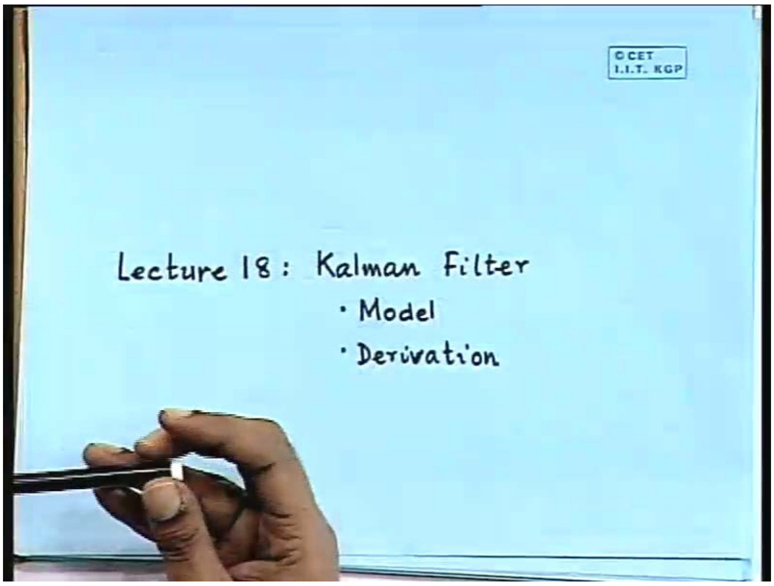
Estimation of Signals and Systems
Prof. S. Mukhopadhyay
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 18
Kalman Filter – Model and Derivation

So as we mentioned in the last class that, in the last class we talked about the observer problem, remember.

(Refer Slide Time: 00:55)

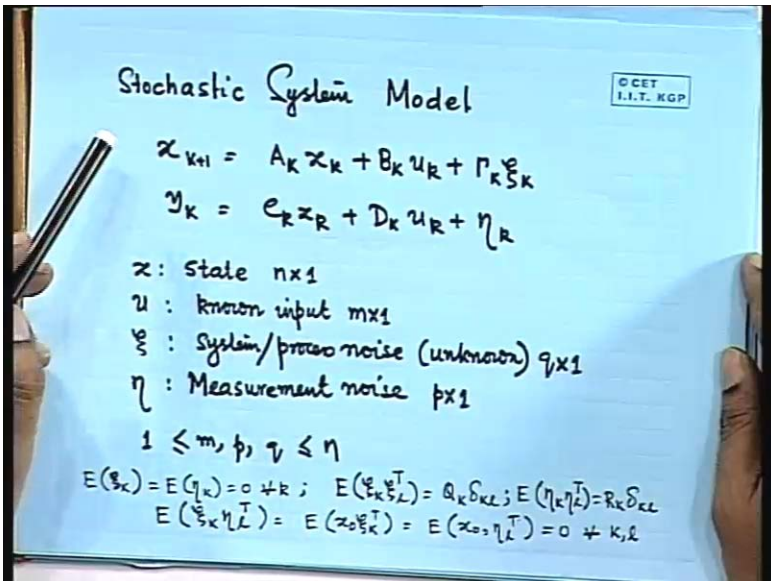


And that was the simple problem; in the sense that, no noise I mean lot of practical things were not mentioned, at least I mean they were not even consider to exist. And at the end we said that, we have to if we want to take care of these things then we have to use the Kalman filter in practise actually. So so today we are going to look at the, start looking at the Kalman filter. So what we will do is, we will first describe the model and then we will start the derivation.

I do not think we will be able to complete it today. There are many derivations of Kalman filters from various points. In facts, I mean Kalman wrote the paper in nineteen sixty by the way on which describe the Kalman filter, and after that lot of work has been done;

especially by Thomas Koilath is one of the I mean scientist, he happens to be an Indian born also visits Bangalore used to visit Bangalore almost every winter, incidentally Kalman is very much alive is so any way that apart. So let us start talking about the model. So now this is the model, okay. This is the stochastic, actually it is mixed stochastic and deterministic. So compare to the previous model, slight change in notation the now I am you know you follow certain text, so some I mean unless you follow its notation it becomes difficult.

(Refer Slide Time: 02:34)



So this x k plus1 previously probably we were writing in brackets; so this x k plus 1 is the state at now, we are talking about discrete time models. We are not talking about; there is a there is a continuous counterpart of the Kalman filter, that is called the Kalman-Bucy filter, we are not talking about it, okay. So we are talking about the discrete time Kalman filter, which assumes a model in this form, okay. So there are certain things to note here; firstly there I am assuming that, the state matrix changes with k. So it is a time varying system, this is the general formulation on the Kalman filter; you can assume that all A k is equal to A n. You can also assume a time in variant case, but the Kalman filter formulation is done for a time varying case.

So that is why I have put this, k is on the even the system matrices. So they are supposed to vary with time, okay. In a in a predetermined manner, it is not non-linear, it is just a function of time, okay. Now apart from A k, x k plus B k, u k which were there in the observer also.

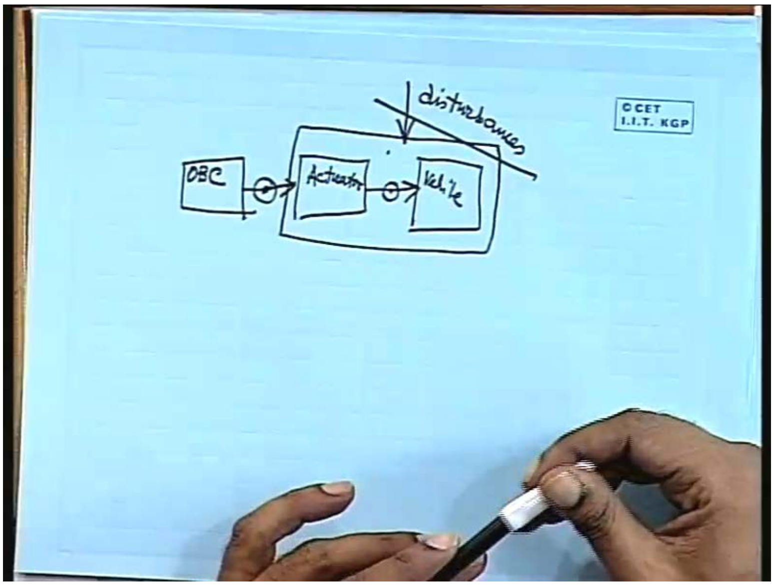
Now we have a we have a third term which is gamma k, this is I think xi or zeta, okay. If this I always get confuse between xi and zeta, probably its zeta okay. So so I will call it zeta, is it is it zeta or xi, I think this is xi, xi.

