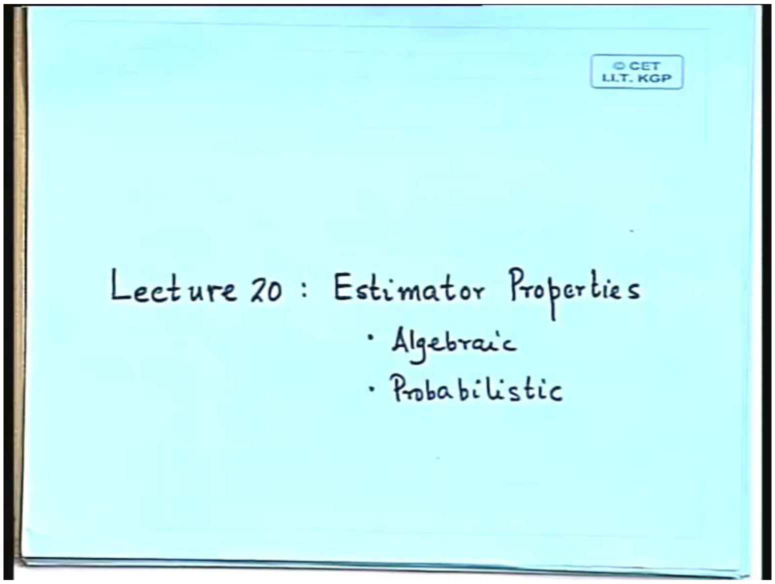
Estimation of Signals and Systems
Prof. S. Mukhopadhyay
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 20
Estimator Properties – Algebraic & Probabilistic

So in this lecture we have seen one approach to derive the calven filter. So what we did is the first we did design the weighted list were function and then we design the way such that, estimator is mean and variance standard optimization approach.

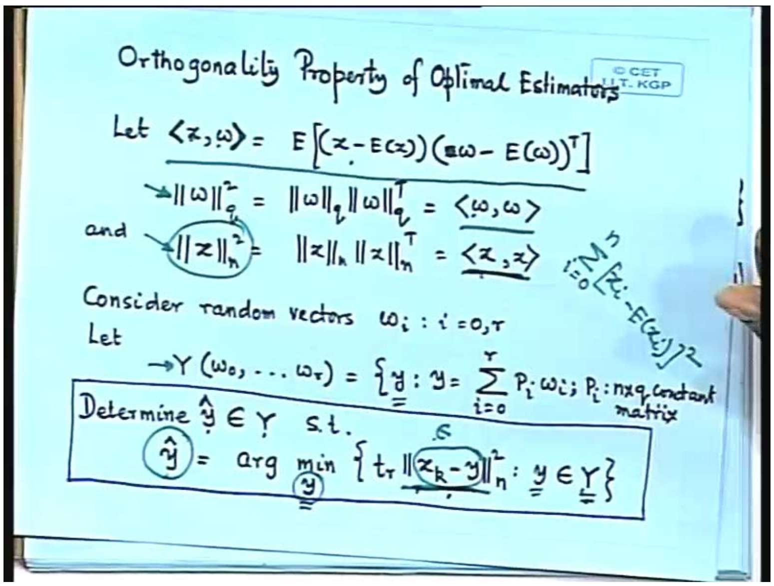
(Refer Slide Time: 01:14)



Now typically you will find the calven filter is derived in to other ways, and these derivations basically originate from two major properties of what a known as, minimum variance estimators. So two major properties one of them algebraic and the other based on probability theory. These two results are also used to derive the calven filter, typically you will find in books. So while we will not re-derive the entire calven filter; to that will involve another sequence of algebraic manipulations, but we will definitely look at this property but this properties are very important.

There were fundamental properties of this vector and we will remember that, utilizing these properties that is if we try to constructed estimated, which will satisfy this properties; we will again arrive at the same same equations and the same calven filter that is why we need to review. So first property we will start with the algebraic property which says that, which is your some sort of orthogonal orthogonality property, we will also encounter this property for.. in the in the previous case; previously when we discuss the a fire filters for predicting the output. So suppose the idea is that, that suppose if we we had some we had some measurements, okay.

(Refer Slide Time: 03:04)



So we are not right now saying that their measurements; we are saying that suppose we have some random vectors w, and now based on this random vector we in some way we want to construct and an estimator of vector x. This random vectors x and w are are related in some ways. We are not right now concern about that, all we are saying is that this estimate that we want to derive, want to generate which am here calling y, is not confuse with the output y; in this case y is an estimate of x. So I am trying to generate an estimate y of x, using these w and not only using them I am using all the W's from i equal to zero to r, and am not only using them am using them in a linear fashion. In the sense, that I am this y that actually constructed by multiplying this w, with constant matrices behind; which have to choose in some suitable manner, so that this estimate is good. But it is and it it is a linear estimation because I am multiplying each of this w's, these this w's are q dimensional while this x all are in n dimensional. So we have to multiply this pi by n into cube matrices. So am multiplying, so basically multiplying a vector by another.. by a by another matrix means; it is lengths scaling other we are stretching the length, and we are also re-auditing the vector, right. So we are basically this called a linear transformation.

So we are passing a linear transformation, using only linear transformations from this w here construct it is y, this is my only constraint; that I am I am not using other nonlinear functions. Now if I do that then the question is that, using this all of a rule that is only using linier transformation; what is the best estimate that, I can create in an in a in a mean square sense. In other words, so now that that sense will have to define; so I am now defining the sense here, so I am define the quantity which have which have which I denoted by this strangular brackets. Sometimes is called a called a inner product or so it is a inner product of these two vector, one of them is n dimensional the other is q dimensional. So derive so then this matrix is becomes an n into q dimensional matrix, okay.

