Dynamic Data Assimilation
Prof. S. Lakshmivarahan
School of Computer Science
Indian Institute of Technology, Madras

Lecture – 28
From Gauss to Kalman-Linear Minimum Variance Estimation

So, far we have reviewed principles of statistical estimation. We covered the basic properties of estimates unbiasedness consistency efficiency, then we looked into two types of estimation problems namely; one within the Fishers framework, wherever there is no prior one of the another is the Bayesian framework, where there is prior in both the case, there are observation.

So, the ultimate challenge to us is, if somebody gives you only observation, what do you do? If somebody gives you observation and some prior information or prior believe, what do you do in the both, the context we have developed least square based estimation methodology. We have talked about deterministic squares, we talked about statistically squares. We also talked about Bayesian least squares.

Please understand the primary theme that unifies the whole presentation of solving inverse problems, static deterministic, dynamic deterministic, static stochastic, dynamic stochastic, all the possible combination, the method that unifies them is the least square method. So, with that knowledge of statistical methods of estimation now, I am going to provide another thought process is within the framework of estimation theory, which is called sequential linear minimum variance estimation.

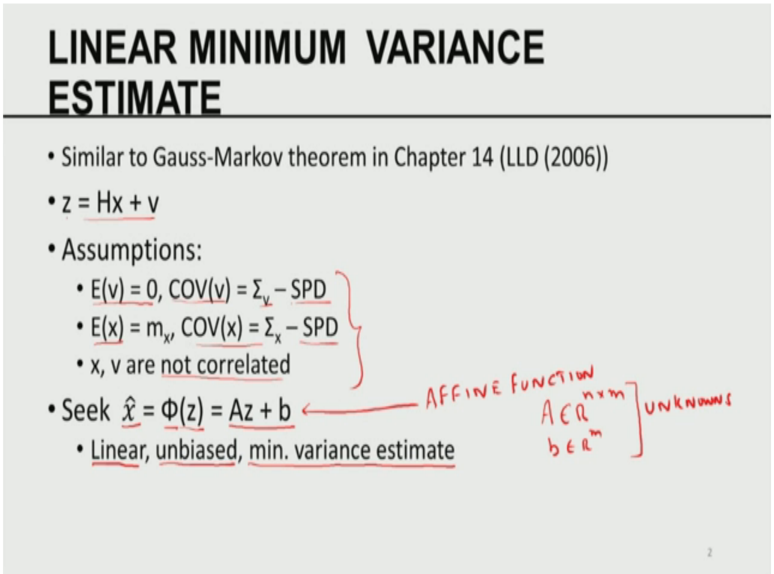
Please understand. So, far we insisted on simple least squares both within Bayesian and non Bayesian context. Now, these are slightly a different perspective for formulating the estimation problem, linear minimum variance estimation, the importance of this linear minimum variance estimation, the importance of sequential linear minimum variance estimation comes from the fact that Kalman originally used this framework to be able to derive the, now famous equation called form Kalman filter equation.

So, by doing this linear minimum variance estimation derivation, we are going to relate that to the Bayesian least square estimation, we are going to show the Bayesian least squares estimation that we discussed in the previous lecture and the linear minimum variance estimation, that we are going to do now are two different facets of the same problem Bible and we will build a bridge to go one to from one to the other, both ways thereby establishing a broader way to be able to look at estimation principles, that are used within the context of general dynamical data assimilation.

So, it is in that spirit, we call this module from Gauss to Kalman, why? Gauss least squares why? Kalman linear minimum variance estimation and. So, we will first develop the linear minimum variance estimation, per se starting from fundamentals then we will try to build the bridge between the Gaussian least squares and the linear minimum variance estimation there by establishing the dual aspects of the estimation problem, estimation methodology that is the goal of this chapter.

Once we do this we are in, we would have in principle completed, the fundamental principles that underlie the derivation of Kalman filter equations, this we are going to, do not within the context of conserving the dynamics, but a very general discussion and it is this generality of the discussion is very attractive and that is what we are going to pursue.

(Refer Slide Time: 04:46)



Linear minimum variance estimate is something similar to the Gauss Markov theorem. Gauss Markov theorem essentially relates to least square estimates. Let Z be equal to H x plus v, the assumptions are the noise remains 0. The covariance of the noise is sigma v, sigma v is SPD, X is random. I have a prior distribution, the prior has a mean mx the covariance of this is sigma x, that is SPD x and v are not correlated. So, these are the basic assumptions that one has to be able to formulate our estimation problem.

Again you can see there is a Bayesian undercurrent of Bayesian assumptions X is random, X has a prior Z is observation. So, prior gives some information observation, gives some other information. We have two pieces of information. Whenever you have two pieces of information, we would like to be able to combine it optimally, the Bayesian framework essentially concentrates on that, we have already seen that.