Conversation between Student and Professor – Not audible ((00:04:17 min))]

Student<I think it is xi>

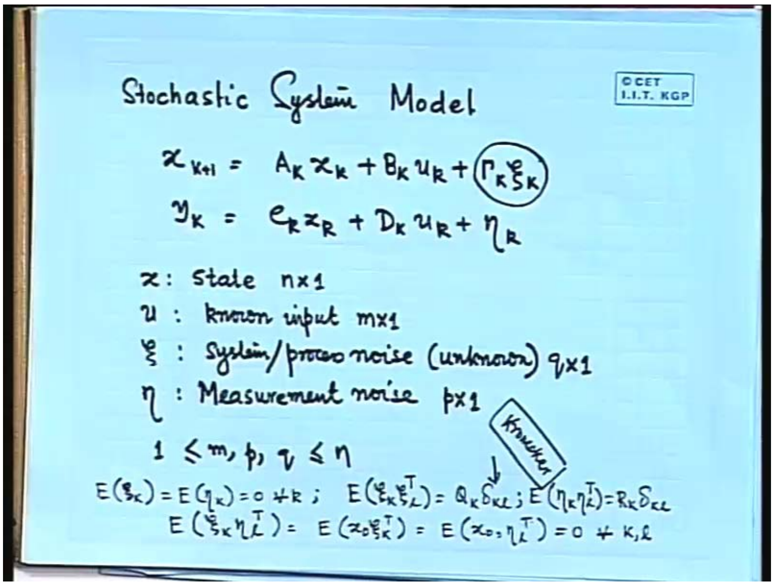
So so any way we will we will we will remove that confusion, I I will check up in the dictionary. So this let’s say is gamma k, xi k. So here you have this is this is the term which is unknown. So we said that inputs, there are two sources of this of this uncertainty; number one is that, see actually what is happening is that, typically what is happening

(Refer Slide Time: 04:46)



That you have some let us say; some control computer which goes to the actuator which goes to the plant, okay. So this is the control processor, sometimes in in aerospace technology this is called OBC on board computer, which contains the autopilot algorithm et cetera. So that gives the command to the actuator and that gives the command to the plant, okay. So this is the say vehicle. So sometimes if if you are using the input here as u then, then then that may not be the same as what is going here number one. Even otherwise, even if you consider this to be the plant you might consider that, even then there there will be other inputs called disturbances; right, which will be unknown. So there are uncertain inputs, the just to capture this this term is included this xi is unknown.

(Refer Slide Time: 05:50)



Similarly the output is normal C k, x k plus D k, u k plus there is a measurement error. So this is the measurement error term, right all sensors will have measurement errors. Now, so these are the dimensions x is n into one, u is m into one, xi is q into one, assumed measurement noise is p into one, which means measurement is also p into one. Once you put these, then you have corresponding dimensions of the matrices, right. So this is n into n, this is n into m and all that. Now the question is that, unless you do some unless, if you if you if you assume it to be totally unknown, then you cannot do anything, you cannot get a result. You have to assume some some things, some bound and values, something; otherwise you cannot obtain any results. So therefore, we have to necessarily we have to assume something about this. So what we assume is in a stochastic setting, there are there are there are other algorithms; for example, there are h infinity filtering algorithms which will not, which will which will assume bounds on this, absolute bounds value of xi should not go beyond this, that kind of assumptions.

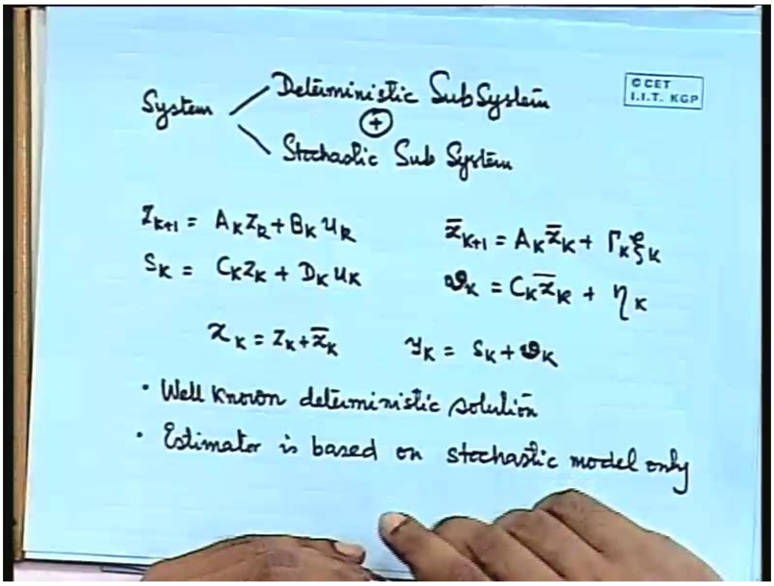
But here we are using a probabilistic description. So our assumptions are as follows; that these two are both zero mean, so this equal to this equal to zero for all k, can you read this? Can you read these lines on the television? Okay. Second thing is says is that, the correlation in this case correlation and covariance are same because the mean is zero. So you see that they are mutually uncorrelated. So if k and l this this delta is sometimes called the Kronecker delta, there are two delta functions; the continuous time delta function is called the Dirac delta, and the discrete time delta function is called the Kronecker delta. This delta means that, when k is equal to l its value is one, when k is not equal to l its value zero; which means that if these k and l are different, then then they are then their correlation is zero. So the successive samples of noise, the same noise are uncorrelated, its says that. While obviously; if if you take xi k, xi k transpose, then you will get a matrix Qk which is assumed to be it necessarily has to be positive definite, okay. Why? Naturally, because x because, if you x transpose this into x; it will be it will be non-zero, l be greater than zero not only non-zero.