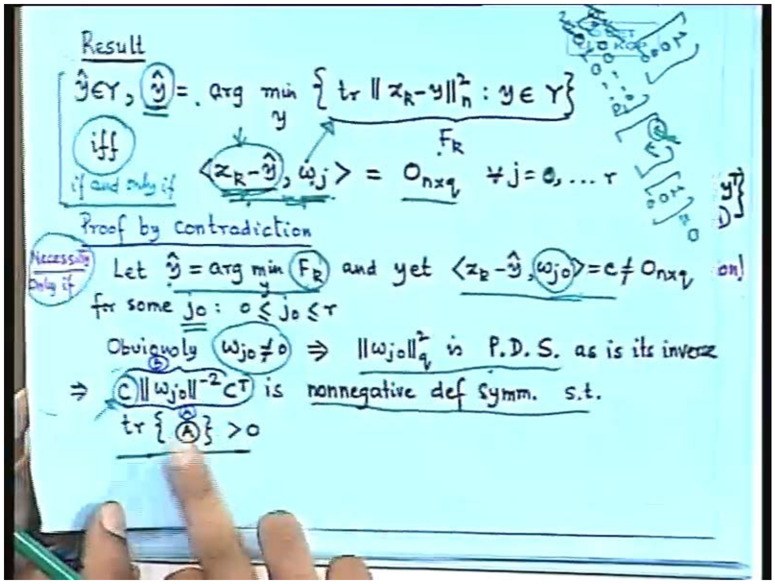
So it is so this inner product is actually define in this manner. So if you define this, it is nothing but the covariance of x and w, this series. So what is the trace of that covariance, it is a it is a what is that what is that trace of this matrix? This is actually this is the here here defining trace will be difficult to get the n into q matrix but but when we have trace of xk minus y where this, so this xk minus y now okay. So first let us let us give this this definition of the inner product, then if this w and x are at the same dimension in in in other words; if we say, if we define x, x or if define w, w then this matrix will become square, if it becomes square it will have diagonal elements.

So trace means, sum of diagonal elements, okay. So what is a what are the sum of diagonal elements of this x, x? So it it will be it will be x one minus expectation of x one whole square, next diagonal element will be x two minus expectation of x two whole square. Only the diagonal elements are considering. So actually it is the various of it is sum of variances of x one, x two, x three; this when when I say trace of this, I basically mean, it is sigma i is equal to 0 to n, xi minus expectation of xi whole square. This is the trace of x x, similarly trace of w, w, rather.. so when I say and this x x I define as the square.

So when I say this, what I mean is what is the if if this can be called as some sort of an error if silent; this xk minus y, y is supposed to be an estimate of xk, if you xk is the true state. So if you do xk minus y, you get the error. If you take trace of this square, you what you get is that, you get the sum of squares of the rather sum the sum of the variances of each components of this equivalent. So so that that is what I want to minimize. So I want to choose a y from this space y. This space y is the set of vectors y which can be generated by such linear transformations; only from that set I must choose y, not from any other set. In other words I must only use the linear estimator; if I use the linear estimator, then what is that y which will minimize this? That I call y minimum variance of estimate. So the minimum variance estimation problem says that, determine such a y hat belonging to y obviously because each each of this are generated from this, so obviously the minimizing solution will also from this.

So find out that minimum, that linear estimate which minimizes the sum of variance of the errors squares. So that is a minimum variance estimate which has been calculated using linear transformations and using all the vectors wi, okay. Now it turns out that this estimator has a very nice property that is what we want to discuss. So what is that property?

(Refer Slide Time: 09:45)



So the property is this, that if actually happen to solve this problem and get this estimator y hat, then it turns out that; the the covariance of this error with respect to the will will be orthogonal to the measurements, to all measurements, right. When this will be a n, this will be a n by q zero matrix, this o means zero; that is each element of this matrix will be zero. So this is so so in that sense the error of estimation, if you have constructed y, using wj's; so it says that, that is the optimal estimate which after the.. after you say, you know sort of taken out all the juice or or or the all the information possible from wj of xk.

So now you have calculated some some y hat, but still you could not get the fool xk. So those error are those, that you cannot get from wj. So when when can you not get anything from wj? When this error become orthogonal statistically to to to wj, then you cannot then then you cannot extract any further information about those those errors from wj. So so these are sometimes called innovations; I mean in a sense... and actually what is it says that, this is the estimation error will become orthogonal to measurements, which have been use to generate this estimate that is the whole idea. And this result rather strong, this result says; if and only if which means that if you happens to generate it is by some other method, this property will be satisfied.

So if this is satisfied this is satisfied, not only that it is the reverse also. If you can somehow construct an estimator for which this is satisfied, then automatically this will also satisfies. It is both way implication, this implies this and this implies this. So if you can satisfy some estimates which will satisfy this, automatically the minimum variance property will be satisfied, right. So this is called a necessary and sufficient condition, by this iff what we mean is, if and only if and now we are you have heard about this. So suppose this is a logically speaking; suppose this is a Claus a, this is some statement which mean mathematical symbols, it is basically a sentence. So if that sentence is a, and if this sentence is b, then what are this say? This say a, if and only if b which means that; a if b and a only if b. So a if b means what? Means b implies a, if a if b. So b is true, a is true. So b implies a logically speaking and a only if b means.. a implies b, okay.

