NPTEL

NATIONAL PROGRAMME ON
TECHNOLOGY ENHANCED LEARNING

IIT BOMBAY

CDEEP
IIT BOMBAY

ADVANCE PROCESS CONTROL

Prof. Sachin Patwardhan

Department of Chemical Engineering,
IIT Bombay

Lecture No – 20

Soft Sensing and State Estimation

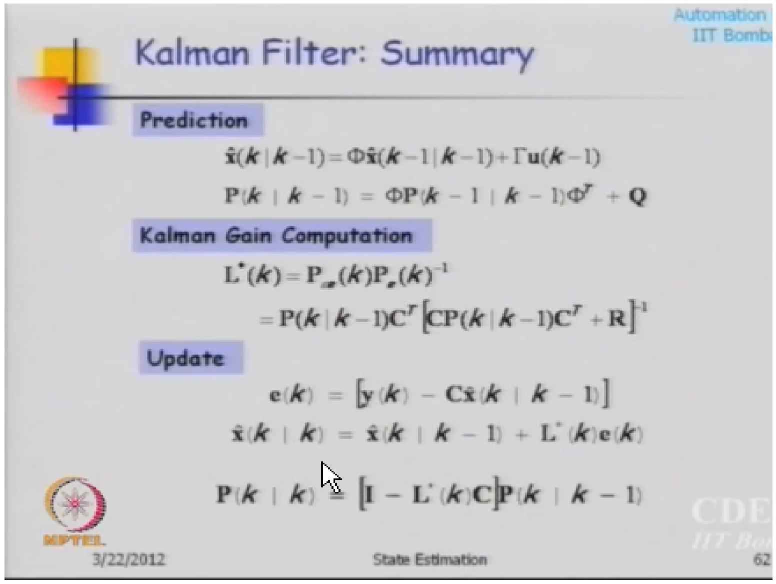
Sub Topics

Kalman Filtering (contd.)

Okay so let us continue with filtering and last lecture I introduced to you a derivation of Kalman filter there are multiple ways of deriving Kalman filter I through I try to do the simplest one simplest derivation to a derive Kalman filter if you see I just went through optimization approach I there is a way to go through a statics and probably density functions and so on I am going to hit.

I am going to make those connections now but those derivations are little more algebraically complex if you through the earlier one was simple then but if you go over it actually I would strongly recommend that it sit and try to derive the expressions in the notes those 5, 6 pages it is easy derive those expressions it is not so difficult, if you derive those expressions it is not so difficult.

If you derive those expressions yourself you will get confidence of about what is happening so let us just summarize kalman filtering algorithm in fact my today's lectures unfortunately.
(Refer Slide Time: 01:37)



Today not many people are able to attendant because of this so next our lecture I think it is equally important are probably more difficult lecture than yesterdays lecture so let me summarize because now in this lecture I want to go beyond what I thought yesterday, yesterday I through this derivation of this kalman filter first of you have to understand that we are dealing with a stochastic process x which I governed by a difference equation okay.

In a stochastic process you can do prediction okay prediction okay prediction is why do you develop models because you want to do predictions okay so prediction is why do develop models because you want to predictions okay so prediction is integral part of a so we predict this state without before receiving the measurement this is my state prediction this is what is expect this is the best value of the state best guess for the state okay.

Now this is as we said as we have seen that this is nothing but a conditional mean this is a the conditional mean of x given information up to k – 1 okay then it is not sufficient to characterize only mean I want to characterize a uncertainty associated with the mean with this random variable so covariance comes into picture so we have this expression for the covariance okay so this tells you that the covariance associated with the prediction is function of the covariance with the pervious updated mean okay.

See x^ k given k – 1 given k- 1 is the pervious updated mean okay covariance associated with this is P(k) – 1 given k -1 okay this only relates the uncertainty at the pervious time step okay with the uncertainty at the current time state okay so this is a stochastic process in which the density at a new time step is a associated with the density at the pervious time state okay it is function of it is a function see in fact the view point I am going to present today is that these densities are nothing but you can be viewed as Gaussian normal densities okay.

So you can view this as a Gaussian normal process Gaussian normal densities Gaussian normal densities can be characterized by mean and variance or mean and covariance so it is sufficient for me to have you know update of mean and update of covariance okay actually what I am doing is I am describing a stochastic process through propagation of it is mean in time and propagation of it is variance in time okay.

So this is predicted mean this is predicated covariance okay now when the measurement arrives okay I want to fuse the measurement with the predicted estimate I want to do a correction the correction step or the update step okay for that you need this gain for kalman gain is computed we saw that optimal kalman gain I am calling it L* in the previous slides I have these L(k) is any obituary gain L* is the optimal gain okay the optimal gain can be computed using cross variance of estimation error and innovation ε is estimation error okay.

And e is innovation this cross variance multiplied by innovation variance or innovation covariance inverse pre course multiplied it is very important here course multiplied these are matrixes okay and this is that quantity we had computed this quantity earlier we had computed this quantity earlier separately I am just putting the expressions now so there is a recursive expressions now so there is a recursive expression to compute the gain okay.

This gain if you notice is the function of the uncertainty associated with the estimate that is P (k) it is also function of now this P(k) given k – 1 is function of Q, Q is the uncertainty associated with the state error what is R, R is the uncertainty associated with measurements okay R is the uncertainty see what are the 3 sources of uncertainty in you difference equation one is the state noise wk okay.

Second is remember initial state this itself see this is a difference equation this is the initial state this initial state itself is uncertain right so this initial state is uncertain so there are 3 sources of uncertainty initial state when you go from k – 1 to k okay second source of uncertainty is wk okay 3rd source of uncertainty is P(k) measurement noise okay now all 3 play role in this gain estimation.

Okay all 3 of them play role now if you see this okay well I should not be talking in terms of rations because this is this matrix into this matrix inverse but in some sense if you here this you know uncertainty predication uncertainty + you know this Q which is coming up w(k) okay divide by in some scene divide by I should not be using divided by because this is a matrix okay but if you do a scalar case you will see that a ratio will come and you know this R here is coming in the you know in some sense it is coming as a denominator.

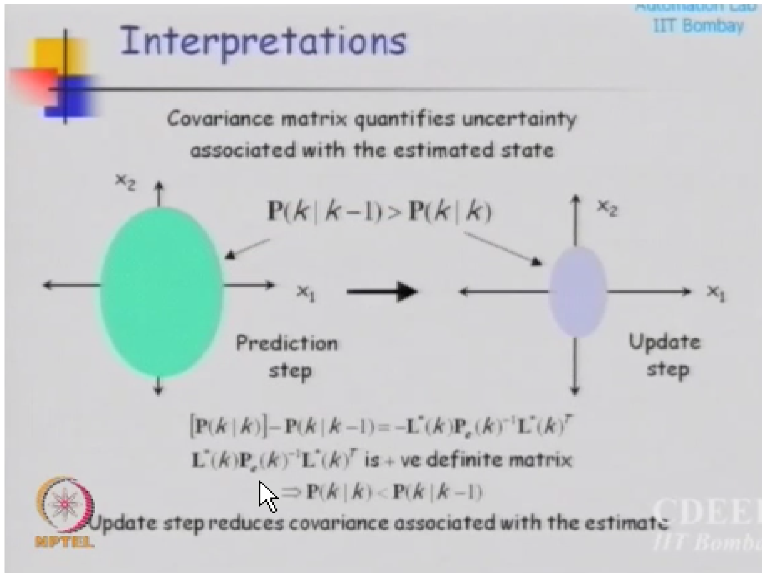
If you take scalar case it will come a denominator so what it means is that we are trying to find out that fraction see this Q appears in the numerator as well as in the denominator okay where as R appears only in the denominator okay so we are trying find that fraction which is because of w okay we want that gain because of w and then we want to use that to correct the estimate okay now update comes through the correction, so L(k) times e(k) and where is the information about w(k) and v(k) contained in.

It is contained in y(k) so y(k) – predicted out invocations will have that information okay some fraction of it what is that fraction decided by this covariance decided by this L(k) matrix gain matrix this L(k) matrix is used to fuse the measurements with the state estimate of the model so my model remember is running in parallel in computer only thing that is coming from outside okay are data one thing that is coming from outside are that is coming from outside is the measurement y okay u is something that computer decides either a controller decides or operator decides you know as a operator u are talking so you know those values of u okay.

So I am going to do a correction here in my state and this is the uncertainty associated with the correction see this is the uncertainty associated with the prediction this is the uncertainty associated with the correction okay now what I expect is when I correction uncertainty should reduce I have got some information from the plant okay I use it information and use it correct the state estimate this is my correction state right.

So ion some sense x^k given k should be a better estimate of the state than k given k – 1 okay so what I would expect is that uncertainty associated with it okay with xk given k – 1 should be higher larger than uncertainty associated with the k given k this is indeed the case okay.

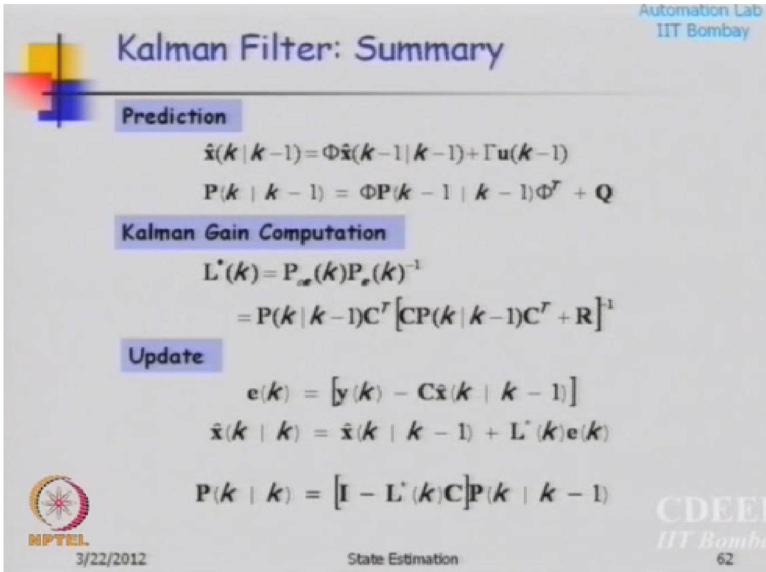
(Refer Slide Time: 10:35)



Yeah prediction step yeah because you have taken the expected value of w(k) to be 0 so we have taken the best estimate that we have brought or mean of w(k) that with 0 we add the state noise something equivalent to state component some compensation for the fact that state noise as actually influence x okay no you the what is unwanted for you is the measurement noise okay.

W(k) is an real input I mean R you can say that it is an imposter for a real input we try to create a model that kind of imitates some real input okay which is influencing the state some uncertainty okay we do not have an explicit handle for it we do not have an explicit measurement we are trying to create a sudo compensation for it through this see yk brings an effect w(k) – 1.

(Refer Slide Time: 11:47)



Okay and then somehow I want to use that information and you now inject it into my estimate okay so this correction is an attempt to make an composition of for that using okay see have we done this kind of a thing earlier we have we are look at series models ARX models box models I am going to do a connection in today class hopefully I am going to show that actually what you are developing is a form of times series model.