Similarly here, so this is the this is called the process noise covariance; this is called the process, system noise or process noise, and this is called the measurement noise. So this is the process noise covariance and this is the this is the measurement noise covariance at time k. Again we are assuming that, their properties may also vary with time, right. But the process noise and the measurement noise are completely uncorrelated, they are identically equal to zero even xi k, eta k transpose is zero. So even if k is equal to l it is zero, they never correlated, okay. So similarly the initial state is is also considered to be a random variable and its correlation are are also assume to be zero.

These these you have to assume; to you know unless you assume these things, then your then your analysis will become too complicated. So you you you cannot proceed, you cannot get a clean result, okay. And in many cases since these sources of noise et cetera are totally different, so they are quite likely to be uncorrelated also. So at this point of time, remember that I am not making any assumption that their that their Gaussian or anything, you know sometimes such assumptions are made in in some proofs of the Kalman filter. So but I am not making any such assumption; but this is what is necessary to assume, okay. So nothing much signals, I mean the disturbance is a generally assumed to be zero mean, right.

There are there there are three uncertain things; One is xi, another is eta, another is x naught is the initial state; which is unknown, which was unknown even in the case of the observer. And what we are assuming is that, the xi and eta are are zero mean, xi is uncorrelated with itself and they are definitely, they are actually independent of each other. so this is the assumption. So there you know, there are like clean signals and white also, because they are uncorrelated, all right. So so so this is my model, now naturally you see that; the it it turns out that we can break, this this has one deterministic component and one stochastic component. So it turns out that, we can really separate out this system, use this system as this system as a as a sum of two systems; one is the one is the completely deterministic one, another is the completely stochastic one, right. And the completely deterministic one, solution is the is known normal linear system solutions. So there so there is no estimation there. So the so the estimation problem needs to consider only the completely stochastic system which means that, I our Kalman filtering problem we can ignore the input u, so it will get so terms will get simpler that is all.

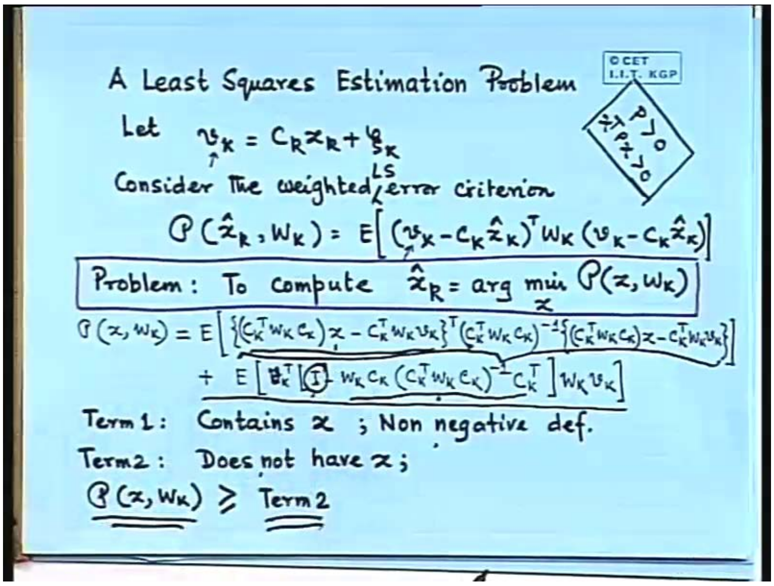
(Refer Slide Time: 12:09)



So I have I can look at it in this way. So if you write that that the state can be seen; that is now we are performing decomposition. So this system state is is suppose to be a sum of two states. One of them is deterministic; this is the deterministic system where there is no stochastic input, everything is known okay. Similarly there is another system which is which is totally stochastic, right. This, it is totally driven by stochastic signals; there is no deterministic signal here, and and and if you take the state of this and the state of this and then add, then you get the total state. So you can easily you can just add these two and you will get the old equation back, correct.

So then we when we are talking about the Kalman filter, we only need to consider this. This we can ignore because, this is a deterministic solution, no no estimation needed. So this this part has the well-known deterministic solution and the and the estimator needs to be based on this one, okay. So we can ignore u for the time being.

(Refer slide Time: 13:20)



Step by step, so we have got our, now before we solve the Kalman filtering problem; I actually I have you know, we have chosen a proof which is from a from a particular text by chuan chen. And it is it is unique way of proving and I mean I personally found it very very easy and simple to to follow, based on I mean common sense optimisation approach. There are there are many other proofs, we can see some of them, for example there is an innovation based proof; which will require you know concepts of gram-Schmidt orthogonalization of signals, et cetera. There are there are some additional concepts, we will have to be dealt with. In here, there are certain assumptions required which are which are again undesirable, that is correct but but let us make those assumptions and, first see why the Kalman filter is is optimal.

Then if time permits, we can look at the Kalman filter from from from from various ways. We can we can look at it as a as a Bayesian estimator, which which we which seen to be as a as a conditional mean or it can it can be seen as an as an in orthogonalizing filter which actually makes the innovations white. So there are there are there could be various interpretations, from which the Kalman filter can be derived under various sets of assumptions. We are looking at one of the approaches of of of proving the Kalman filter, which is very simple and does not require any any extra knowledge, requires some assumptions; but we will we will be a I mean, you will you will easily be able to appreciate that, how it is optimal without any extra knowledge.

So from that point of view, I have chosen this proof among many others which exist in the literature. So now before proving the Kalman filter, we are first solving a simple problem.

You know the Kalman filter actually; the the the problem I mean the Kalman filter is supposed to supposed to give you a state estimate which will optimize which will first of all take care of all past measurements and we will minimize the variance of all those measurements with that estimate.