So how do you prove this result? This is a this is a very very fundamental result of regarding the calven filter, because the calven filter is also a minimum variance estimate, that we already seen, right. So so for the calven filter this will be true; in fact you can, that is what I was saying that, in fact you can construct the calven filter, by trying to construct the estimator for which, this will be true. Then then automatically this will be true. So that is another line of proof or or another line of derivation of the calven filter. So this is a I mean also told I mean want decided to discuss the proof, because this is a proof by contradiction. I do not know you you must have seen proof by contradiction where where you first assume that, you know you assume that the theorem is wrong and then you will arrive at some obvious fallacy like that; like may be the sun rises in the west. So then you will say that, sun cannot rise the west therefore the theorem is wrong. So you know you know this kind of arguments.

So the idea is that you first assume that, the theorem is wrong and you can show that, if you assume that, the theorem is wrong; then you can arrive at some obvious obvious contradiction like, I mean something like one equal to two that true may be, which is obviously wrong. So so we are going to do that, okay. So first we prove necessity; necessity means we want to prove that, if this is true, this is true. That is this implies this, a implies b you want to proof that. So we first say that let a not implies b, which means that this is true but this is not true. We have to assume that; first it is proof by contradiction. So we are assume that let y is this, this is this whole thing and and this cumbersome to write so I call it as F k. So let y is a minimizing solution of this

F k, but let let let there exist some wj0, one of them actually actually this is this should hold for all wj j equal to 1 to r, but let there exist if it at least one for which it does not, oka. So if that is true there is there is a some j0 where j0 belongs between 0 and r. This this will be 0, such that this is not true I am assuming.

So then you have to prove that, you have to then then actually what it will show is that; if that is a case then then obviously this cannot be the minimizing argument; in other words we will show that under this condition, we can find some other vector y, some other vector y which is also linearly united from wj zero, but for which which will give a minimum which will give a value of F k which is actually less than that of y hat. Therefore y hat cannot be a minimizing argument, this is our line.. So what a we have going to do? So so basic idea is that, so naturally if if this is not equal to zero; obviously this is not zero, because if this was zero this would be identically zero, correct.

So so then this is not zero and if this is not zero, this is this is positive definite because, why? Because this can be written as coherence matrix, so so it is a positive definite. So if it is now if so then it is positive definite and if it is positive definite, its its inverse is also positive definite, okay. Because of this, why? Because of u d u transpose factorization, you can always inverted it. So if you have, u d u transpose factorization; if you invert it you will get, u inverse transpose. U universe transpose is u then d inverse; then so you will find that again you will get u.

So if you inverse this matrix, then you will get you will get another u d u transpose factorisation with the d matrix now is d inverse. Now d matrix is a is an diagonal matrix of of Eigen values. So if if if all the Eigen values of d were positive, then then then all the Eigen value of now going to be one by lambda one, one by lambda two, there also positive. So therefore, this this inverse will also be positive definite. If this is positive definite then then this is positive semi definite. Why? Because, why why this semi definite? Why has it true? Why is there this is a this possibility of this being zero? This possibility of this being zero is because of this c matrix.

This c matrix has now n rows and q columns. If you can see; c row, c matrix as n rows and this n into q. So so n into q means and n is greater than q; generally this state dimension is is greater than the measurement dimension. So it as more rows than columns; which means that it is wrong definition. So because its this wrong, can never be more than q but it has n rows. So therefore; it itself is wrong definition. Therefore therefore there will definitely be a exist some vectors y such that; this will become zero but it so it will be either zero or positive, because this is positive definite, okay.

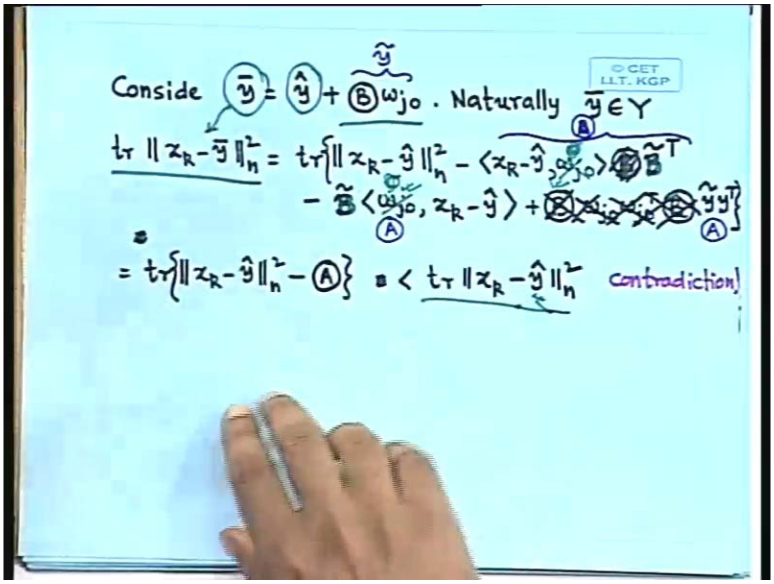
So so therefore this is non-negative definite. See you have to there are some little little you have to go steps very very very minute details; because we cannot mathematicians never never leave any any thread loose, they will tie up all the threads, okay. That is the beauty; nothing is left to chance. So so this is non-negative definite, fine. Then now now there also telling that, even if it if it is non-negative definite, the trace of this will be greater than zero. Here there are not putting greater or equal; though this is not this can be zero, this x transpose something can be zero but the trace which is sum of all the diagonal elements that can that must be greater than zero, why? Because, if the sum of if the sum of this is you can you can find out, even you can actually, we can solve solve if you are interested, because solve solve this n... It it is the skill of manipulating matrices, it turns out that that, if you if you have this trace as; for example, now this is nonnegative definite which means that, if you multiply it by y transpose, this into y it will be either zero or positive.