Now, we are going to be looking at another way of doing the estimation Bayesian like estimation, but within the framework of linearity an unbiasedness and minimum variance and that is the theme of this, first part of this discussion. So, I have an estimator which is phi, phi of Z, gives you the estimate. Estimate is x hat

In the Bayesian framework we concocted a cost function and then we had used that cost function defined what is called the Bayesian cost; and we tried to minimize the structure of the estimator that minimizes the Bayesian cost. Here, we are going to do totally differently, we are going to insist that this estimate x hat is linear in the observation. This x hat is going to be unbiased estimate, this x hat also simultaneously possess the minimum variance property.

Earlier, when we did the Bayesian estimation we showed that the gundy, the conditional Bayesian cost can be minimized by choosing the estimator to be the posterior conditional mean and after choosing that we then demonstrated that estimate is unbiased and it is also minimum variance. So, we first optimize the cost and then study the properties the optimal estimate as being opt as being unbiased and of minimum variance.

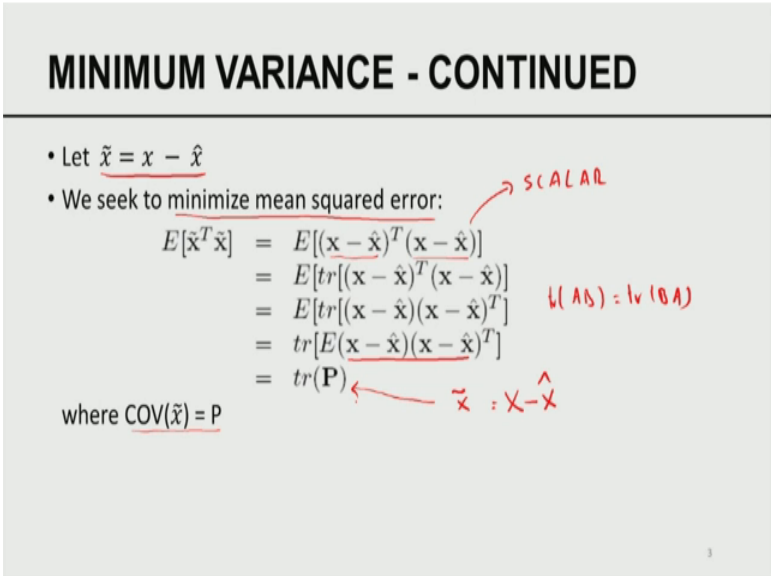
Here, we are going to start by explicitly talking about the structure of the estimate. So, let just, let the estimate x hat be dependent linearly on Z well x hat. So, phi of Z is equal to AZ plus b is such functions are not called linear, they are called affine in. So, if b is 0, is linear, if b is not 0, it is a linear term plus a constant term that is called affine function, but we called such a structure for simplicity linear. So, the estimate has a linear structure and I would like this estimate to be unbiased and minimum variance.

Now, please understand, we only know Z, what is A? Is A matrix? What is size of A? The left hands dimensional vector Z is m dimensional vector. So, A belongs to r n by m

B belongs to r m. So, in A and B are two unknowns. So, we are going to simply require let x hat be A Z plus B. There are two parameters A and B. I would like to be able to written A and B, I would like to determine A and B, such that you have properly chosen x hat becomes unbiased you have properly chosen x hat in addition to being unbiased, it is also minimum variance.

So, we are going to determine two parameters, such that the resulting estimate will satisfy two conditions, one being unbiased another, being minimum variance straight away.

(Refer Slide Time: 09:59)



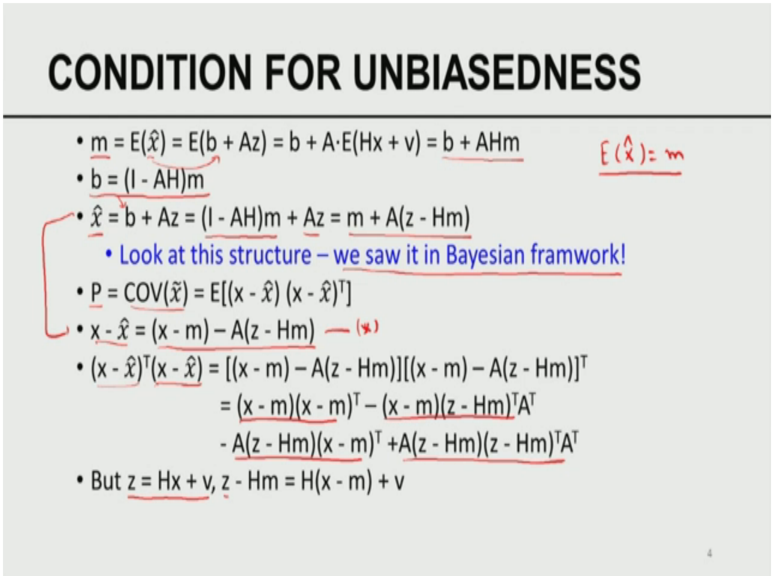
So, let x hat be the error in the estimate 2, we seek to minimize the mean square error. So, x hat transpose X is given by X minus x hat transpose X minus x hat X. This is an inner product that is a scalar.