So first of all let us see that, if we are given one value of measurement v at any instant v k, how to choose a state? That is what how to calculate an estimate x k, such that, the variance of that measurement is minimized and something else is minimized. So first let us try to solve the problem for a single measurement case, then we will try to solve the problem for many measurements and that will give give us the Kalman filter. So the first problem is much simpler; the first problem says that, suppose this is your equation. So here you have a measurement, this is this here you have a state, so C k, x k plus I have used the term xi k here for some reason, because it will come later on.

It in your earlier equation, you will find this as eta k but in this case, we have used xi, k okay. So so so suppose this is the true system equation, this is the true state and the measurements are being generated in this form, this is a random process, okay. Now now what is my, what is my objective? First of all I want to find an estimate, I mean the moment I find, what to find an estimate. I mean that estimate will have to will have to optimize some performance criteria, so have to define some performance criterion; all estimates or all optimal estimates will be optimal with respect to some performance criterion. So what is my performance criterion? So my performance criterion is this, I have defined it like this. So you see what is this? This is at this is the weighted error, v k minus C k, x hat k. Suppose x hat k is my estimate, then what is my error, v k minus C k, x hat k? So I am calculating a variance of error but it is a weighted variance in between, I have putted W k, some W k I am choosing.

So my so I have chosen a a a weighted estimation error with some W k, which is a priori chosen, for the time being let us assume to be fixed.

We will also see, how to choose W k in a second step? So we are going step by step; first we want to minimize this, we have we want to choose such an x hat k which will minimize this for some choice of W k, okay. So the problem is to is to compute x hat k, this a r g min means, you have to find out the minimizing argument of this function. So this function, so you have to find the minimizing argument x of this function. In other words, you have to choose x here; such that this becomes minimum. This is a you know, we have writing a r g min.

So now, so now what we do? So we first thing we do is that, we we see see what we have done here. What have we done here? Look at this, if you multiply this is see; I have from it will define like these, now suddenly I am writing this, how? So what I have done is. Now if you multiply this, C k transpose, C k transpose W k, v k transposed, okay, this correct. What is p x, W k is v k, v k if I put here, it will be let us multiply this to this, then I will get x, just one moment. Let me see actually, what I am trying to do is .See what I am trying to do is, I have to minimize this sum, I have to choose such an x hat, such that this sum is minimized. So so what is the trick? Trick is that, you you separate it out into into two terms okay; one of them will contain this x, the other one will not contain this x. See the see that the second term does not contain x. So so there is no optimization involved here, you you are supposed to choose x. So this term is independent of x and I have written this term in such a manner, that this term is always greater than zero. So it is supposed to be positive definite thing, why?

Because this is positive definite W k is W k is chosen as positive definite. So, therefore this is this is a positive definite matrix and this is some some some matrix transpose, x transpose this x. So therefore, this is also positive definite. So this is a function, this this first term can never less than zero, actually can never be less than zero. Now there is a there is a question, which I am expecting here. It is not coming, is that firstly is these is this a scalar, that I am talking that it will be the that it will be less than zero greater than zero is this term a scalar? It is not a scalar; because this is not a scalar this has dimension p, so then this is not a this is a p by p matrix. So what do what do I mean by it it it be greater than zero?

So whenever I say that are whenever I say that a matrix is greater than zero, I mean that it is positive definite which means that; if you choose any x, then x transpose p x which is a scalar will be greater than zero. So whenever is say a matrix is greater than zero, I mean this okay. So so this is in that sense, right. So with so this means that this matrix is it is a positive definite matrix. Now we are we are we were confused about, we can we can just multiply this straight away, okay. If you multiply it, we can do it term by term this will come. So let us not do it I mean right now, if we get time we will we will do it.

This will come you have you have to just multiply this term by term this will become x. This will become let me see, some of these terms this is v k transpose I, v k transpose there is a term v k transpose w k, v k okay okay okay this has been made into a whole square that that is what, here what is the term which is not involving v k, not involving x? There is only one term that is v k transpose w k, v k. Now here that has become so long, how? Actually this v k transpose, W k, v k here only this you take this I and then open out this bracket, then you will get v k transpose w k, v k. Why this term is added, because this term is added here; this term is added and subtracted here, I think it will require little algebra and and why it is added here? To to to make it into a whole square, to to to get it into the positive definite form; it is a its a its a manipulation trick. So using this term here, they have made it into a positive definite form, okay. That that is that is the whole idea. You can actually multiply it out, and see whether term by term it cancels, okay, it will.

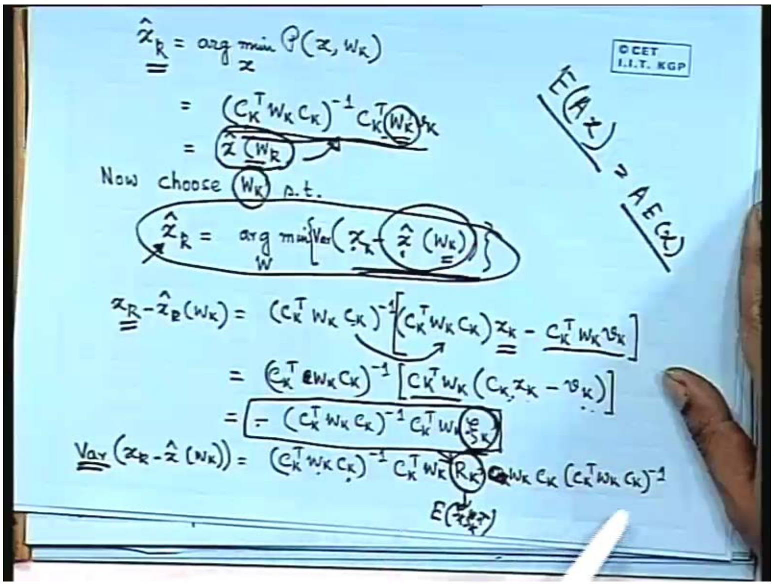
So the idea is that, term one contains x and it is non- negative definite, actually it is non-negative definite. It can be zero, if in fact that is our solution. So then this is always non negative definite and and this does not contain x. So that is why this will always remain in the performance index. So term 2 is does not have x, so therefore this function is always greater than or equal to term 2, correct? Is there is is is agreed or not? Now what is following one? Two plus one