And I also show that the steady state kalman filter I am going to take about something steady state kalman filter, so steady state kalman filter is nothing but box model okay so what was the attendant time series modeling unknown input modeling right unmeasured disturbance modeling okay here is a same thing okay see in box model or in R max model we had this ek right ek was innovation.

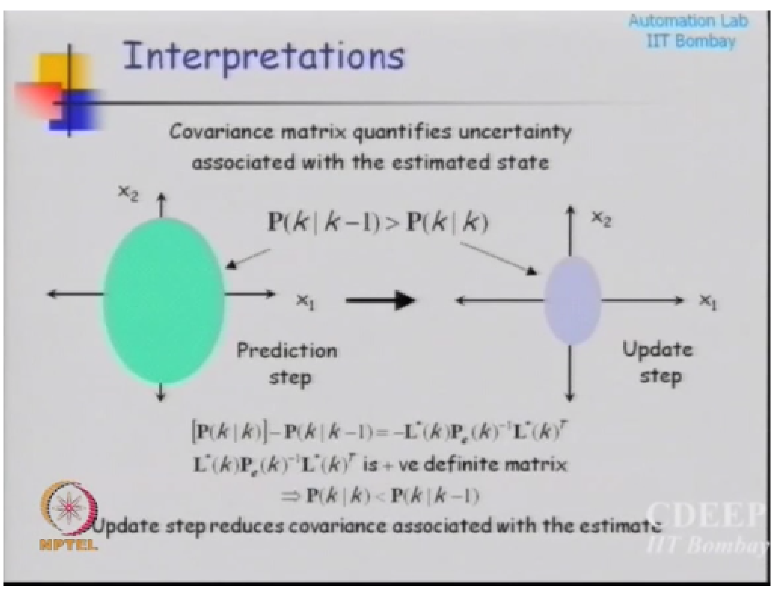
Innovation was yk – y^k given k – 1 okay this ek and that ek are not different I will do the concretion then you will realize that actually you know where taking same thing into 2 different languages okay is the same idea different ways parameterization different approaches actually finally led you to the same fundamental idea okay.

So actually in your assignment when you develop a let us say R max model okay and when you convert it into state space form actually you have identified a kalman filter from data okay so I will come to that but you know historically these 2 things have probably immerged completely differently and then now you can see those global connections you know or many be those who developed it knew all those connections.

We are late in understanding those so if you see lunges book so if you see lunges book we will talk about connections with kalman filtering very clearly only thing which I am somewhat disappointed is that he has put it as a exercise problem hidden in so it should have been permanently you know there in the main part of the text saying that actually this is the kalman filter it is not.

So unless somebody goes and solve those exercise problems you would do not realize that you know actually a box model or air max model is a parameterization of kalman filter okay any way we will do this connections, so is this clear okay so we are hoping in time okay we are going from time k -1 to k and every time we are finding the new density function associated with it okay. Now where is a density function coming in till now I only talked about mean and covariance.

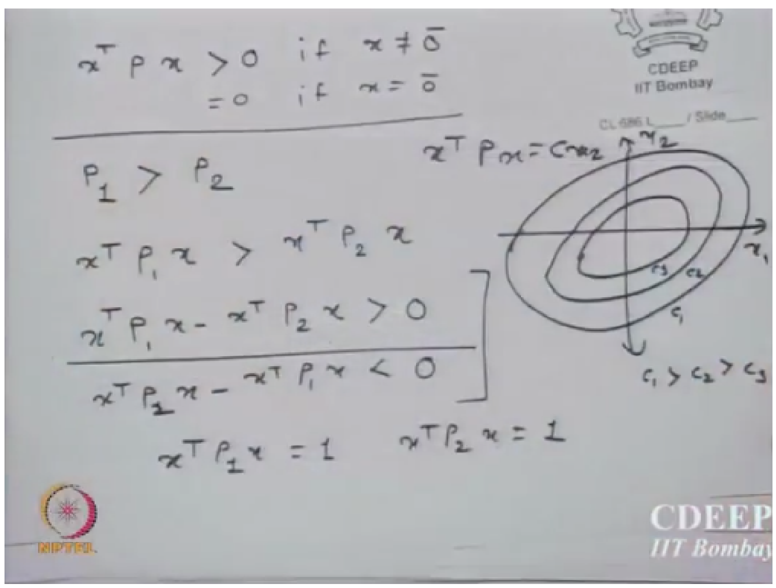
(Refer Slide Time: 15:08)



let us now try to associate some real density function which is before that let me say something about covariance reduction okay the nice thing about positive definite matrix is okay is that you can compare them you can say that this matrix is greater than this matrix only for positive definite or for negative definite of course in definite matrixes you can do this positive definite matrices you can talk about a matrix one positive definite matrix being > the other positive definite matrix okay.

So actually a positive definite matrix can be associated with some kind of closed counters in the x1 suppose you take a simple case of x1, x2 okay and if you draw a the locus of points see what I am trying to show here is that a no it is x transpose px comparison.

(Refer Slide Time: 16:39)



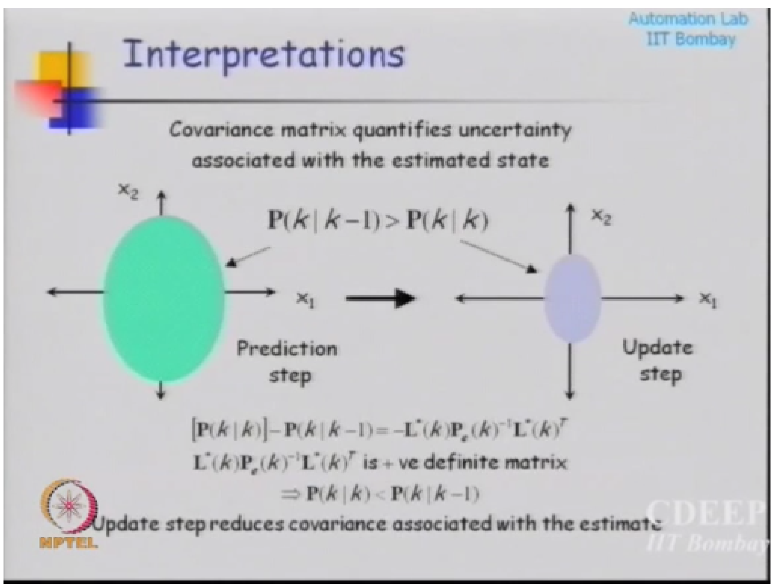
So here see what is a positive definite matrix is $x^T P x > 0$ is $x \neq 0$ okay and this is = 0 if x = 0 at that right, this is the positive definite matrix okay I can actually compare 2 matrices if we say that matrix P1 > P2 if $x^T P1\ x > x^T P2x$ or $x^T P1\ x - x^T P2\ x > 0$ okay or other way you can put that is x^T P1, P2 x - x^T P1x is < 0 either one of them okay so actually if you take set of all points such that $x^T P1\ x = 1$ and $x^T P2\ x = 1$ okay.

If you take this locus actually in general first of all know that if you take a locus of points $x^T P x$ = constant okay so this will turn out to be ellipses this will turn out to be closed ellipse like this if you take let us say x1, x2 this is in two dimensional case so this is x1 and this is say x2 then in 2 dimensional space you know so this will be some c1, c2, c3 and actually c1 > c2 > c3 and so on okay so locus of these points for a given positive definite matrix are ellipses okay.

And when you have this in equality which means that if you camper the ellipses of P1 and P2 okay ellipse created by P1 will be larger than the ellipse get but P2 okay so actually this ellipse size for a fixed say if you take $x^T P2\, x = 1$ and $x^T P1x = 1$ if you compare these sizes see then you can get idea of size of uncertainty see if the ellipse is smaller uncertainty is smaller if ellipse is larger uncertainty is larger.

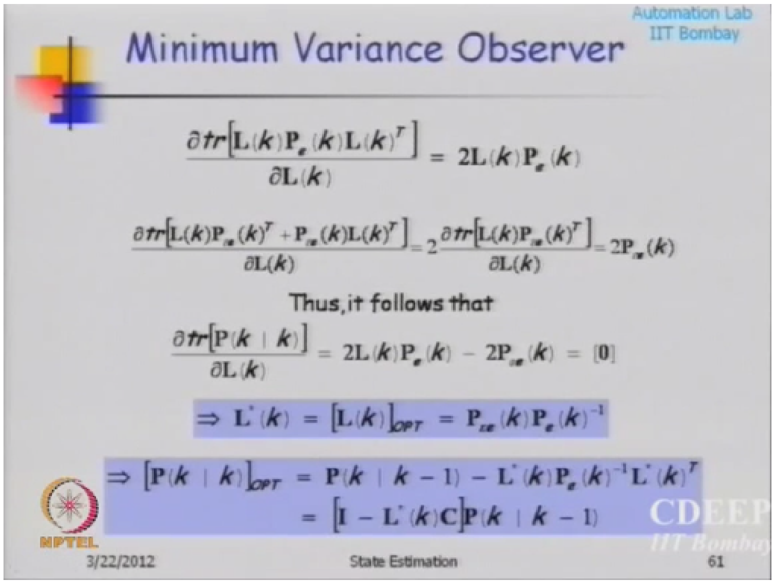
Larger covariance will give you larger ellipse smaller covariance will give smaller ellipse okay so what I want to show here well I have not drawn exactly the ellipse have drawn here.

(Refer Slide Time: 20:10)



Are aligned along the access in general for a general positive definite matrix they will be aligned they might be tilled okay so I have just create a carton which is convenient to draw it was difficult for me to tilled this if someone one of you can explain how to tiled this ellipse I could improve this carton but the idea is that what is happening is that this matrix P(k) given k is actually smaller than P(k) given k -1 why am I saying this.

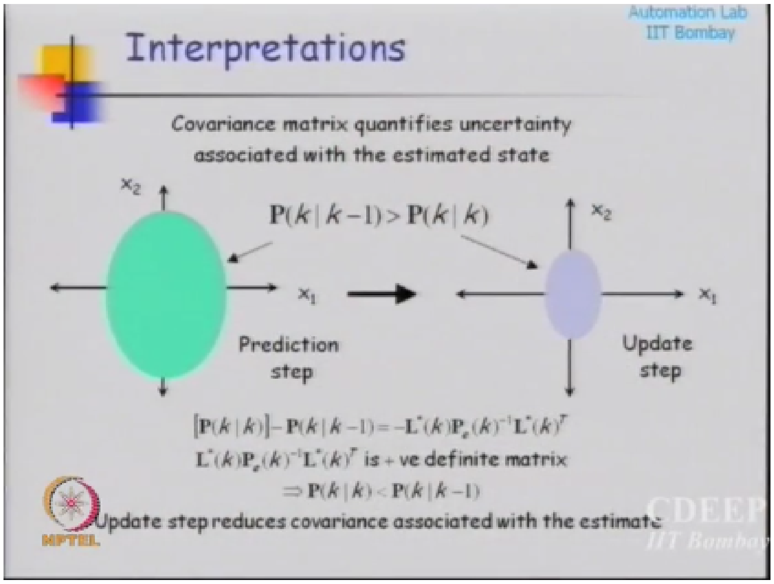
(Refer Slide Time: 20:41)



See let us go back here okay if you okay you look at this quantity here P(k) k is P(k) given k -1 – this quantity now this quantity which is here is this always positive definite L see here this is Pe is covariance matrix is positive definite okay so it is inverse is positive definite okay then if you take a matrix okay multiplied by a positive definite matrix and multiplied by transpose of that matrix Lk this whole thing is a positive definite matrix okay so actually you are subtracting a positive definite matrix from a another positive definite matrix okay.