So so so let us choose one vector y, as first element 1 and all other zero; using this if we calculate y transpose, this y what will you what shall we get? We will get that first, will get this matrix, one, one nth element, can you follow this? If you multiply a matrix by this transpose, that is a one zero, zero transpose some matrix in to 1 0 0, what you get? You get this gets picked up, correct. So if this sum, so each of the elements are are positive. So if the sum is if this trace is greater than zero is is equal to zero, it means that each element is zero. So it means that, now you if you multiply by this vector, it will be this will be equal to 0. Next time you will multiply by 0, 1, 0, 0. Then the second element will get picked, that is also zero.

So which means that, this 1, 0, 0, 0, 1, 0, 0, 0, 1, 0 if all this vector if you multiply by x transpose, rather the y transpose that matrix y, will be zero. Now these vectors form a form a , basis form a set of basis vectors. So that means any vector which you choose can be can be written as, sum of these vectors, which means for all vectors; if this is equal to 0, then for all vectors if you multiply y transpose c y, it will be zero. That is clearly not the case. So slightly involved argument, but that is why this we can assume to be greater than zero, because these c matrix at least has some non-zero rank, if this c matrix has rank zero; then only this will be equal to zero, but that is not the case at least there is one measurement, so this rank will be either at least one, okay.

Now we are going to do that, so after we have proved this, now we are going to say that, we can now construct; see previously we had assumed that, y hat gives the minimum variance solution, now I said no no no if if you find such a c then I can give the another y which will give a lower value of F k. We are we are we want to prove it by contradiction, right. So now I will construct that, somebody actually thought it out; somebody actually spend lot of time and actually thought out that, if you cleverly make this y bar like this then so you just have to you know you are you first have to see the trick, which somebody thought out.

(Refer Slide Time: 22:15)



So somebody said that, okay lets construct this y bar as this this so called minimizing argument plus this one. So he has added another term, so obviously he had generated a new solution. So now he will show that, this we using this trace if you calculate, it is going to be less than the the trace, that is supposed to be given using the optimal one; but you initially what did you we assume you assume that, that that this optimal one. So this trace can never be lower than for any other estimate that is what you assume.

So obviously you are assumption was wrong. So why your assumption is was wrong, because you assume that that matrix is non-zero. So therefore, if this has to be minimum that has to be non-zero, right. So so this is the this is the this one side of the proof, where do you started assuming that that is the minimizing argument and then will say that if it is the minimizing argument, then this must to be zero. It is such a thing cannot exist; it cannot be seen non equal to non-zero, this this one way. Other way is to start from the other way. The other way is now; suppose this is zero

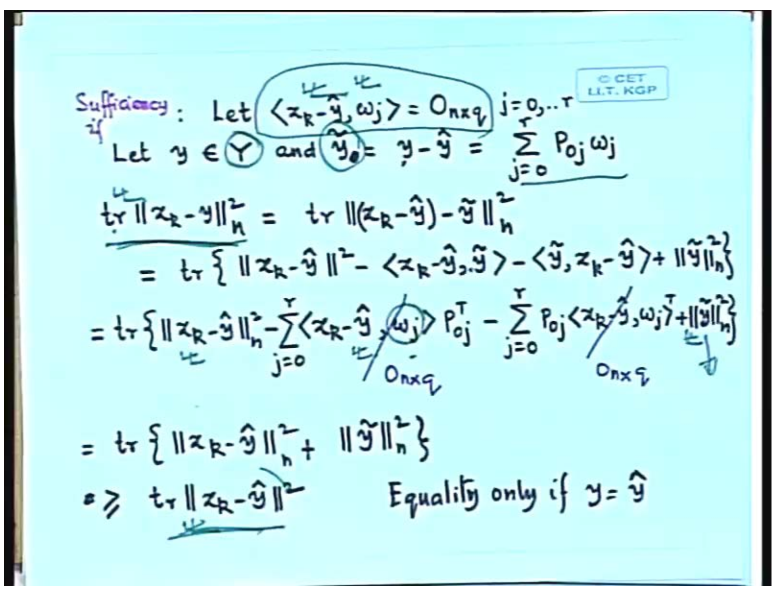
[Conversation between Student and Professor – Not audible ((0:23:40 min))] ya < a_side >

Student<what is this b term? < a_side >

This b term may actually taken actually is very difficult to write. This b term means just this part; leaving out this c transpose, this a term is the whole term. I just you some symbols so that I do not have to write it every time, okay. It is a just a clever way of constructing and then they will this this this very simple; this y bar they will put this and they they will break it up and then and then they will show that, this is a this a and this is a. So this is this minus a and then a is positive; so therefore this must be less than this. Just simply simply decomposition, okay.

So anyway this is this proof is not to you know line by line remember; this proof is to just to just assure you that it it it indeed comes and to give an exposure to a way of proving, that you can prove things by this. This as this as the very popular way of proving things; many many theorems are proved by contradiction, okay. Either you can prove by contradiction or you can proof by construction, right.