[Conversation between Student and Professor – Not audible ((00:24:10 min))]

term two plus something

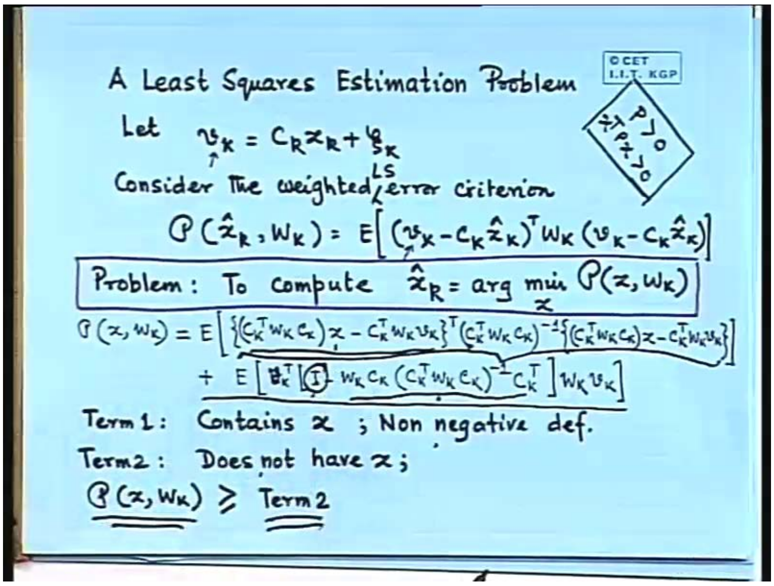
So so when will it be become, so minimum it can become is when it is equal to term two; when will that happen, when this is zero, when this is a this is exactly zero. That gives my solution of x, that gives the optimising solution of x, okay.

(Refer Slide Time: 24:32)



So x is simply you take from here, so this should be zero. So this inverse this; remember that when I am saying this inverse, this I am existing, I am I am assuming that, this inverse exists. So that is number one assumption, that this inverse exists, need not exists but we are assuming. So this is the weakness of this proof, that it requires some additional assumptions. The the the Kalman filter can be proofed without these assumptions also, but we are using this proof because it is simple. So under the assumption that this inverse exists, otherwise you cannot write this like this at all. So under the assumption that, this inverse exists this problem has a solution which is this. So this is the least square solution, given the weighting matrix W k is simple least square solution. Any least square solution will will always be obtained like this, there is also another way of doing this;

(Refer slide Time: 25:47)



for example, you simply break it up open it out and then simply differentiate it, with respect to x, that also will give the same thing. Remember, we when we when we optimize FIR filters, we did that. So it is a you you you can do either way; either I did as a some of squares and then say apply this argument or you can simply differentiate with respect to x, that also will give the same same solution.

So now we have chosen, this is my solution for a given w k to that problem, okay. Now the question is that, I want to choose W k, because because where from this w k has come? We we do not want w k, okay. So we now we want to choose W k; now obviously you see that, if you you cannot say that, you cannot say the for example, if you say that now now now let us choose W, let us choose a W k such that this becomes the minimum. See this has the minimum for a given w k, so I will say that now which is the W k which will give the minimum of minimums, I can ask for that; but but that in in in this case gives you a nonsensical solution, that will that will be I mean you will always get the minimum of minimum, when you have W k equal to zero, because these the quadratic form. So it so it is so its minimum value is zero, when does that happen? When, W k equal to zero.

So then W k is equal to zero, does not get you anywhere. You you cannot get a solution from there. So therefore we cannot we cannot use that formulation, rather we are using this formulation. That let us try to choose x hat k such that, this one the true state minus the estimated state; see for any W k, I can get an x hat W k which is this. This x hat is actually a function of W k, right. So this solution is is actually a function of W k. So what will be the final error between the state true state and my estimate that is also a function of W k? So I can minimize that function. So I can say that, what what is that W k such that if I solve a least square problem with respect to that W k, then I will get the minimum variance of my estimate with the true state. That is an in fact we want, in fact the Kalman filter is a minimum variance solution.

It is a minimum variance estimator. So we are so so now we are trying to bringing that minimum variance. So now if you, so now what do we have to do is that, you will you simply write it as a function of W k and then and then you have to minimize it, minimize it now with respect to W k, right. So now you have so now in this place, we have already put that function. See C k, C k transpose W k, C k inverse. So this into this multiplied will give you just x k, which is this x k. And this into this is your x hat W k, this function. So so this is yours, so this is your function and then finally you can take this C k transpose W k common here, then you get this common and then you get this as xi, k. So this is your error, still as a function of W k, this minus this it is still a function of W k. Now you want to minimize the variance of these. So so what is the variance of this? Variance of this will be, what it will be expectation of this into this transposed, right.

So you go on doing this into this transpose minus minus will become plus, so on both sides you will get these two will be symmetric thing, if you multiply. So this into this transpose so this transpose means A, B, C transpose is C transpose, B transpose, A transpose. So now first term, you will get xi, k transpose; this is now what what is this? Remember, this is expectation of xi, k. xi, k transpose, can you read this? So this xi, k and when you do the transpose of this? This this xi, k transpose will come in the beginning. So this xi, k, xi, k transpose all other matrices are constant matrices. So therefore the so therefore the expectation operator will go straight in.

What is expectation of A x, where where A is a constant matrix; it is A into expectation of x. So if you take expectation outside, the expectation will go through all the constant matrices, anything which does not contain a random quantity, expectation can be taken inside of that. So all these matrices are constant matrices, constant are known. So there is there is so there is nothing random about them. So when you take expectation outside, that is this variance; the expectation will go right in and will settle here. That is this is the only uncertain quantity here, only thing which is random. So the expectation will go right in,

[Conversation between Student and Professor – Not audible ((00:31:26 min))]

System no no system is changing but remember that, that we are now first of all there are there are two answers this question; first question is that system is changing but it is changing in a known way, we assume that A k, B k, and C k are known at every instant k. So there is nothing random about it. Second thing is that, we are considering it at an instant k, right. Right now we are trying to solve the problem, at an instant k. So therefore our C k is constant, correct?

why do not you think.