So this matrix should smaller okay this matrix should be smaller than you know P(k)k it should be smaller than because your removing a positive term in some sense okay you are removing some okay your are removing some positive term form the original matrix, so the covariance associated in the update step the covariance associated with you know with the estimate reduces okay.

(Refer Slide Time: 22:16)

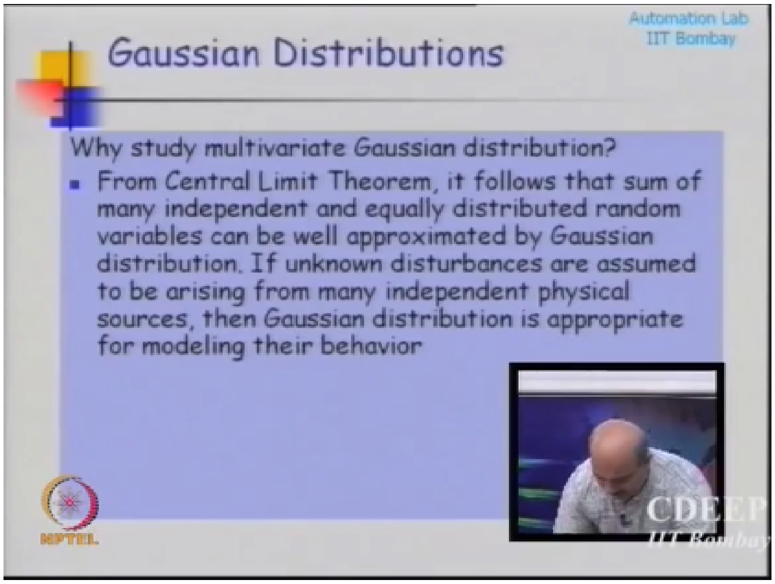


So this is very critical this is very nice thing that happens okay so now actually we have developed a filter which systematically takes into account stochastic description of stochastic deception of the you know unknowns signals we have a way of doing optimal filtering okay there are some questions to be answers okay if this is optimal but is this stable okay stability concept stability idea is still not answered okay but at least we have established that this is the optimal filter okay.

I will show you that this is at least under certain you know simplification you can establish the stability establishing the full stochastic stability is little complex and it can be done but I do not want to do it as a part of this course but I want to give you at least level of house stability and system better is this idea clear that the covariance is reducing okay movement I do update okay covariance reduced so every time actually I do an update more and more information I get the variance associated with the estimate shrinks okay.

So I should get better and better estimate every time I more and more data I collect now till now I never made an attempt to associate any specific distribution to it.

(Refer Slide Time: 23:50)



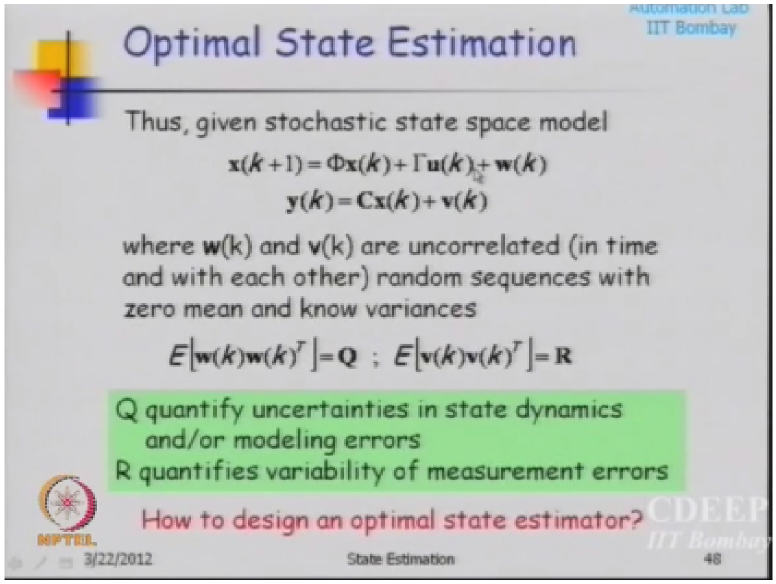
I am going associated now the most wonderful distribution most easy to handle Gaussian distribution and then relook at the same thing through Gaussian distribution view point okay actually Gaussian distribution have some wonderful properties why multivariate Gaussian distributions well multi variant Gaussian distributions can be expressed in analytical form I will show that form what that form is.

Then this is a very important set made I do not know how many of you are first course in statistics but if you have done it you would have you know stumbled up on central limits theorem says that if there are multiple random variables simultaneously influencing some system okay where then companied effect is like a Gaussian random variable okay each there are multiple random variable each one of them have different distributions none of them is individually Gaussian but the combined effect can be shown to be Gaussian okay.

Now if you go back and say that when did you see this statement what is it let me have I am saying that some of many independent and equally distributed some of these words are very important I am just I try to give a very simplified versions in Calabria language but the right way putting it is that many independent and equally distributed random variables can be approximated by Gaussian distribution.

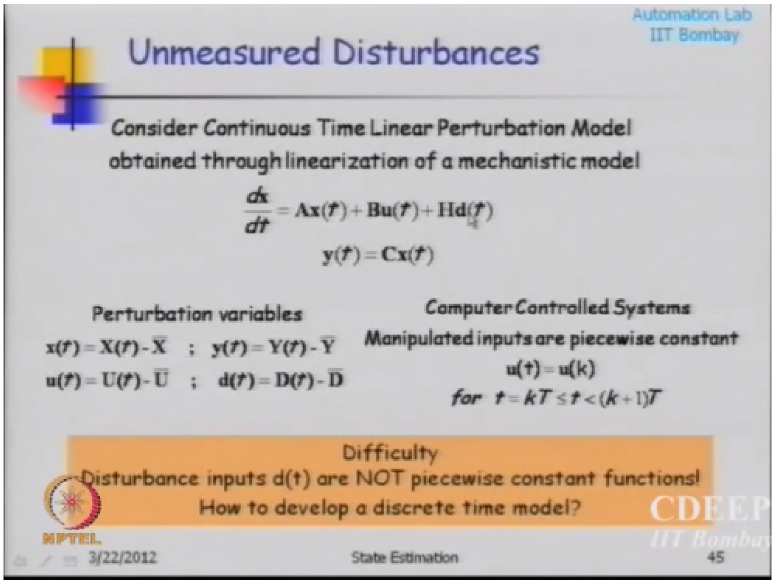
This is a you know very powerful is it so if you know that uncertainty is arising from multiple physical sources it is justified to model it as a Gaussian distribution okay that is why now why am I saying this let us go back here and let us look at our original model.

(Refer Slide Time: 26:04)



Yeah just look at this model okay let us assume see this is a liberalized model right what is u here is the none input okay actually I try to give some rational to w by saying that you know we start from liberalizing you know we start from a linear probation model okay.

(Refer Slide Time: 26:50)

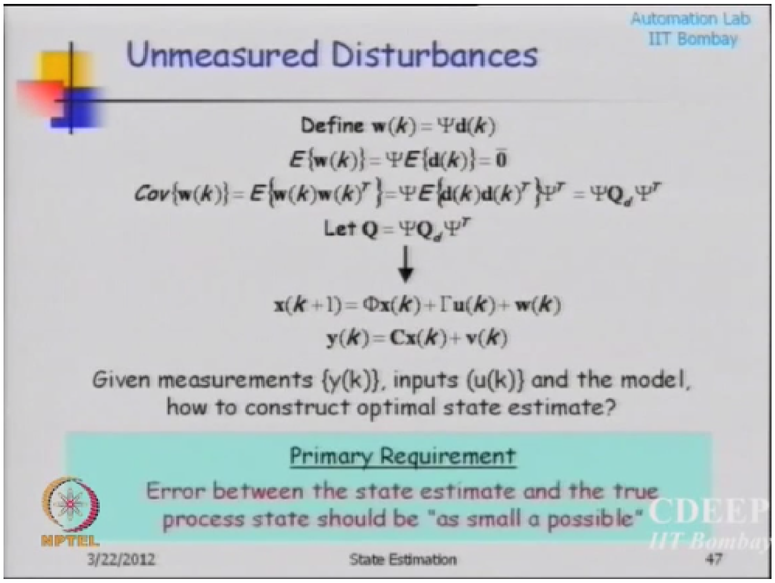


And then you have these disturbance terms okay and then you make an assumption that these are pieces vise constant but just imagine when your developing a model for particular system actually you cannot develop a model with respective all disturbance that are affecting you will develop a model with relevant disturbances okay see let us say I ask you to develop a model for temperature distribution in this room okay.

You will develop a model with you know prominent disturbances for example I would say prominent disturbances people walking in by opening the door okay random disturbance which is somebody walks in with opening the door or walks out with the opening the door but you know there could be some disturbance which is coming because you know there electrical fluctuations which is causing some problem with the you know air distribution.

And you know you cannot model everything that is happening okay and all of these are independent sources of disturbances okay they are combined effect okay actually can be through of as if Gaussian distribution okay so even though I have given here this rational it is not always possible to model all disturbances that are actually influencing the plane okay.

(Refer Slide Time: 28:14)

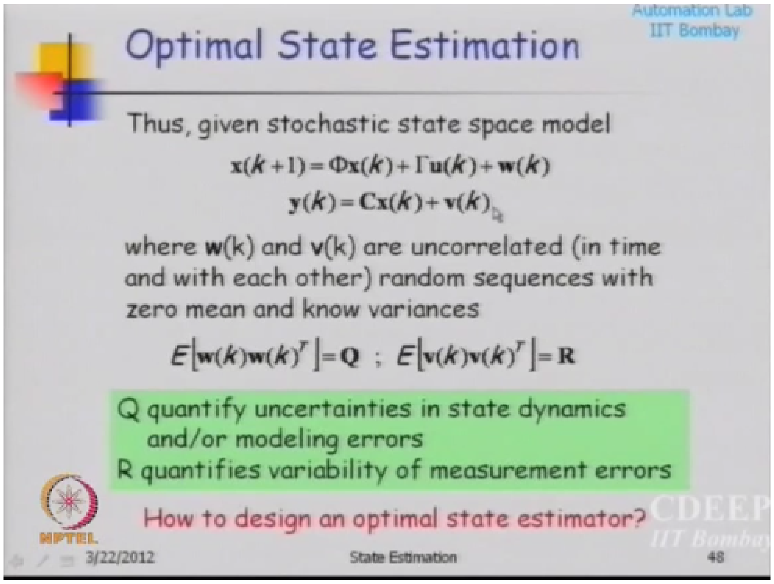


So actually when you come up with this model here so this is all fine you know we discretize and when we come up with all this model here but movement I come to this w(k) model here you can what is the other source of uncertainty when you are actually doing this see you actually started from a non linear differential equation you linearized okay and then you discretized right and then you got this equation.

