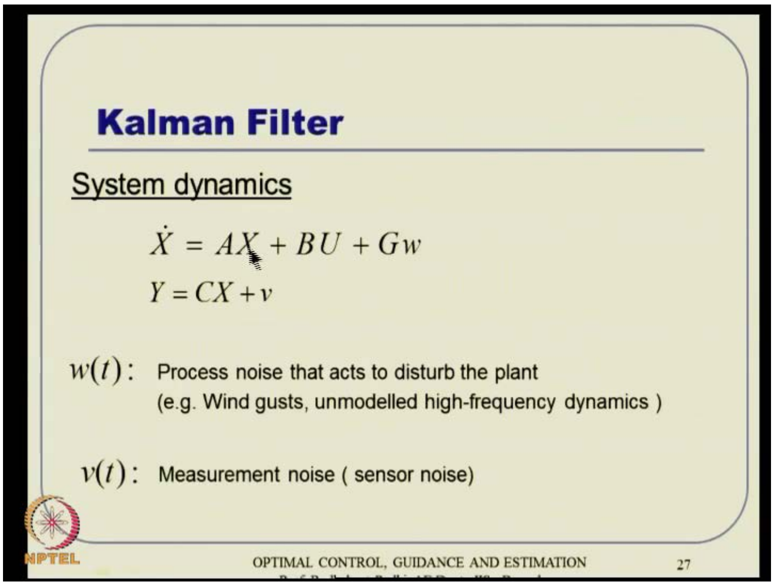
(Refer Slide Time: 51:36)



So, information required is something like a system model, and we need a linear model forsake, if they, if it not linear it needs to some linearized first, and then you needs some measurement and statistical behavior and also need statically models charactering the process and measurement noise and typically will assume that they are 0 mean uncorrelated white noise in typically, now it should what mean by 0 mean what you mean you by on correlated signal and what by means white noise, should all make sense now basically.

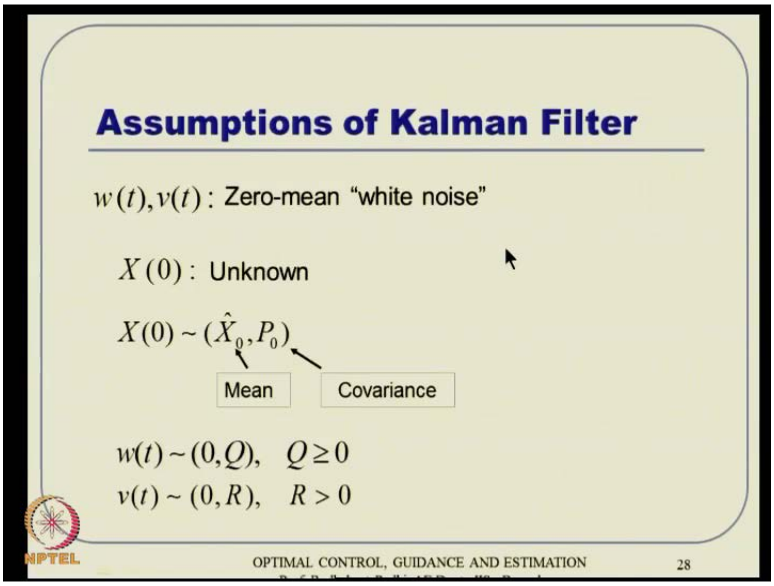
Then, also we will need initial condition information for the states actually, what is the task? Task is to estimate filter the state by processing the measurement data and using the system model.

(Refer Slide Time: 52:32)



So, how do you do that, very quickly, we have X dot X dot equal to A X plus B U plus G w now let us w g is something like process white noise process is white, for example, wind guests un modeled high frequency dynamics something look like that, and we have this measurement Y which is nothing but C X plus v, this v is measurement noise, basically they are process noise then we have measurement noise.

(Refer Slide Time: 53:00)



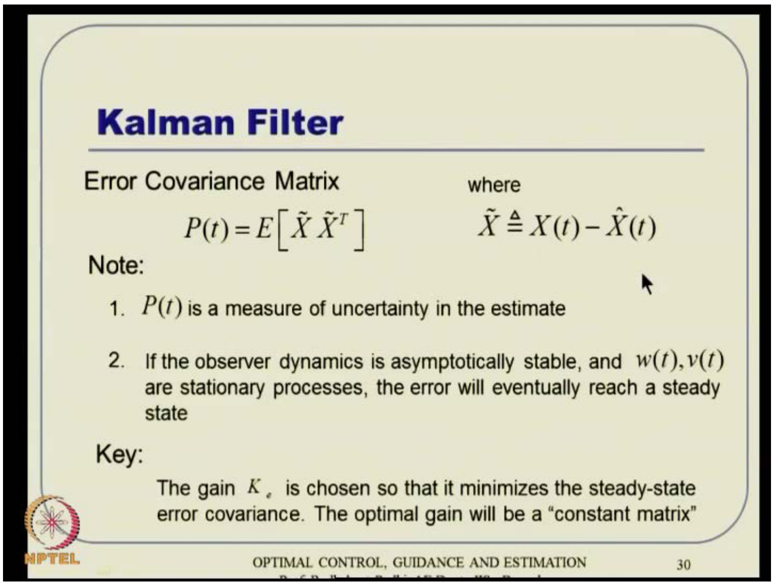
So, assumption is this noise is are 0 mean white noise, now again it will should make sense now what is what white noise and all that. Auto correlation from sense turns out be till the delta function actually and except 0 is unknown initial conditions are typically not known really, wheather it is characterized by some sort something like in mean value and its co variation actually, how much it will distribute away from the mean value.