[Conversation between Student and Professor – Not audible ((00:32:01 min))]

Because, we wanted to wanted to finally come to solve this. So we are doing it two steps. We want to finally want to solve this, but we do not know how to solve this. So therefore we have first solved the problem and then we are choosing, this is this is just a way of proving, right. So I mean remember that, we are not using any any any Bayesian estimation, nothing. We are we are just simply using normal optimization. Normal least square optimisation, I mean just simple least square optimization and and matrix algebra. We are not using any statistics, we have not made any assumption that that these things are Gaussian or anything like that, that is a that is the advantage; I mean you can you can follow these steps, after all this is simply normal I mean algebraic manipulations that I am doing.

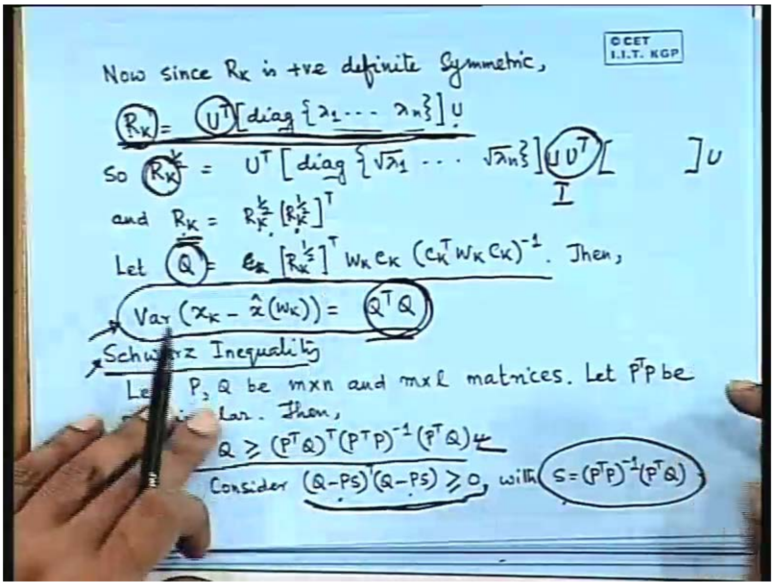
Otherwise, I would require some theory if I want to if I want to for example, if you read some other I mean very well-known text, you will find that they will first prove some some conditional estimation results without going to the Kalman filter and then, they will pose the Kalman filtering problem as the conditional mean estimation problem and then apply thos results. And then out will come those Kalman filter equations, that is another way of proving but then you first need to know I mean I mean, how to calculate a conditional mean.

Student< Coming right from top and not changing its coming in a >

[Conversation between Student and Professor – Not audible ((00:33:32 min))]

So far so far we have not changed it, you see we we have so far; first we have tried to find, see that if we that if we choose a W k, how to obtain the optimal solution? That is very simple. Now we are find so so then for for every choice of w k, I will get an optimal solution. Now we are actually we want to minimize these, you know. This is our actual, we want to find the minimum variance estimate of the state, that is our objective. So now we are saying that, okay now of these optimal solutions; if I now change W k, which one gives me the best solution? So first I keep W k obtain a solution then see look at it the function of w k and then try to choose a good w k, I am doing it in two steps, okay. So so this is my variance, which now the now the question is what is that value of W k for which this variance will be minimum, okay? So that is not a not, does not appear immediately apparent its.

(Refer Slide Time: 34:50)

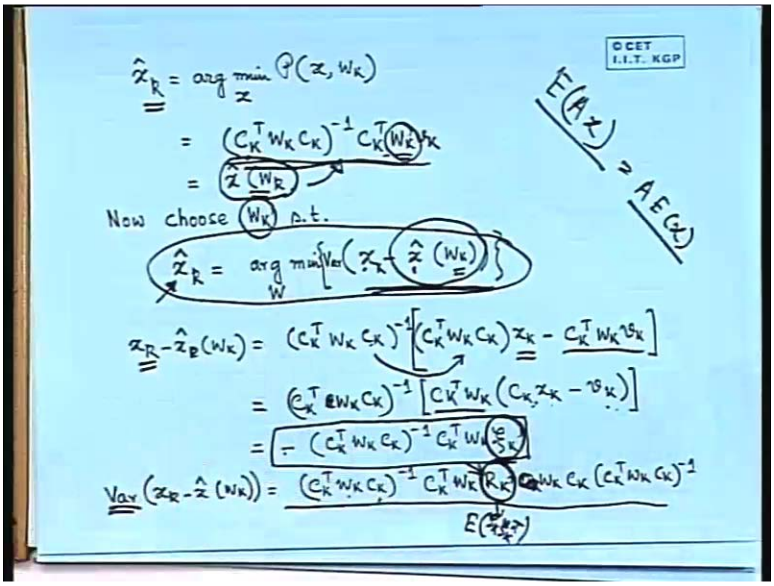


So now again some this is this is all matrix algebra. So now this now what they are to they are trying to do is; okay that I want to express it as, there are you know there are certain matrix algebra results. So I want to cast it in a form where I can use some matrix algebra results. I want to get the minimum value of these that is what I want to get. So so how do I do it? So first I say that since R k is positive definite symmetric. So I can break it down, remember that when we when we made coordinate transformations on of for the for the adaptive filters; we had we had made a made a coordinate transformation based on based on Eigen vectors. And then we said that then then it will become decoupled. So so what we are saying is that, any positive definite symmetric positive definite matrix, can can always be be broken up like this.

This is called a UdU transpose factorisation of the matrix, this is this is always possible; a standard matrix algebra result. So these U's are called unitary matrices; where U U transpose is that is U transpose equal to U inverse, they are such matrices. And this is a diagonal matrix and this is U. If this is true then then I want to now actually what I want to do is; I want to I can hear size because people are you know people are lost in matrices, that is the Kalman filter, you cannot you cannot be afraid of matrices and derive derive the Kalman filter. So so now what we are saying is that, so we have we have expressed. So this matrix can be can always be broken up like. If it can be broken up like this, now I want to take a square root of this matrix. Now again what is the square root? It is a it is nothing but a notation, it is a its a symbol. So by by by square root of Rk; I mean such a matrix such that that into that transpose will give me this matrix. Where are when like a square root; like like root x into root x gives you x, so in that sense.