Now you one thing is that when I got this equation I made an assumption that a disturbance is a piece wise constant but they are piece wise constant so there is a residual disturbance which is molded right so that assume that is an independent another disturbance residual disturbance which is unmolded is like some other disturbance okay there is one more source of uncertainty actually my original system is nonlinear I linearized so there are approximation errors coming from linearization okay.

That is another source of uncertainty in this dynamics right so there are multiple such source you know then there are so many unmeasured disturbances which we cannot model so if you view this w as a raising from multiple independent source of uncertainty then it is justifies to model this w as a Gaussian distributed random variable okay that is what the central limit theorem assures you okay.

(Refer Slide Time: 29:55)

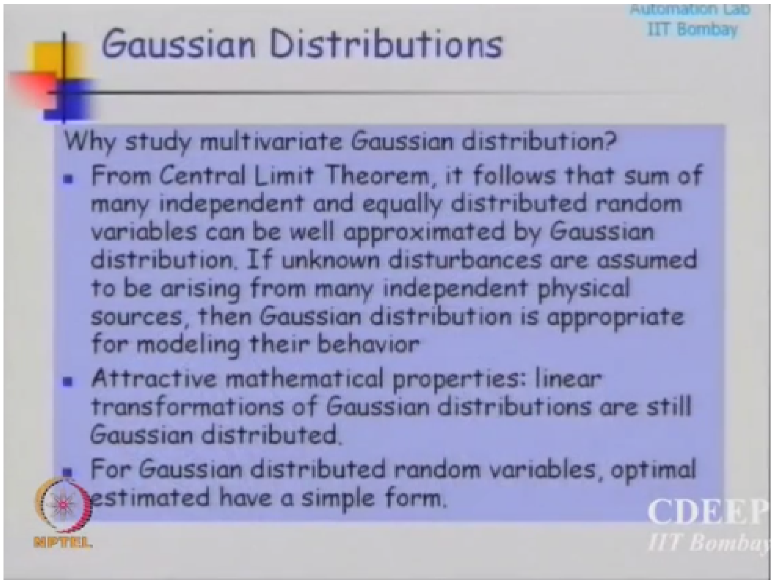


So all that I want to say is that when I develop this model same thing is though out the v(k) when your touching a measurement okay there will be errors committed through multiple sources see you have a single first of all you have signal conversation you know you will convert some temperature into some voltage the you will have amplification circuit right will amplify it okay filter some noise or something and that will add some you know superior signals into the then you have transmission right you have to transmit the data either digitally or through analog say 4 to 20 mille amps or some digital transmission.

It will pick up noise there then if it is analog transmission then you have did a conversion okay in your computer it will pick some noise there and all these are independent sources they are not a combined effect of this can be view as a Gaussian normal distribution quite justified in doing that an actually you will see that if you try to estimate distribution for a measurement noise I have shown you in the earlier lectures right it comes out to be almost Gaussian distribution okay.

So this Gaussian distribution are very important now there is something mold to it than just the central limit theorem they are also very mathematically convent okay they have woudnderful mathematical properties and that is why.

(Refer Slide Time: 31:41)

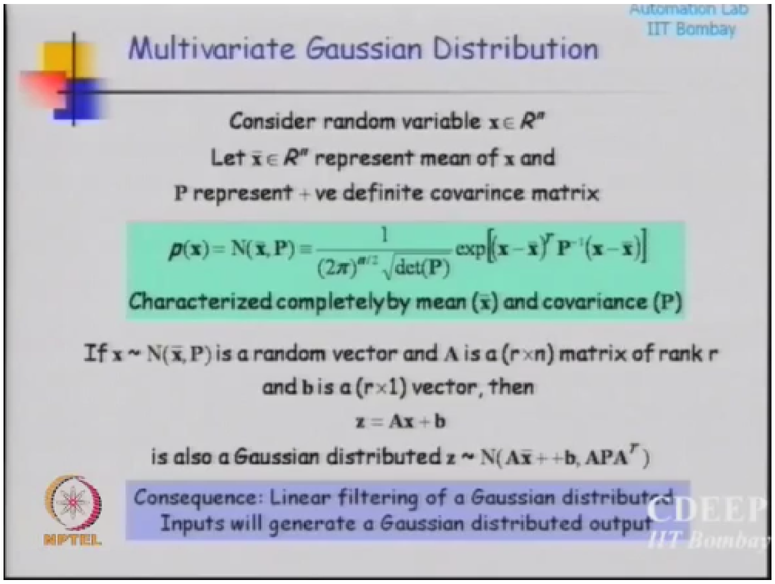


Also by the way we have not missed out anything about Gaussian distributions till now because Gaussian distributions need only 2 things mean and variance and we have formula for doing mean and variance okay so attaching Gaussian interpretation to what we have done is very easy only things required are required to characterize Gaussian distributions okay so this is the first thing that uncertainty is arising from multiple independence sources can be conveniently molded as Gaussian distributions firstly.

Second thing is we have very attractive mathematical properties if you take a Gaussian random variable transform it through a difference equation okay the resultant is also Gaussian random variable that is very nice about Gaussian random variables any linear transformation of a Gaussian random variable is also Gaussian random variable okay this properties you know woudnderful and then it helps us to that is not true about other distributions.

Gaussian distributions have this amazing property that when they are transformed linearly the linear transform Gaussian variable is also Gaussian variable okay and then you know for Gaussian distributions optimal estimates perhaps simple forms so what I am going to hint at is that if you take a view point that W(k) v(k) are Gaussian random process.

(Refer Slide Time: 33:09)



Okay and initial estimation error x0 is also a Gaussian random process all of them are Gaussian not a process initial estimation as a Gaussian distribution and wk and vk have are Gaussian random process okay then xk also as Gaussian distribution xk given k – 1 or xk conditioned on y up to k -1 and xk condition up to y k both of them have Gaussian distribution because you know Gaussian distribution transform through linear equations we will again give you Gaussian distributions.

That is the woudnderful property that is ravel you look at Gaussian distribution what is a Gaussian distribution requires only 2 things mean and variance so if $\bar{x}$ is the mean and P represents a positive definite matrix then probably you remember this complex looking formula so this is exponential of x - $\bar{x}^T$ P inverse x - $\bar{x}$ you can probably make a connection with some of the square of errors x – $\bar{x}$ okay is like an error okay it is weighted by a covariance inverse P inverse okay.

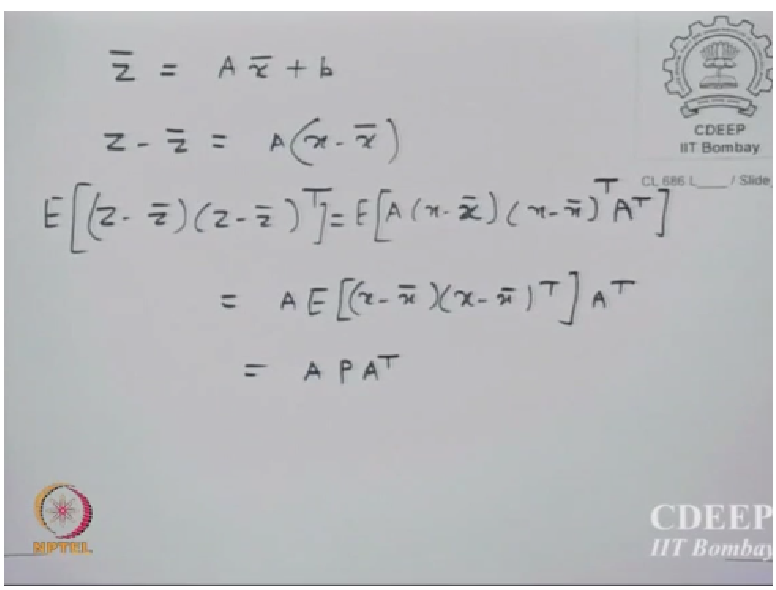
And okay this is an exponential term what do you expect if a if x is away from the mean this quantity will be large or small if x is closed to mean this will be 0 right x – $\bar{x}$ will be 0 okay so exponential of that will be large term okay so actually this is try to say that large errors are less you know smaller errors are more they are closer to the mean errors are and you know this is only a normalizing factor.

This particular factor which comes here it only make sure that what is a property of density function area under the curve should k be one okay so this is this actually this part is only a normalizing factor okay and if you look at this, this is like somehow the square of errors okay if it is like exponential sum of the square of errors what is the waiting function of the sum of the square of errors P inverse.

Covariance inverse is the waiting function okay so the nice things is that if you take x as some random variable Gaussian random variable and then A is some constant matrix and b is a constant vector and you create this new variable z which is Ax + b okay you can show that z also as a Gaussian distribution okay.

It I very easy to prove can you just do this can work this out what will be the mean if $\bar{x}$ is the mean what will be the mean of z it will be Az bar okay what will be the variance z – z bar into z – z bar^T what is it expected value of z – z bar, z bar will be A $\bar{x}$ right + b.

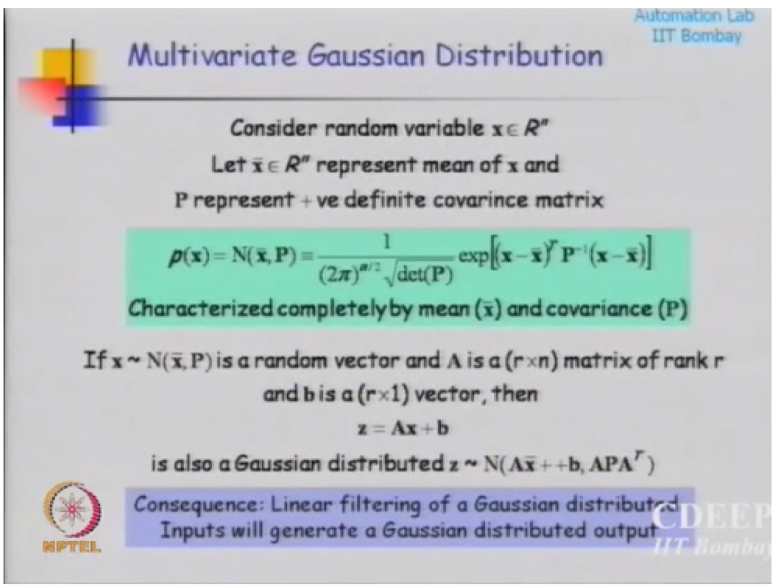
(Refer Slide Time: 37:31)



A x – $\bar{x}$ so z – z bar into z – z bar ^{T} will be A x – $\bar{x}$, x – $\bar{x}^T$ A^T now if I take expectation on both sides okay expectation of this is expectation of the quantity okay which is = A expectation of x –

$\bar{x}$ into $x - \bar{x}^T A^T$ this is APA^T right APA^T okay so this is a very crucial property that if you have a Gaussian random variable.