(Refer Slide Time: 24:53)

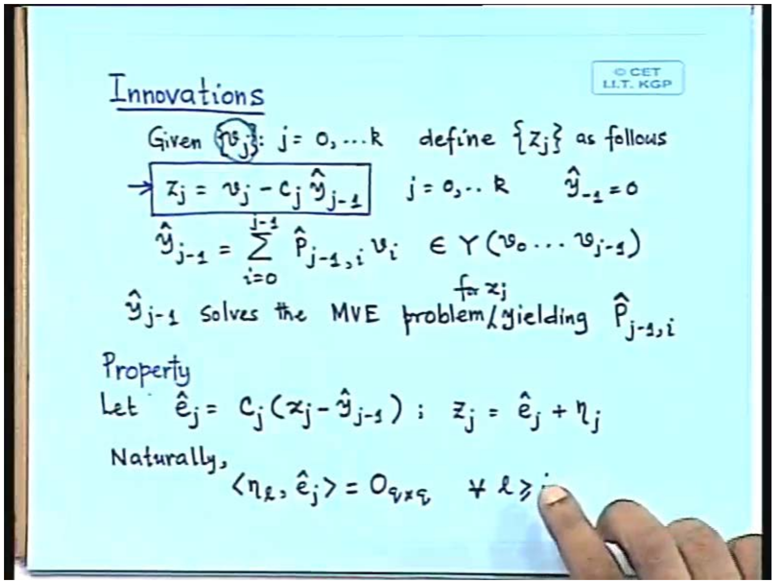


So so so here is a prove by construction not contradiction. So here they will say that, they they will directly prove that; if if if this condition is satisfied, now they are starting from this condition then they will prove that, for any estimate y which is not equal to this y hat, the this function going to be larger than that for y hat, directly they will prove. So how they will prove this? So they will say that, okay suppose y y hat is suppose y is different from y hat; in that case y tilde is not zero, it is some vector. So therefore and and obviously; since y is also chosen than using some matrices of wj and y hat is also chosen from the same space, so therefore y minus y hat can also be chosen from in this manner, just you now abstract this matrices.

So y y tilde also belongs to the same space, this y, correct. And now then simply calculate this; so again same thing, write it as y is equal to y tilda plus y, y head. So after then again break it up and it turns out that; since y tilda is see now now you are using this property which we assumed, so this this inner product can be written as a inner product, sum of inner product assuming with wj because y tilda is this. You just substitute this thing here and then calculate the inner product. Now each of this inner product at zero by this assumption, so therefore this two terms should go away; they will become zero and only this will remain. Now this is obviously positive because it is the same vector square; the same vector square must must be positive. So which means that; this is this plus this, this one is positive which means that; this is this is greater than this. There there is no contradiction, they will directly proving, that is we will assume this you can show that for any vector; for which this y tilda is non-zero this this will be more than for this, right.

So it follows both ways you know. So so this is the very very important property that; the error the estimation error becomes orthogonal to the measurements which have been used for generating the estimate. This is a very important quality of minimum variance estimates, okay and that is used for generating the calven filter.

(Refer Slide Time: 27:28)



So in a I am actually for generating a calven filter because you know properties can become very complicated, so therefore first what they do is what is known as orthogonalization; what is orthogonalization? What they are saying is that actually you are supposed to generate, suppose this is v j's okay. So these v j's are the; remember that what what were our v j? v j equal to, this was the equation; c j x j plus eta j. So the v j's are actually the measurements, so we are supposed to generate our estimate using the v j's, right okay.

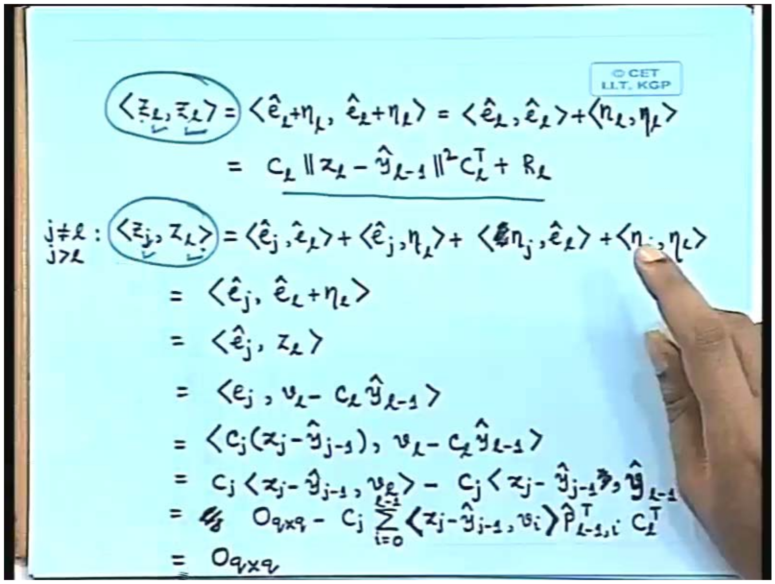
But rather than doing that in this we have you know, complexity of proves. So rather than doing that, can we transform this e j's to another equivalent set of sequence. So what they are doing is there, they are now defining in terms of e j; that defining another sequence z j. This z j are called innovations because they have a nice property and once you have this property, you can show that first you can show that; estimating based on v j's is the same as the estimating based on z j's. It is it is identical because, why it is identical? Because z j's are nothing but linear transformations of v j's, is is you y hat has been generated using y. y hat has been generated using information from v zero, v one, v two, up to v j minus one and here is v j.

So this z j have been again generated by linear transformations; using v0, v1 up to v j minus 1 here and v j here. So there also linear transformations of v j; so therefore this z j sequences are are nothing but a new sequence which is generated from linear transformations of v j. So therefore generating based on v j's is the same as generating based on z j's, that is the that is the argument which is being given; but this z j's have a nice property, what is the property? That property is that, so now obviously you are generate, what is this? This is a this is a minimum variance estimation generated using v i's. So what they are saying is that, first take the take minimum variance estimate of x j; generate the minimum variance estimate of x j using v i then what's what is c j, x j subtract it from v j, then you get what is known as the estimation error. Now this estimation error is called the innovation, because it has certain property; what is that property? That property is that the property is this.

The property is that, the co-relation wise, now we are talking about this z sequence which is the innovations; that property is the only z l, z l co-relation is non-zero, for all others z j, z l is zero. So you have sort of you know made made a made made a wide sequence. So again what is this? This is the see, this is the estimation; this is your estimate of the output, what is your estimate of the output? The output is suppose v is to equal to c x plus eta; you have generated x, you have got an estimate of x hat. This is totally uncorrelated with as as we assume that, it is uncorrelated with past measurements etcetera etcetera.