So if you have so I I want to find out what is this. Now you see if you multiply this by this, what will happen? That this if you take this transpose, what will happen is that, you will get U transpose here. Now U U transpose is I, so so this will become I, right. Then next you will get this diagonal root lambda 1 thing, next you will get U. Now this diagonal root lambda 1 into into diagonal root lambda 1 will give you diagonal lambda 1. So you will get U transpose diagonal lambda 1, these two matrices U. So you will get R k. So so this is such a matrix that, this into this transpose gives you R k. So it is always possible to express R k like this; why I want to do that? Because then if I choose it in this form, if I choose another if I call Q to be this matrix, this R k to the power half transpose w k, c k then this whole expression, this whole complicated expression, this one becomes equal to Q transpose Q.

(Refer Slide Time: 38:11)



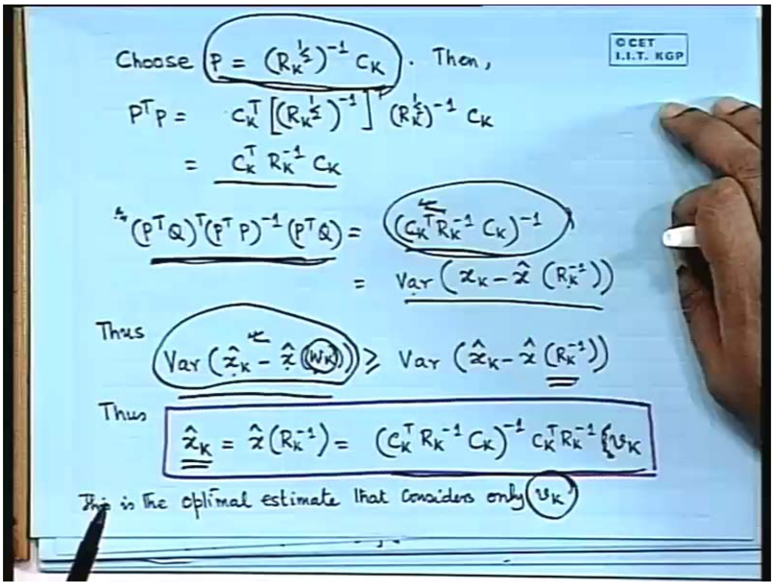
This is why I want, I did I too call the trouble of breaking up then doing root over et cetera. So all this trouble was done because the final variance can be expressed as Q transpose Q. That was that was that was fairly obvious; because you see this is a symmetric matrix, so this and this inverses are same, I mean they are they are they are transpose of each other. See this is this side transpose, this side not transpose. W k is again symmetric, so therefore transpose transpose same and and this is symmetric. So it was in in it was in a symmetric form all the time. So we just sort of you know, formally split R k into a symmetric form and then said one part is Q and the other part is Q transpose. Nothing, it was it was it was already in a symmetric form, only this was sort of this is this itself is symmetric but we broke into a product of two matrices, that is what we did.

So now we have got these results. So so where does it get us now? We use what is called as Schwarz inequality; see finally what we want to do? We want to find out what is the minimum value of this, for which value of W k will this become minimum. I am looking for a minimizing solution of this by choosing W k, right. So again I I will I will express it into a part which contains W k, which is square and I mean I mean that kind of a thing. So I will have to somehow find out, that what is that part which contains W k and is greater, so so that is what I am doing. Now so imagine, that for example first of all see that this is greater than zero always again again this is a matrix. So when I say greater than zero, I make positive definite; why this is greater than zero? Because again if you do do x transpose, this into this into x it will be greater than zero. Why?.... Because something transpose something, x transpose x is always greater than zero. So if you take, suppose you take x into Q minus p s as a vector v, choose x in such a manner.

So then what will happen? This will become what v transpose v, so v transpose v is always greater than zero. It is because of this form; whenever you have something m transpose m matrix it must be positive definite, so therefore this is this is always greater than zero for for arbitrary P, Q, R, P, Q, S. Now if you choose, so now if you you just break it up, you can always break it up; here you will get a term Q transpose Q, you will get other terms. So the other terms you can take on the right hand side, I am not written the whole proof. And then if you choose S as this; you can choose any S after all this P Qs is free, for any P Q S you can do it. See if you choose S is equal to this, then you will get this result. Can you follow this? If you break it term by term, what will you get? If you multiply this by this, you will get Q transpose Q. If you multiply this by this, you will get Q transpose P S. If you get multiply these by these, what you do get? What you will get? you will get S transpose P transpose Q. If you multiply this by this, you will get S transpose, P transpose, PS. So all the other terms other than Q transpose Q, you can take on the left keep on the left hand side. Put all the others on the right hand side and then put S is equal to this. Then you will get this.

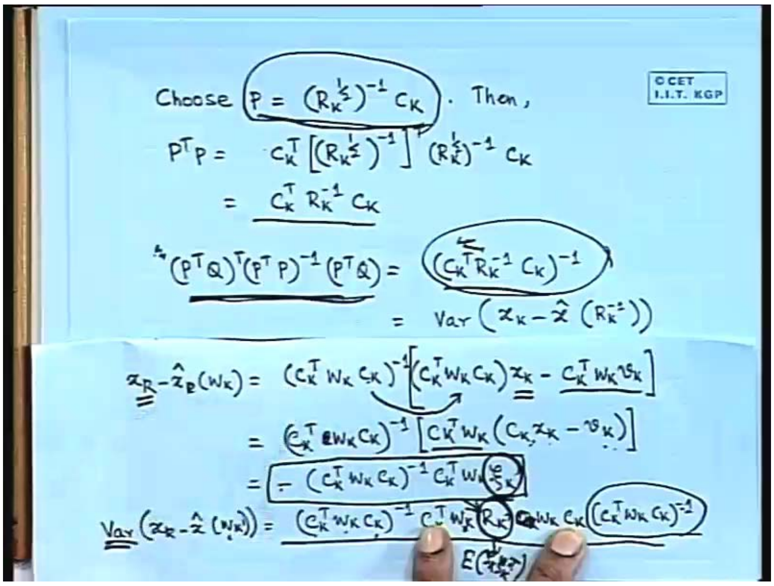
So now Q transpose Q, this is a standard matrix algebra relation again. I hope you can understand it, so now we have to choose with respect to we have to map it with respect to these. So now we have this with respect to this, we have to map this, so what is that mapping?