(Refer Slide Time: 38:32)



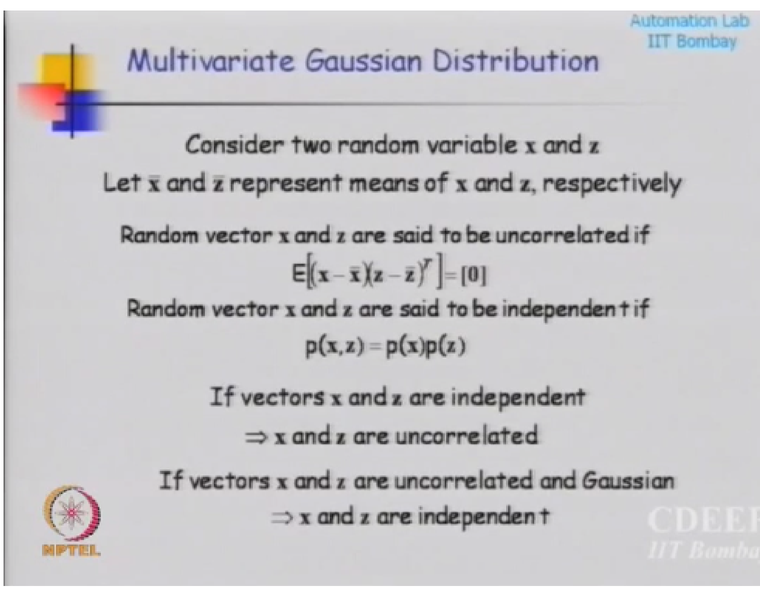
I am talking about vectors okay x is a vector z is a new vector Ax + b, give this mean z this is a new vector if x as Gaussian distribution normal distribution z also as Gaussian normal distribution and from knowing one you can estimate the other okay that is very important yeah Gaussian white noise now wk you can take Gaussian white noise other than Gaussian see white noise does not require know what are the 2 properties of white noise no mean is constant okay and there is no time correlation that is important okay so if you can construct see mathematically it is possible okay.

What we normally use conveniently is yeah Gaussian distribution I think it is possible to construct Gaussian distributions which are multivariate for examples I will give you example of trunked Gaussian there is something called trunked Gaussian distributions okay so it I possible to construct the but the I mean white noise need not always have for example I think you can create a white noise to uniform distributions.

At least one dimensional it is you know should be possible to have white noise so it is uniform distributed yeah no not Ax+ so actually you know the filtering which is happening to difference equation is actually linear flirting we are all having linear difference equations so with an algebra using this particular property this fundamental property okay Ax + b is not filtering what I am saying is that a consequence of this property is that linear filtering of a Gaussian distribution.

So using this property applying it to our difference equations you know this is the fundamental property which can applied and then you can. Yeah oh yes thank you I have missed out the fundamental error thanks okay now sorry about that okay so I am going to this distribution as okay then notion I am going to use this N, N stand for normal distribution $\bar{x}$ and P so only 2 things are required for normal distribution mean and variance okay.

(Refer Slide Time: 42:17)



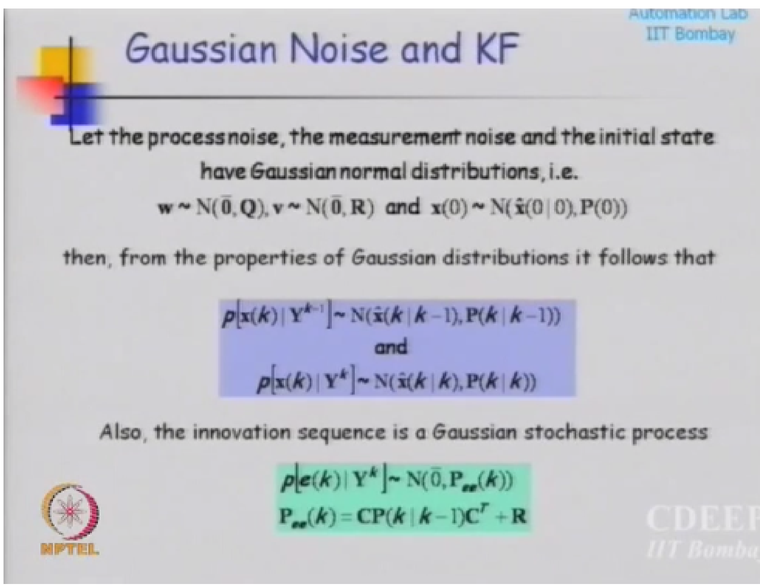
There are some other properties of multi variant random distributions if you take 2 random variables okay which are Gaussian distributed x and z they are called as correlated this is just general definition for nay random variables they are called as uncorrelated if expected value of x

– $\bar{x}$ z- z^T is 0 then they are called uncorrelated I am just giving definition here okay this is nothing to do right now with Gaussian random variables.

And random vectors x and z are called independent if probably density function of x join density function of x and z can be expressed as p(x) p(z) okay now if x and z are independent that means given any 2 random variables if x and z are independent that means they fallow this property okay they are uncorrelated that you can show okay but if they are uncorrelated does not mean they are independent.

With the exception of Gaussian okay in Gaussian random distributions if you 2 vectors x and z and if x and z are uncorrelated they are independent if they are independent they are uncorrelated this true only for Gaussian random variables okay in general it is not true Gaussian random variables have special property independent simples uncorrelated simples independence okay they are 1 and the same properties for Gaussian random variables okay.

(Refer Slide Time: 43:58)



This is just a property which I am stating here all these properties have to be used when you do the analysis ms using this interpretation okay now where am I going to using this I am going to make fundamental assumption that w is a stochastic process is a white noise with 0 mean normally distributed okay normally distributed random variable with covariance Q, v is a 0 mean.

See this is this symbol implies that ve is drawn from okay normal distribution v~ is v is drawn from okay that is how you read this and x_0 is drawn from $x_0$00 and this is the initial covariance initial uncertainty about the estimate that you giving at time 0 okay so all 3 of them are normally distributed multi variant random variables if I make this assumption okay then using the properties of Gaussian random variables you can show that I am not doing this proof here okay.

What I will do is will upload a some you know references and I will send you some references which you can refer to for further studies but I am not going to this proofs what you can show is that this conditional see we wanted to find out conditional mean right if you make this assumption that this is Gaussian w is Gaussian random variable v is Gaussian random variable and initial condition is Gaussian random variable you can show that conditional density of x given yk-1 is nothing but this as a means and this as covariance we know how to compute this covariance and mean okay.

Same thing is true about the conditional density of xk given yk okay so if these 3 are Gaussian these are also Gaussian okay this proof is fairly simple and I mean lot algebra but you can prove this using this properties and I will give you some appropriate literature on this so that you can peruse this so one could actually derive everything assume Gaussian I do not think kalman did this kalman dis the way we have gone about he did not assume Gaussianity but what is woudnderful is that you get the same result if assume Gaussian distribution.

The finally result update rule you know kalman gain calculation nothing changes okay so completely different view point will give you same optimal result okay that is very woudnderful okay you can also show that innovation itself a Gaussian random variable those are few who have lecture can move do not worry so this innovation sequence is also a Gaussian random variable innovation sequence.

And then you it is also 0 mean we already found out mean and variance for innovation sequence you can show that even innovations is a Gaussian random variable okay so with this assumption another fact that linear transformation of Gaussian random variable is of huge rise to Gaussian random variables you can show that conditional density of x is also given yk – 1 is Gaussian conditional density of x given yk is Gaussian conditional density of innovations is Gaussian.

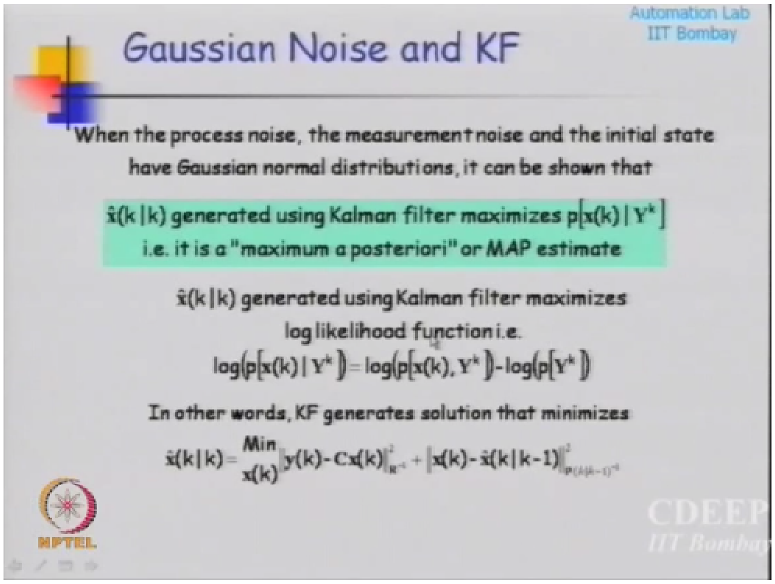
So very nice this any work out for another distribution this works out only for Gaussian linear transformation of Gaussian variables gives rise to Gaussian variables okay one other covariance is you can estimate them okay. Then well I am just leaving some of some more interpretations here you will need to reflect over it before you can get all the meaning see there are 2 things that you should note see actually what we are saying here is that xk is a random process okay and by this assumption where actually able to find the probability density function of xk right.

Conditional probability density function of xk okay so this, see this actually gives you the density functions okay so what is the value that xk take it is a random variable okay you want to give a point estimate actually if you give the distribution xk is characterize is not it sk is completely characterized if you are given the distribution is it so actually it is like saying that this the temperature in this room today okay as a certain distribution okay.

If you take temperature in this room as stochastic process temperature in this as a stochastic process temperature as distribution what value it as taken today is a realization okay that is a realization, but if you characterize the entire distribution you had a tell I mean then you know you have said everything about that random process okay but what happens is we need a point estimate you know we need one value to say what is the best estimate okay so if I give you distribution and say that if this is the random distribution but you will ask me okay but what do you expect it to be today okay.

So that is a point estimate so actually once you have distribution characterizes you have entire information of about x okay yet we need a point estimate so how do you construct a point estimate okay so there are different ways of constructing a point estimate.

(Refer Slide Time: 50:45)



I can say that I want that estimate which maximizes the probability density function conditional probability density function okay which value of x okay maximizes the density function that is the probable value I am will to except that as a most probable value okay that is called as MAP estimate maximum a priority estimate so if I give you lot of information density is lot of information.

You know I cannot interpret I want one number okay so or I want one vector okay so that one vector can be constructed from the density information by multiple ways one of them is this maximum so maximize this density okay so you can show that a estimate that maximizes this density function okay is nothing but kalman filter estimate it turns out that the kalman filter estimate is the MAP estimate.