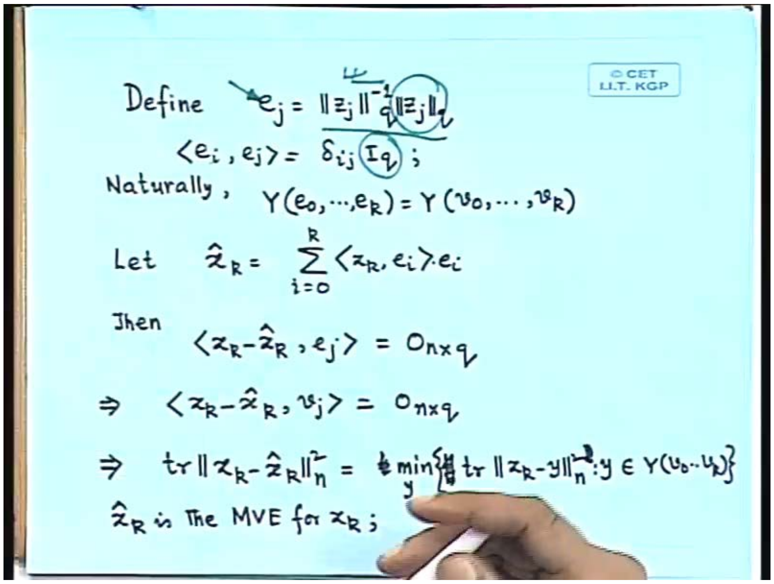
So if you have an x hat, what is the estimate of v you can generate, obviously c x hat. This is your estimate of v using x hat. So but but you got v, so then v minus v hat is the error. So that v minus v hat is nothing but this z. So it says that, this z's become white; so if you use the estimates then the output estimation errors will become white, in the sense that their co-relations functions will become derived delta functions, only if it is z l, z l it will be non-zero, if it z j is z l, that j and l are different then it will become zero.

(Refer Slide Time: 31:58)



So again number of manipulations and skipping them, so this is the so so what have you done in effect, what we have done in effect rather than using the measurements; we have created an alternative sequence called innovations which is linearly generatable from the measurements but which is white. Now obviously if if you deal with white sequences, all all our analysis everything will become simple that is the advantage. And now what we are going to do is now you you see that and then you have you know, you have normalized it normalized it; it means you have you have found out, you have sort of scaled it.

(Refer Slide Time: 32:43)



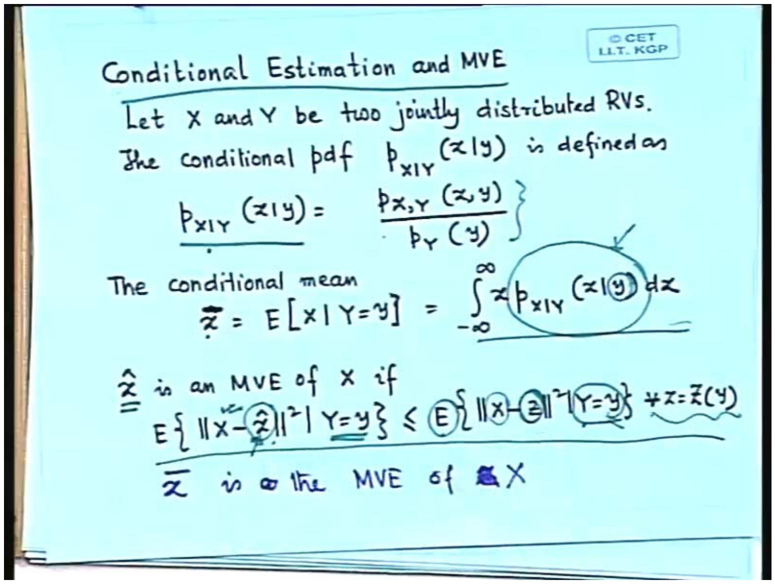
This z j's you're scaled using this matrix. So once you are scale them then and and you generate this you know; sort of normalized error then what happens is that that this co-relations becomes identity matrix, previously it is not identity it will have this this term. You have scaled it, so now what you are going to do is now if you generate it, now you can show that if you can generate your estimate; now we we want to generate a minimum variance estimate, that our are ultimate goal that that is the calven filter.

So so an alternative approach will now say, that if you generate this one; so once you defined this e i's, now so so you see that even the measurements you can generate this e i's, even the measurements you can first generate the v i's, the the z j's and then from z j's you can get the e's. That is recursively you can do it. So if you can generate this x hat k, based on these in this fashion; not this is not a this is not a computable form, this is this an abstract form. This this cannot we are not going to compute in this form, you are not finally compute the estimate but if the estimate whatever form you compute; if it becomes equal to this then we can show that it will be minimum variance, how?

Because, we will show that, this will be zero that is very easy to prove. So, if if now now now we are in previous theorem, previous theorem what do we say? That if this is my estimate, so if this is my measurement error; so if the out if the estimation error becomes orthogonal to the measurement all measurements then my corresponding estimate is minimum variance, that have just proved. So it it shows that if you can generate the estimate in this form then it is very simply shown to be zero. So therefore this is the minimum variance estimate in terms of innovations and then then I am skipping the next part. Then what they will do is the, they will they will actually try to find the nice form, nice computing form which will be equal to this that is a lot of algebra. So they will actually compute all these e i's, it is is using the using all the a b c d matrices and then they will show that; if you really want to compute this, you get the same calven-filter actually the calven-filter compute this.