And w of t v of t is tiled of 0 mean white noise and 0 mean, first mu are 0, and the kind of covariance is nothing but for w t and Q and for e it is R. Alright, so kalman filter talks about something like this, we have got the estimate of something like observer dynamics actually, and it is very close to what we know this is dynamics here, A X plus B U plus G w; obviously, this quantities G w and v cannot b processed, because they are noise quantities something look like that, they are directly use for some processing only in particular filter like they are, everywhere else these are concepts we are used computational concept.

This is estimate of dynamic value that A X dot plus b u plus B U plus K e times y minus y dot, where y dot is nothing but excepted value of C X plus v here that is nothing but Y actually - I mean - this by definition Y dot respective value of this.

So, then expected value is linear operated again, so you can expand this and then expected value e is 0 means noise, C can turns out to be something like this actually, so Y dot is nothing but C X dot. So, that is something like expected output sort of thing, this is a true output and this is expected output, these is a different between them, and that is how it will look this the dynamics or estimated dynamics operate actually.

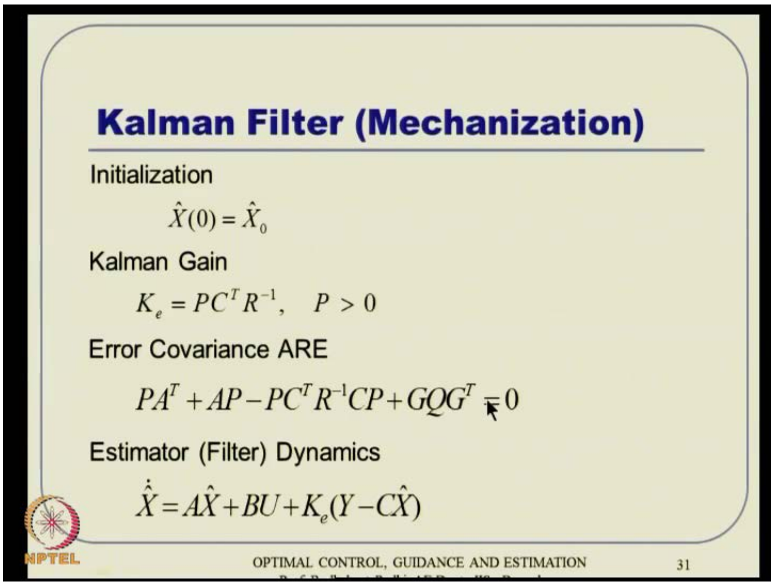
(Refer Slide Time: 55:19)



So, kalman filter also define something covariance matrix and covariance matrix here turns out to be definition, by definition expected value of X tilde dot times of X tilde transpose, where X tilde is X minus X of t minus x tilde of t this is no two state and no two state difference between that is X tilde. So, P of t is a measure of uncertainty is this, in this in the estimate that is why we are more interested in P of t basically. We estimate something, but we also want to know major of uncertainity in that particular estimate, that is why p of t basically.

If the observer dynamics is asymptotically stable turns out to be like that and w of t v of t are stationary process; obviously, the error will eventually reach a steady state value, and the value of here is that need to design this k view this thing still needs not, so what is K e it does not given that, so the gain k e is chosen so that it minimizes the steady state error covariance matrix and then optimal gain will be acting like a constant matrix, we will see that in the next class, we will derive all that actually.

(Refer Slide Time: 56:36)



So, the implementation sense we have some of the cell condition initial condition X - splitted initial condition - are there starting from the we can compute the kalman gain, where this P nothing but positive definite matrix where this filter Ricatti matrix solution sort of this thing. So, the filter Ricatti equation matrix is something like this, its needs to be solved for p, and hence if you get for P k e is ready.

Finally, K e is really than estimator of filter dynamic is done that way, so starting with this initial condition we can propagate it based on the actual measurements actually - that is it. So, thanks for attention and then will continue the derivation in the next actually; thank you bye.