(Refer slide Time: 42:52)



So that mapping says that, you choose P is equal to this; these people have simply you know by observation and by thinking; people have found out I mean, if you think hard as to you know how to prove, you will get it. So then people have found that, if you choose P is equal to this then you will get that P transpose P is this and this will become this, see how it comes P transpose P is, okay. This into this will give you give you C k, R k inverse C k. This half and half will become R k, that is obvious enough. And now that now the now the interesting thing is that; if you take this thing, then final form becomes this. Now you see, what was yours x, w k? What was your x hat minus, let me just one second… compare with compare with here, compare with this.

(Refer slide Time: 44:01)



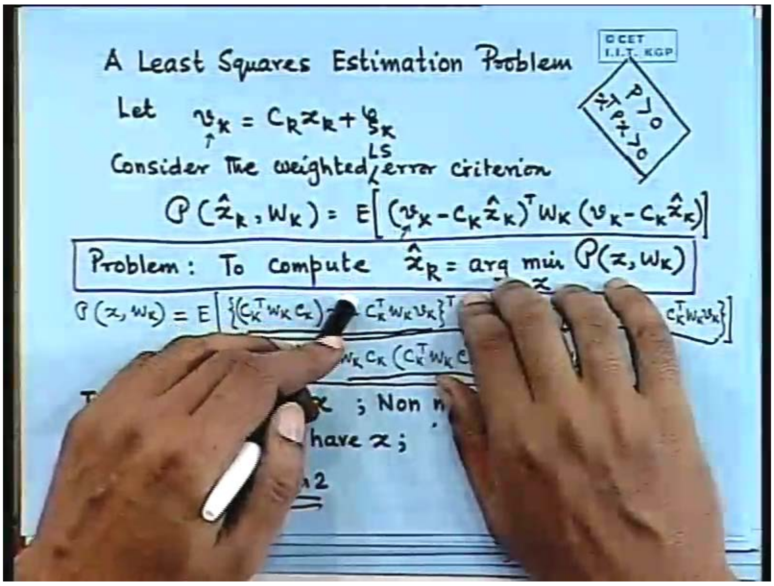
This one, here this P transpose Q inverse is becoming this. Now the interesting point is that; if you put suppose you choose W k equal to R k inverse, we are actually looking for choices of Wk is it not? So if we choose W k equal to R k inverse, then you put you put w k equal to R k inverse everywhere. So here you have R k inverse, so R k inverse and R k will get cancelled. It will I will have I, there will be one R k inverse left here. So these will be C k transpose, R k inverse c k, and this is C k transpose, R k inverse, C k inverse. So this and this will get cancelled; so you will be left with only this which is C k transpose R k inverse Ck.

So which means that, if you put R k inverse here then this expression becomes C k transpose, R k inverse, C k inverse and that is here. Which means that; this quantity is nothing but the variance of x k minus x hat of R k inverse. So now what do we prove? What have we proved? We proved that, we have proved that variance of x k minus x hat W k is always greater than variance of x, x hat k minus x, x R k inverse. It cannot be less than that,.. okay. So which gives me the my, so which tells me that; I have solved the first problem, that if I choose W k as R k inverse, then I get the solution which minimizes this variance. So this is my solution of that of that old problem. Now only thing is that, fine so I have solved the first part of first part of the problem; only thing is that we have I have done all that with respect to only vk, I have not considered any past, anything. But if I have to do the Kalman filter, have to have to find a minimum variance estimate for; that is it it should be it should be something which takes care of all the past data, because all the past data is available.

This is this is an estimate which does not take care of any of the past data. It only considers v k one measurement, but but ideally speaking; if you if you really want to construct something really optimal, then it should consider all the past data because they are available. So whatever information has to be extracted from there, you should extract that. So so now we have to show in the second stage of the problem. Now that we know that, how to solve this problem for a for one measurement. We have to see in the second part that; how to solve this problem, for which will optimise this kind of variance for the whole past at every instant. So that will turn out to be the Kalman filter and incidentally; it may me let me again I mean reassure you in fact, that is sounds you know a lot of you know jugglery with matrices, but this is an this is an extremely effective algorithm. It has been applied very widely, gives fantastic results, you will find I mean what amount of noise it can clean.

So it is extremely practical and used in all aerospace applications, so is is not that, that is why it it it is so famous because it works, okay.

(Refer Slide Time: 48:23)



So and in fact, we will realise how it works because at least on on this particular thing, we more than having a tutorial, I mean because this is this is a recursive algorithm; whose greatness you will only understand by writing a program and then running it on thousands of data points, you cannot do, you can probably do some theoretical kind of problems in tutorials, but you can never do it in tutorial because you need a computer for doing that.

So we will have an assignment and I will formulate it, in which you will have to write a Kalman filter program; and try it with different process noise, different measurement noise and see I mean, how good and how bad it is, okay. Only then you will get some get some feel of the algorithm.

Student <Sir, how can we know that other>

[Conversation between Student and Professor – Not audible ((00:49:07 min))]

Student<We are choose them, wk is equal to Rk inverse, how can we know that other solutions is possible?>

Other solutions are other solutions are definitely possible, but then we have proved that if we choose any other W k; you will get you will get higher variance, that is what we have proved now. So we have found, that is what we have proved in the last step; that variance of x k, this this one for any other choice of W k is going to be larger than this one. So you you so you cannot find the W k which will give you a variance; you cannot find a W k and then solve the earlier problem, which will give you a variance less than less than this one. That is not possible, okay. So that is here we stop, and in in the next class; we will we will we will consider the second part of the proof which will be the…..