So that is another derivation of calven filter, so we are not we are not going to do that derivations but all we want to say is that; this is another approach of deriving the calven filter which uses the whiteness of the innovations property. So we want to so so in that approach we we want to construct a filter such that the innovations become white; and in turns out that if we can construct a filter such that the innovations are white then we arrive at a minimum variance estimator which is the calven filter, that is that is an alternative way of derives, okay. So I am skipping all the you know painful other details.

(Refer Slide Time: 36:01)



Now we now we are going to prove another one; which is this is this algebraic one, so just remember this result that the algebraically the innovations will become white or the estimation error will will become orthogonal to the measurements which is used for generating the estimate. In other way as if you know it will it will it will get it will it will only leave that part, about which no no information it is it is possible to extract from the measurements. See the estimation error becomes orthogonal to the measurements; so if something is orthogonal to the, what is orthogonal mean? Orthogonal means that, when are when are two sequences orthogonal?

[Conversation between Student and Professor – Not audible ((00:36:49 min))]

Right, inner product is zero. Inner product is zero means that, that that the covariance is zero. so the so their variations have no co relation, that is sometimes this is this is positive that going negative; you cannot act you cannot get any information of the variance of one quantity, looking at the variance of the other, that is where it become orthogonal, in a statistical sense, right. So you cannot extract any further information from it, you can only extract information when when two quantities are statistically co-related; if then un-co related, you cannot get any information of one about the other.

So now we are going to do the next one, that basically say is that another way of generating the minimum variance estimate is to generate the conditional mean; that is another way that is other than computing, if you can compute the mean estimate of the conditional distribution then then that also is a minimum variance estimate. So you will find I mean especially if you assume that if you assume that, all these you know all these noises are Guassian then you have very simple formulae for not simple but, when I mean a in very nice compact formulae, such that you can you can estimate the conditional mean; that the conditional mean of of of of Guassian sequence are are actually known and their formula are very nicely given. So once you know that the minimum variance estimate is the is the conditional mean, you have to just just estimate the conditional mean.

So it turns out that that also leaves to the calven filter. So that is another path way of leading to the calven filter equations, so we are we are looking at three ways but there are three different concepts, okay. So here what we are going to show is that, the minimum variance estimate is actually the conditional mean that is what we are going to see. How? So we are we are starting like this, that let x and y will be jointly distributed random variables and we are we are we just recapitulating what is conditional mean. Basically conditional, what what what are the conditional probability, probability density say? That suppose there are two random variables and somebody says that, okay you know that random variable y, because value two; if you have this the information that that the two random variables x and y, they are related in some way to their join probability densities. And you suddenly got some information that, one of the random variables took a particular value, now now obviously you can if you knew the join probability density, you can use this value.

You now calculate the now, I have have the fact that are given this information, what is the new conditional probability density function? It will get changed, right. So therefore probability density of x will not be the same as probability density of x given y. This we are we are also seen in our earlier, right. So and and they are related by the by the basical rules. So the so the conditional probability density function is actually written by this assuming of course that; p y y for the given value of y is is is not equal to zero. Now now now what is the conditional mean? We are you are interested in finding the conditional means. So what is the conditional mean? Conditional mean is the is the is just the mean, according to the conditional distribution.

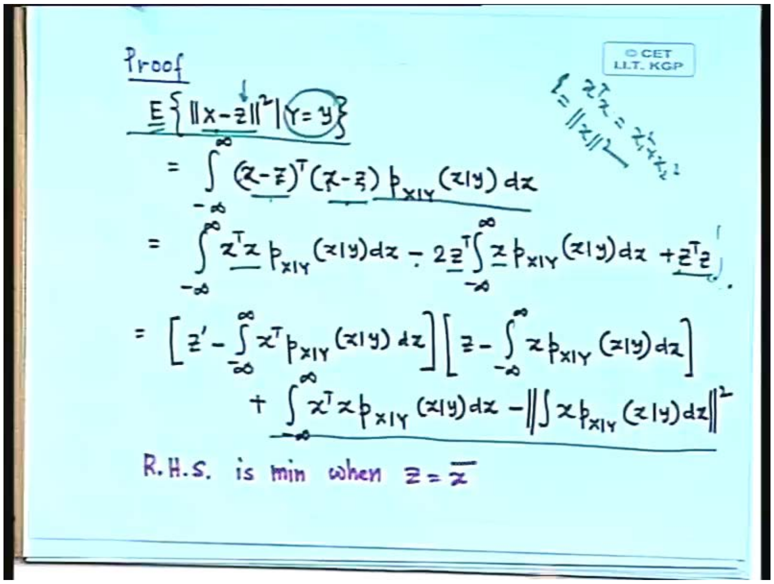
For any distribution, what is the mean? x effects d x, so only since it is a conditional mean; you have to put the conditional distribution, right. So in this distribution, the value of y is fixed; so it becomes a pure function of, once the value of y is given, frozen at a value that becomes a pure function of x, right. And if you want to get the mean, you have to do that normal normal mean calculation only with this conditional distribution now, that is the that is the conditional mean. Now, why should it be a m v e? Why should it be a minimum variance estimate? That is what is very nice about this result; that if that x hat is the m v e, how is the m v e defined? We are we are defining a m v e like this, that is x minus x hat whole square, this known means x minus, x one minus, x one hat square the basically the sum of the vector elements, this this is a vector norm; given that y equal to y will be less than this, for any other value of z.