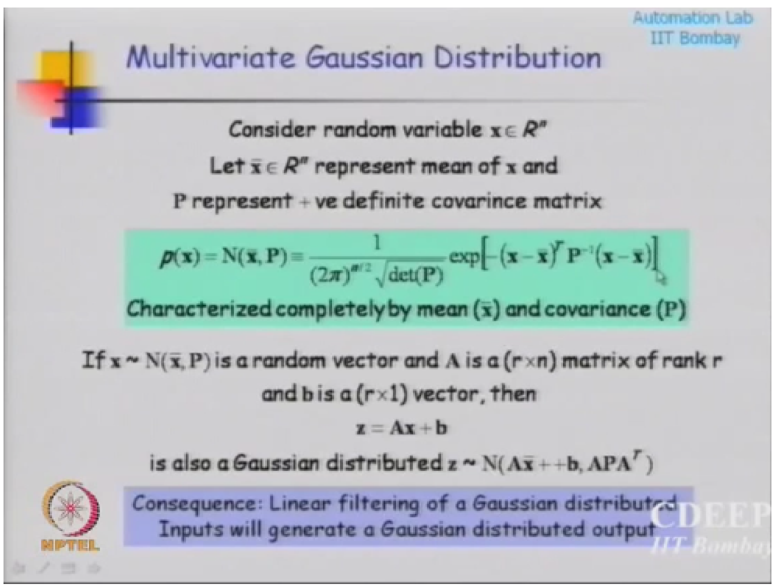
MAP estimate is maximum a posteriori that is the value of x that maximizes this density function is nothing but the kalman filter estimate okay so you can derive kalman filter from this view point okay you start with Gaussian density functions you develop this you know density function then you maximize with respect to estimate you will get the same formula you will get the same update rule.

Everything will be same okay finally but a different view point completely different view will give you same kalman filter okay that is actually beauty of kalman filter multiple viewpoints actually arrive you arrive at the same formula okay there is one more view point is you know maximize this density function other one view point is maximize the likelihood function okay so given the density function that are different ways of you know constructing point estimates.

One of them is just likelihood function this likelihood is log likelihood function actually says that your maximizing this or you are minimizing this some of the square of errors what are the some other some square of errors you are constructing that estimate which minimizes true – predictor estimate okay, this is Nome this is you know waited Nome which this inverse coming in okay this waited norm between yk true – Cxk with R inverse coming in okay why this R inverse and P inverse are coming in.

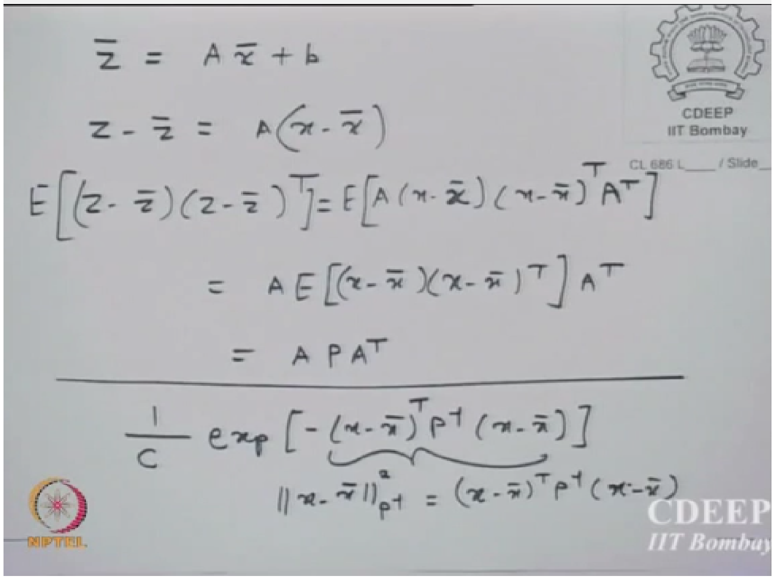
No they are Gaussian these are Gaussian densities so when I take log likelihood okay I will get log of see what is Gaussian density.

(Refer Slide Time: 53:54)



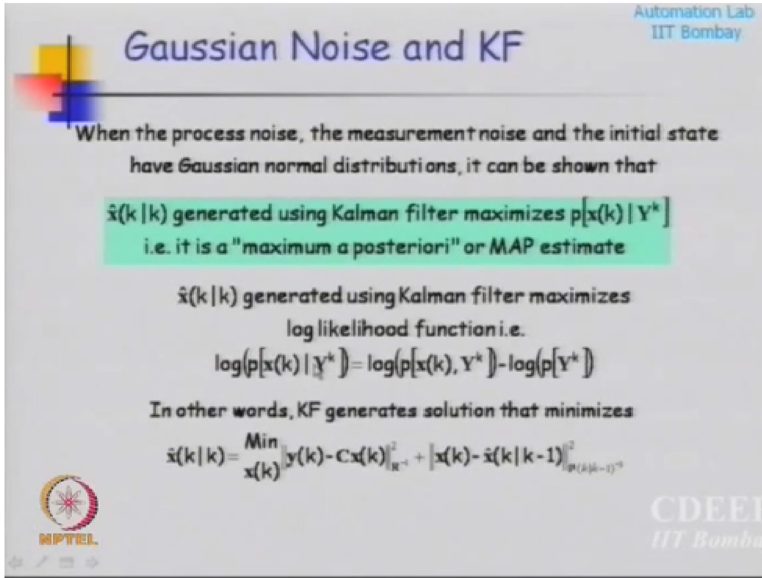
Will be x transpose P inverse right so what is x transpose P inverse x I could right this see Gaussian density I can write as see this is 1/ some constant exponent – x – x̄P inverse x - x̄ right.

(Refer Slide Time: 54: 32)



This quantity I can say that this is norm of x - x̄ sorry transpose comes here yeah so x - x̄ this waited norm this is waited norm square okay so waited norm square is x – x transpose P inverse x - x̄ this is a norm function right you can say that this is a norm function norm square actually not norm so it is like distance square okay it is distance square if distance from the mean value okay that is what appears here in the okay so going back to our slides.

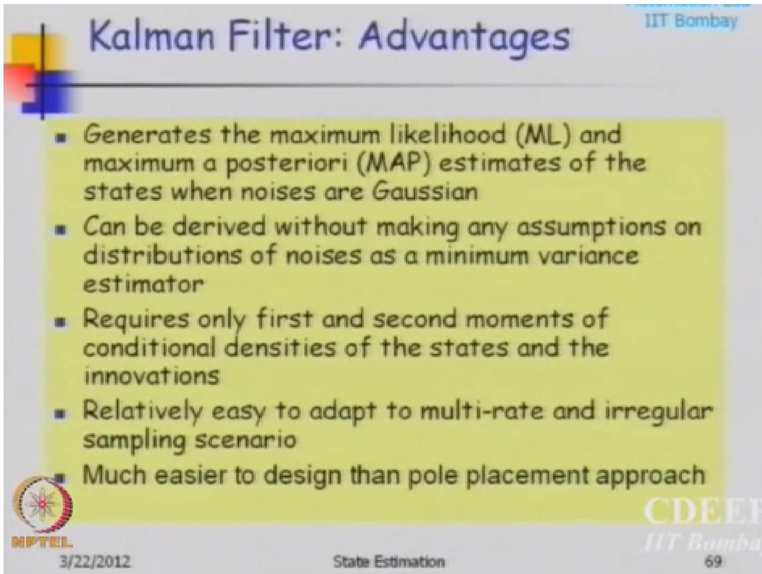
(Refer Slide Time: 55:29)



Okay when you take see Pxk yk or Pxk given yk all these are Gaussian densities and when you take log if that okay you will get this norm which is R inverse or norm because individual densities you know you can use all the property of Gaussian densities of you know uncorrelated and independence all those things will come into picture linear additions of I mean I am skipping all the in-between steps okay.

I will give you those material on that but this can be viewed as you know estimate that gives you maximum likelihood estimate or one that minimizes this objective function so this derivations I will give you those derivations are also equally important so what is the advantage of the kalman filter it generates maximum likelihood estimates it generates.

(Refer Slide Time: 56:31)



Maximum a posterior estimates when the noises are Gaussian okay if you do not want to attach Gaussian distribution interpretation okay even then it gives you minimum variance estimate we derived everything through minimum variances of variance I never used nay Gaussianity there okay so forget about Gaussian distributions we do know what distribution it is but it is yet a minimum variance estimate okay so minimum variance estimation derivation did not require Gaussianity anywhere okay.

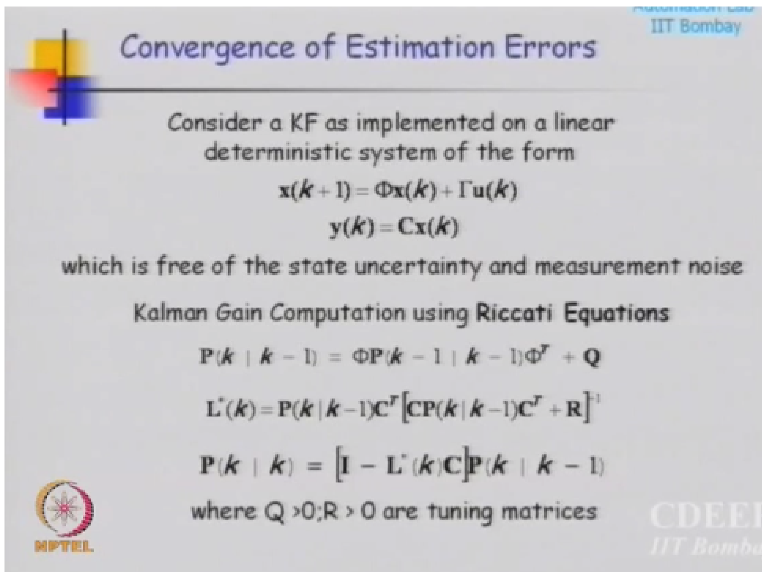
So that is very important so it is a minimum variance estimate it only requires 2 movements 2st movement and 2^{nd} movement okay I am just done with so that is why it is very nice algorithm it is very easy to adopt for irregularly sampled systems are very much there today in which kalman filtering is used you have a data see let us take an example of a mobile phone okay in which the data why is the signal received okay.

If you are have a kalman filter and if your estimating the state, state is the speech which is we want to reconstruct okay and deliver it to the you know person who is listening so if the data is missing okay if I have a model I can do prediction okay I can fill in the missing salable or missing data using x^ given k- 1x^ k + 1 given k – 1 okay when ever data comes I can fuse it and correct the estimate.

So you know you can sue all these ideas for signal reconstruction you know when the data is missing and packed missing you know it happens when you are moving with a mobile phone suddenly the packets miss but yet if you have a sound algorithm which can reconstruct the missing stakes in between okay using kalman filter okay you have to uncertainty molded for measurement and for the sate noise okay.

And you can develop a I mean I am trying to give a simplest view point on this whole thing it is probably much more complex than what I am saying but you know it is possible to do this at least something equivalent can be done in process control I have done this it works so and this is much easier than pole placement approach believe me finding out poles that will give you best rejection of disturbances is very difficult is not so easy okay.

(Refer Slide Time: 59:28)



Another difficult part is now conversations right we talked about the optimal estimate we said that maximum likelihood view point is gives you best estimate and so on okay now comes as a control engineer we are first bothered about the stability then performance right now I talked about a performance first I said this is best estimator we can take care okay so you should ask me what about stability okay.

With it converge okay it might be the best estimate okay but if it does not converge does not make sense how do you prove convergence now this is a little trickily task because this is a time baring system LK is a matrix which others functions of time you cannot use simple Eigen value analysis we have to use this lab stability analysis I am going to take a simplified view point I am going to take I am going to look at kalman filter as a deterministic system I am going to ignore WK and Vk okay.