Rather than taking x hat, if you took any other value of z; the right hand side will become more than the left hand side, now what is the right hand side? What is this expectation? This expectation is the is is over what? Expectation is are is is I mean always over something. So this expectation is over what? This expectation is over x; obviously because why, because y is frozen there is no question of expecting a y. y already taken taken a value small y, small y is the number. Once it is taken a value, now if suppose one one way is to choose x hat and another is to choose some other z, now this x hat and z are basically functions of y. Given a y, you will have to define some rule by which will you have to arrive either at x hat or at z.

So they are all based on value of y, if you are given one value of y; you will calculate one value of x hat, if you are given another value of y, you will calculate you will obviously calculate another value of x hat. So you have to have you have to you have to have rule by which you will you will able to calculate x hat or z from y, so so actually this are actually functions of y.

So but once once y is frozen, this this becomes a value and then you have to you to take the expectation of x; and then it will become so the minimum variance estimate will be, such a rule such that this will be minimum, okay. Now it turns out that the conditional estimation will give minimum variance estimate, why?

(Refer Slide Time: 43:32)



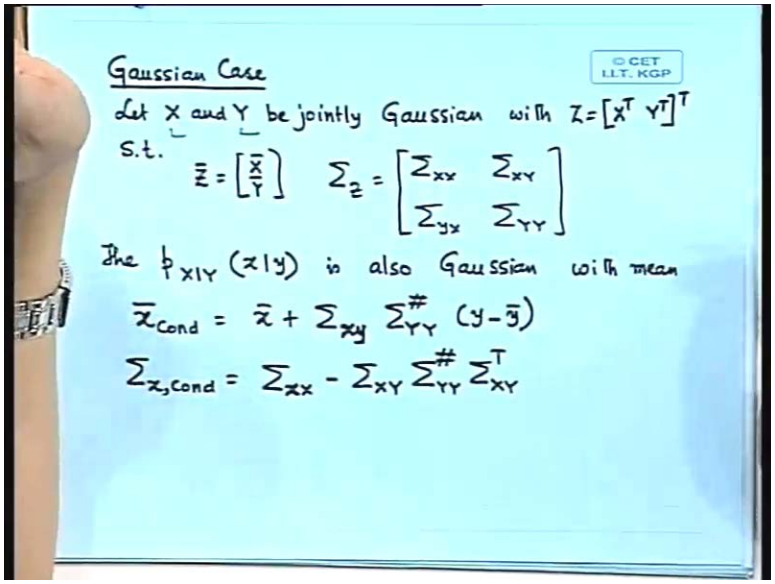
So that that proof is very simple. So so so so suppose this is my this is some estimate z, I am not necessarily saying that, this is a conditional mean, some estimate z you have calculated from y based on some rule. So obviously this norm equal to this transpose this, what is the norm of the vector, when a x transpose x is is defined as norm of x square. Norm means, this is equal to x1 square, plus x2 square, normal a clear norm, length of a vector, okay in in in dimensional space. So it is x minus x transpose, x minus z which is norm of x minus z; only thing is that you have to take its expectation. So we will take the expectation with respect to the conditional distribution, because y is the y.

So you can just just just open it out, this will be x transpose x, minus z transpose x, minus x minus x transpose z but plus z transpose z, this termed by term multiplication. So x transpose x then then z transpose x, actually z transpose x x x and x transpose z are same. So therefore two minus z and plus z transpose z, just just open down the multiplication. And obviously z transpose z z does not depend on x, it is a number; so therefore z transpose z will come out of the integral and you you will get minus infinity infinity, p x y, x y, d x which will be one, because of the area under the distribution is one.

So therefore I have not taken that integral, here that turns out to be one. And then this is just a again algebraic manipulation, you can you can again you see always if you want to prove minimization of function; you must write it as one part which which does not depend on the minimum, one part which which does not depend on the thing which you want to tune for minimizing and the another part which turns out to be a square. This you have done many many times in this question. So exactly I am doing this, see I have written it as something, once second, so this is basically; see this is something transpose something, so this is a whole square term this first term is whole square, now for making it whole square I have I have I mean I am I have to include some extra some extra terms which were not here. So I have now adjusted those terms and I have written it here.

So this plus this if you just open out and then multiply you get, this they are same things. Only advantage is that now this thing that not contain z, it is out of optimization, right. This thing is a square which means that, it will be either it will be either zero or positive. So the so the minimum value it can have zero because its a square, when will it have when will it be zero? When this is zero, when will it? So the minimum value of this is possible when when when this is zero, which means that z is the conditional mean this is the conditional mean. So only when z is, so we started with some arbitrary number z and we found that, this function will be minimize only when z is the conditional mean. So so this is another nice property because, this this property is also used for deriving the calven filter because it turns out that

(Refer Slide Time: 47:26)



That, if these are supposed to be this assumed to be jointly Gaussian then then there is a nice formula for this conditional mean; you know you know nice formula I will level for the gaussian case. So now you have to just you evaluate this sigma x x, sigma x y, sigma y y, etcetera for the case of the a b c d matrix. And then if you just start evaluating them and finally assume make these things, you will you will again arrive at the calven filter; that is another way but we are not deriving that.

So our in this class we just wanted to prove two major results of minimum variance estimators; one is that the innovations become white and the estimation error become orthogonal to the measurement, this is a very well-known algebraic property which is used in several minimum variance estimators and the second is that the conditional mean, if you if one can one can estimate that also that also gives a minimum variance estimate. So with this we will conclude today's class and then the next class, we hope to see let us see it all depends on how you can organize but we I mean a lot's of theory. So I say I said that we will we can we can actually show by some numerical examples, how how nicely the calven filter works; why it is so popular and we will have to see that, at least have some demonstration of that numerical demonstration which have to cannot be taken here. You have to go to another class, we will do that; when I want to show you that that in some practical case, that the calven filter in the very useful, thank you.