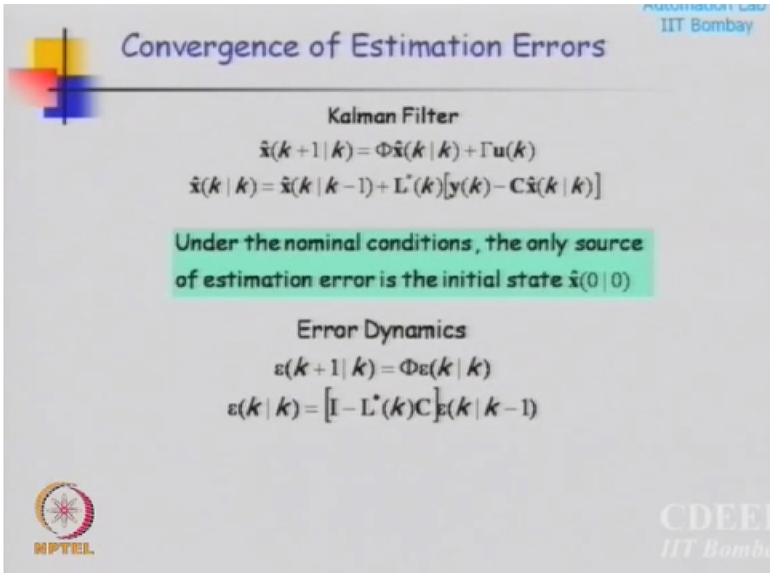
There are 3 sources of uncertainty initial state okay state noise measurement noise are inputs because inputs let us say if they are bounded you know then we do not have too much worry about the bounded output if bounded input will give you bounded output I am first worried about convergence this respect to error in the initial guess okay.

So I am going to consider restricted problem I am going to say that is a deterministic system there is no error in the measurement state there is no noise only possible error is in the initial estimate of the state okay x00 and x0 true x and estimate of x at 0 are different okay will error go to 0 between true and estimate okay I proved it now no expatiation at 0 does not mean it converges so slightly they are different notions we have a class we can okay.

So let us look at KF implemented is a deterministic system okay so let us look at this simplified form okay and so I have naked of measurement noise I have naked of uncertainty in the states okay kalman gain computations are still using those equations recurrent equations you where called as riccati equations so I have this Pk given K – 1 is this ϕ k – 1 k -1 to ϕ transpose + Q then kalman gain is still computed within the same formula okay.

And update is the using the same formula so this equations are still the same and Q and R are some positive definite matrices okay which quantify uncertainties some uncertainty right now no interpretation from the view point of noise some tuning matrices let us take them as a tuning matrices.

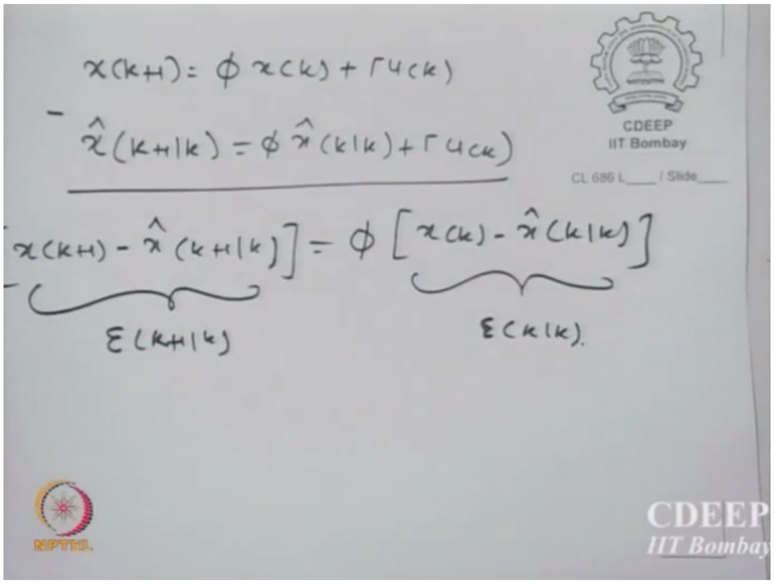
(Refer Slide Time: 01:03:24)



Okay so my questiuion is if I have this kalman filter okay will you know this dynamics of error will it be stable okay see this is my error dynamics can users derive this and you just take this what is the error dynamics this is plant this is the true plant okay and this is the estimator okay what is the relationship of prediction error and what is the relationship for just check whether you get these 2.

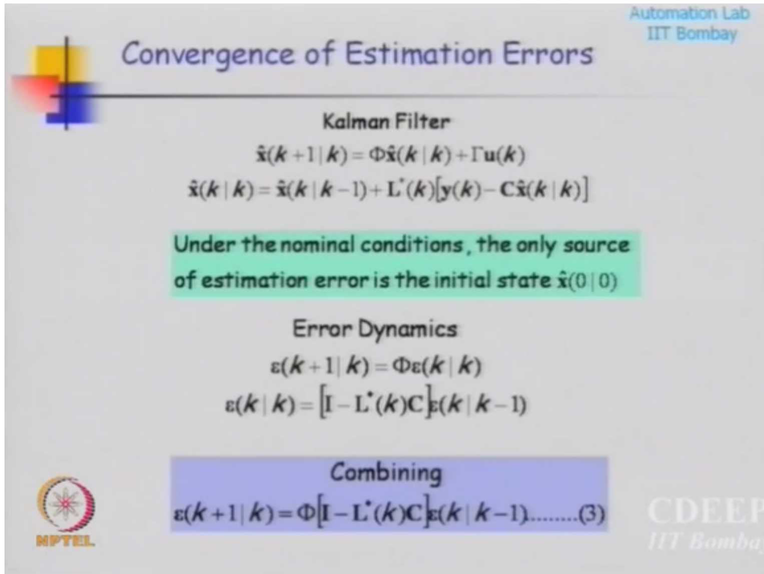
You just subtract plant dynamics from this subtract this equation from the plant dynamics okay you should get this equation UK γ uk γ will cancel okay yiu will get (ϕ) just see if everyone is with me one this.

(Refer Slide Time: 01:04:57)



See I have this equation xk+1=Φ xk+γuk subtract from this x^k+1 k=Φx^k1l+γuk +subtract I take x k+1 –x^k+1given k=Φ*xk-x^k1 okay so this is my ε k+1 given k and this is my εk given k okay.

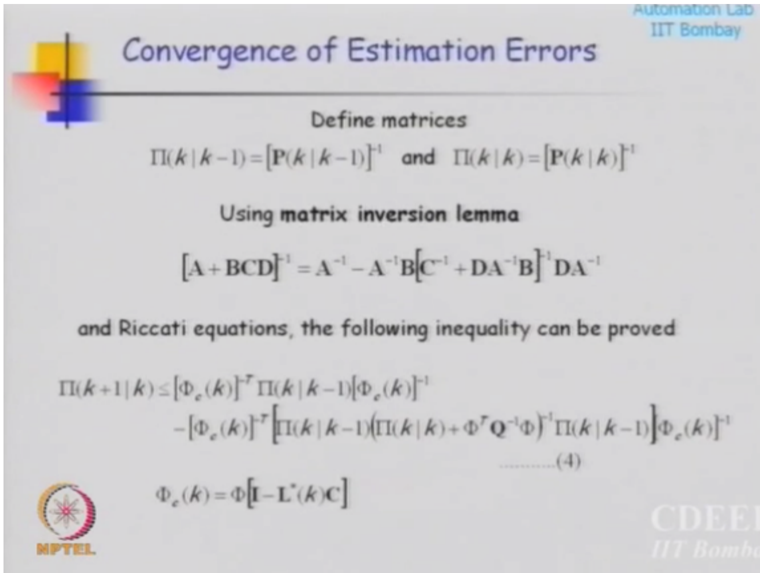
(Refer Slide Time: 01:06:01)



So that is what I have done okay second equation follows from the update equations it is not difficult to derive okay so these are my these two coupled equation define the error dynamics okay I will eliminate here εkk I will combine this into this equation these two can be combined into eliminate one of them I am going to work with this prediction error dynamics okay.

If I show that prediction error dynamics is stable then this also stable because these two are written by constant but through matrix this is only linear transformation of this is only linear transformation of this so if this goes to 0 this is also go to 0 that is not possible okay so one of them I have to show that if I show this dynamics is stable now I cannot apply normal analysis of values here why? Because this matrix is time varying okay so I have to use something else okay.

(Refer Slide Time: 01:07:16)



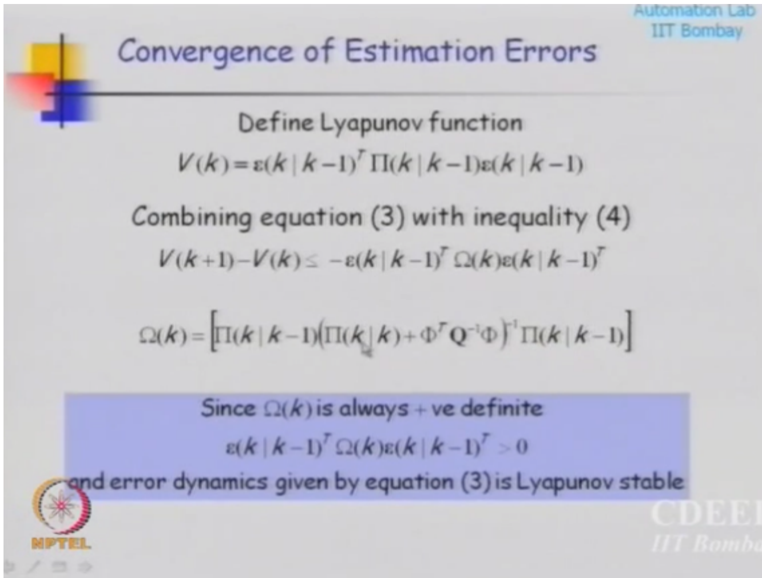
What is a linear algebra okay so first of all I am defining two matrices here which are πk k-1 which is inverse of this matrix okay and π k which is inverse of pk given k okay that is the first thing I am defining this inverse second thing is I m skipping huge amount of algebra okay I am just giving you one result okay this particular result will be very famous result in linear algebra this is called as matrix inversion lemma okay very, very rapidly used in derivations of Kalman filter okay.

A matrix inverse of lemma talks about A+BCD inverse and if this can be proved very easily just do multiplication cross multiplication okay and you can shot ht left hand side equal to right hand side it is not difficult to do this okay you just have to be patience multiply all you know matrices properly you will get i-I okay this result will not difficult to but this result is very useful.

And I want to use this result to prove a very complex inequality okay what is this term here this is inverse of P k+1 given k okay I want to use this term okay I am showing that this term okay this smaller than you know this quantity here plus another quantity now this is the positive difference matrix this is another positive difference matrix and this is positive difference matrix because why because pk is poly definite matrix so inverse of pk is also poly definite matrix okay.

So this inequality with lot of algebra okay you do not trust me with this I can put up those derivations and you can have look t those the end result where I want to come to is more important so there is some complex inequality which you have proved using this okay.

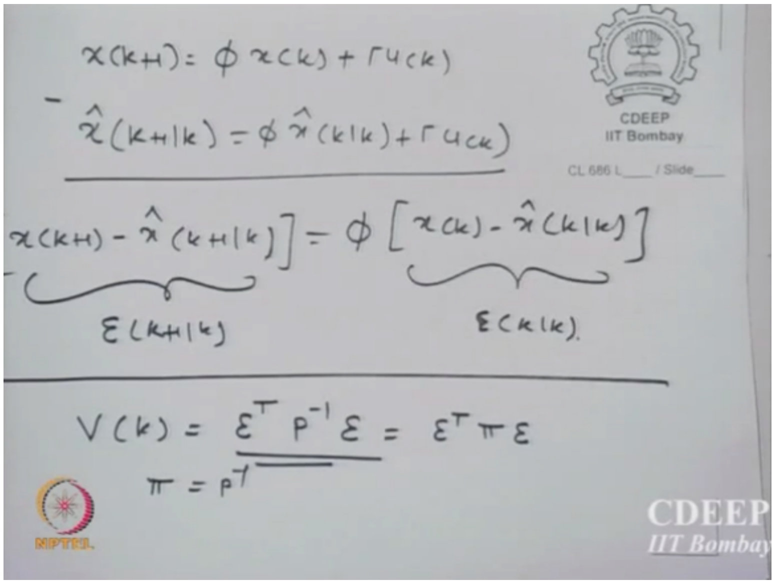
(Refer Slide Time: 01:09:57)



So how it going to helps me I want to use to define a Lyapunov function okay and I want to use Lyapunov stability analysis to true the convergence okay. Let us go back to Lyapunov analysis what is the Lyapunov analysis Lyapunov function is poly defined function and it is gradient should be negative infinite okay so I want to show that vk+1 –vk okay is negative or it is negative infinite okay so as time progresses okay you know vk reduces okay as time progresses Lyapunov function reduces.

The second constructive Lyapunov function okay where I can how this okay then I establishes the stability of the okay I am going to define Lyapunov function using see.

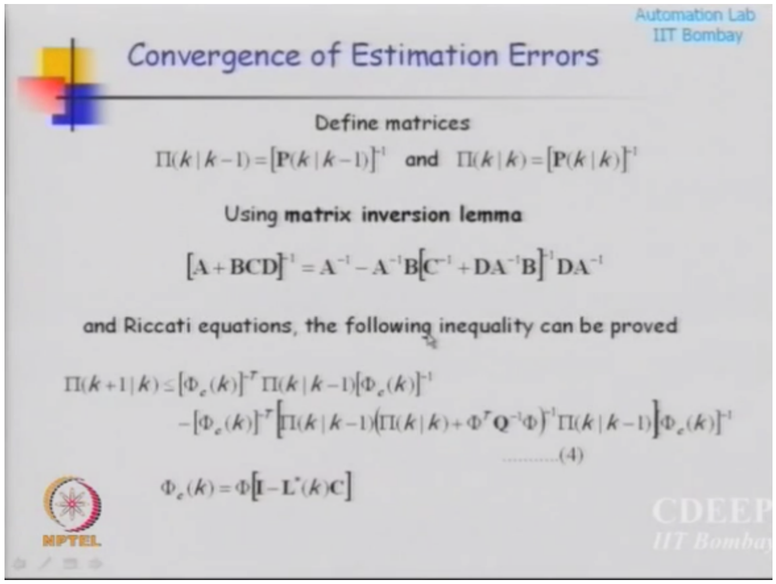
(Refer Slide Time: 01:11:10)



I am going to define a Lyapunov function vk is ε transpose P inverse ε what is P? P is the co variance associated with the ε okay but we have said that π=P inverse okay so this is nothing but ε transpose π ε okay I have just removed all K , K given k-1 mand all that this is what I am doing I am defining the Lyapunov function actually this is the term which appears in the Gaussian distribution okay.

I am using the same term to define Lyapunov function okay so I am defining Lyapunov function like this I am moving on here I have defined this Lyapunov function vk which is error at k given k-1 transpose π is the P inverse into error again at k given k-1 what you can show is that vk+1-vk is a negative function this difference is always a negative function. Because this can be shown to be equal to this term in the right hand side this term here appears to be nothing but this complex matrix and how could I prove this.

(Refer Slide Time: 01:12:34)



I could prove this because of this inequality this inequality follows from matrix inversion lemma I am not giving you a proof I am giving you a sketch of the proof okra I am giving you a sketch of the proof I can do Lyapunov stability arguments I can construct a Lyapunov function okay through which I can show that Kalman filter gives me a stable filter if I view it or deterministic filter.

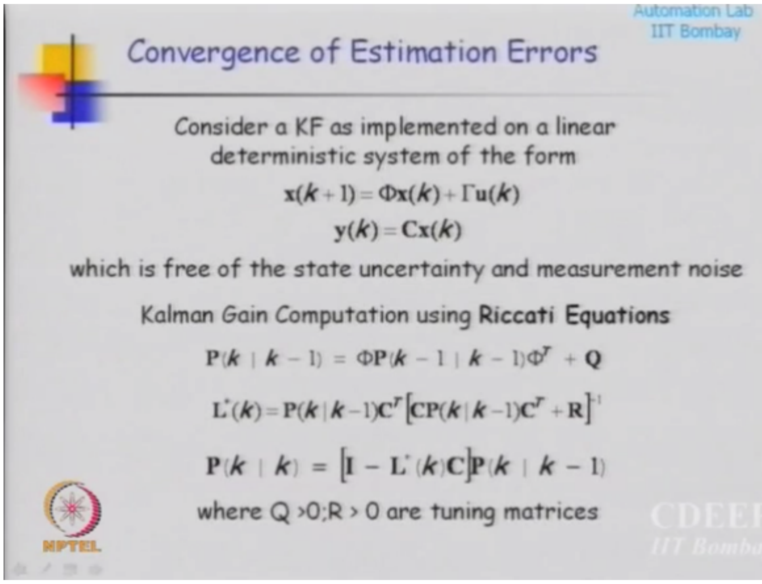
I only take initial estimation error into account I forget about measurements noise I forget about the state noise okay I have consider Lyapunov function which is negative infinite okay vk+1 –vk okay is always a negative value why because this ω is poly definite matrix why ω is poly definite matrix you have to go back and argue this is see what is this function of what is πk? πK is P inverse p is the poly definite matrix.

So P inverse is poly definite matrix okay so this is positive definite matrix okay what is Φ Q inverse Φ is this poly definite matrix you saw the poly definite matrix some of the positive definite matrix are poly definite matrix is the positive infinite matrix okay so I can show this vk+1-vk is nothing but this term which is always negative because this positive definite matrix so this term will always positive definite for any ε okay so – of this will always negative so this vk always reduces this Lyapunov function will always reduces so this is a Lyapunov stable.

So I have proved stability of Kalman filter okay at least in a restricted sense okay I have denote the state noise measurement noise but I have proved at least for the initial condition the error will go to 0 okay at least the sketch of the proof is yeah, yeah yes, yes no, no. So I am saying that,

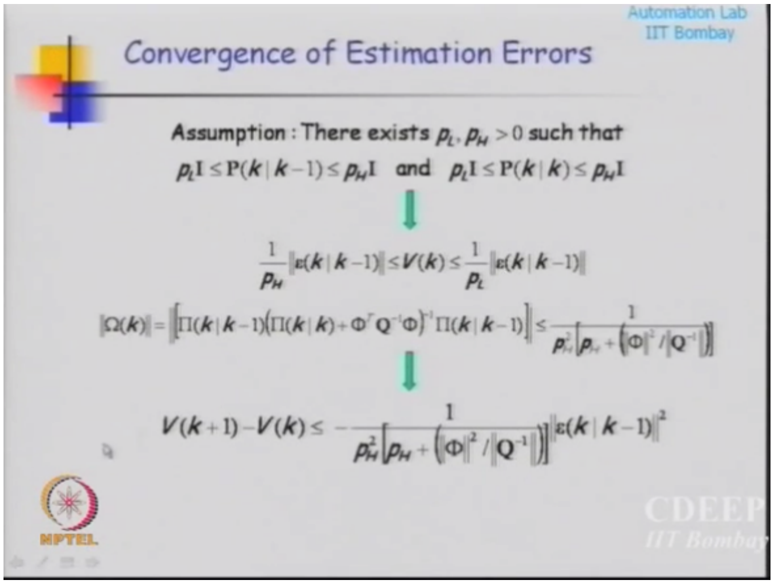
that is the formula P and Q are some matrices okay and there is the formula for computing Kalman gain what is that formula this is the formula okay.

(Refer Slide Time: 01:15:20)



This is the formula for computing the Kalman gain so P and Q are some positive definite matrix which re some tuning matrices now do not associate view point of co variance okay so P and Q are some tuning matrices well I m trying to compress things that are understood over several years those lectures and you also just think that this is the sensitizing if you pursue in this line you will understand this much more deeper much more what I have understood okay.

(Refer Slide Time: 01:16:12)



So this is the Lyapunov function and then if you make some more assumptions okay you can show that there is not just stable observer by it is asymmetrically stable of the and you can even show that this is exponentially stable observer so I have just outline that here that if you show if you can say that this co variance is bounded between some lower limit and upper limit then you can show that this vk is also bounded between lower limit and upper limit.

And you can prove you know asymmetrically convergence of error will not only the error dynamics Is not only stable but you can show that asymmetrically stable okay so that we can prove so this particular quantity you can show is always positive quantity and rate of error will asymmetrically okay so I have just put here the conditions for asymmetrically stable I will show you that if you make one additional assumption that co variance are bounded okay they do not blow up.

Then you can prove stability of I just said that it is stable in the sense of Lyapunov I just said that is stable in the sense of here strictly negative definite yeah, but I m saying here is less than or equal to 0 so it can admit actually positive case probably yeah, but if you say if you admit this yeah but equality will no, no but what if this ω is rank definite and ε is line in the space of yeah then if it is rank definite positive definite but rank definite okay.

I mean I am just thinking a mathematical possibility of okay so for strict asymmetrically stability you have to show this that vk is rounded okay and then it is not sufficient to show what I have shown earlier if you can establish this then you can show asymmetrically convergence so let me stop here it is already quite complex but what I want to do next is to connect these Kalman filter with time series models that is my last class in the series of the lectures.

Once I have done that you know I close the book so we started with data driven models and it come to Kalman filter and I will now show that actually Kalman filter have a connection the data driven models and how they merge where they merge where is the merging function okay so will come to that in the next lecture and then after that I done with that I want to start control okay I am done with so the next part of assignment is computing some implement a Kalman filter okay.

NPTEL
NATIONAL PROGRAMME ON
TECHNOLOGY ENHANCED LERNING

NPTEL
Principal Investigator
IIT Bombay
Prof. R. K. Shevgaonkar
Prof. A. N. Chandorkar

Producer
Arun Kalwankar

Project Manager
M. Sangeeta Shrivastava

Sr. Cameraman
Tarun Negi

Sr. Online Cameraman
Sandeep Jadhav

Digital Video Editor
Tushar Deshpande

Technical Assistants
Vijay Kedare
Ravi Paswan