Therefore the transpose of a scalars, it is the trace of a scalar, is itself. So, I am going to express this as a trace of that quantity, because a scalar is a one by one matrix I can think of a scalar and the trace of a scalar is itself. We also know trace of A B, the trace of A B is equal to trace of B A therefore, the third line comes from the second line expectation operators trace operator, they commute one can readily see.

So, expectation of the trace is simply the trace of the expectation of this term inside there is an outer product that is that is X tilde times X tilde transpose and s that is the covariance of the error we are going to call it P therefore, that is equal to trace of P where P is called the covariance of the error, it is this quantity the covariance of P is what we seek to minimize.

So, when we talk about minimum variance we are going to be talking about the following X is the unknown x hat is the estimate X minus x hat is equal to X tilde I would like to be able to minimize the sum of the variances of the individual components of X tilde, that quantity is related to trace of p. So, that is the quant that is the target that we are going to be working towards.

(Refer Slide Time: 12:26)



So, now having explained, what is that we are looking for. Now, we are going to go back to the expression. Now, please remember x hat is the estimate x hat is the estimate x hat is equal to B plus AC, let m be the mean of the estimator.

So, m is equal to E of x hat E of x hat is equal to the expectation of B plus AC, B is a constant, Z is equal to H x plus B. So, I can substitute Z is equal to H x plus V A being a constant H being a constant I can pull them out if X is m, because we have already made assumptions about, we have already made assumptions about the mean of X being m, that curve, the mean of X being m therefore, m which is the mean of the unknown, because unknown is a random vector, unknown as a mean m and such A mean by virtue of my estimate being linear in Z must be B plus A H times m.

So, m must be equal to B plus A H times m and if that is equal, it is unbiased. So, this is where the unbiasedness comes in. What is the unbiasedness condition? E of x hat is equal to m. M is the unknown I am trying to estimate, m is the mean, we are going to estimate x hat is the estimator. So, unbiasedness essentially requires E of x hat is equal to m. So, that is what we start with and. So, unbiasedness condition leads to m must be equal to B plus A H times m. So, that essentially tells you B must be equal to I minus A H times m.

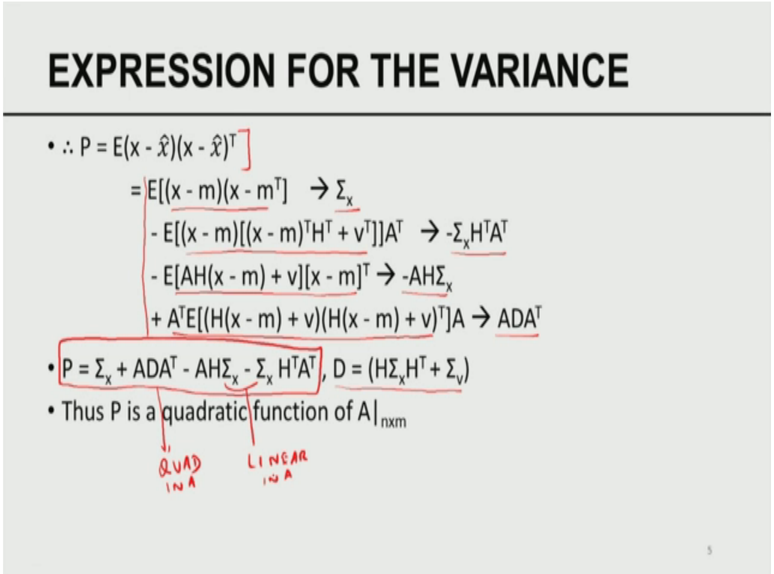
If I substitute this value of B in here. I get a new structure for x hat X hat is equal to I minus A H times m plus A Z, which can also be written as m plus A times Z minus H r look at the structure. This looks truck, we look at the structure, we saw the structure of the Bayesian framework already; that means, the estimate is equal to the prior plus A times Z minus y.

So, Z minus is the innovation A is the weight. So, I had two unknowns, I have the element, I have decided what must be the structure of B in order that the estimate is unbiased by substituting the condition for unbiasedness. The estimate done by as the structure the unbiased estimate must be given by this m plus A times Z minus H m.

And therefore, the covariance of P, the covariance P of x hat is equal to X minus x hat times X minus x hat transpose the expected value that of X minus. So, I know X structure of X hat. So, I can compute explicitly. What is X minus X hat? So, if I substitute to this structure in here, I get the error structure to be this. So, that is the expression for the estimation error it is this matrix whose expected value is P. So, let us compute that matrix explicitly by substituting star in here, we have two terms you multiply the two terms and simplify you get as a result there are four terms that affect the inner product x minus X x hat transpose x minus x hat.

Again Z is equal to H x plus V. So, Z minus H m is equal to H of X minus m plus V, again there are lots of algebra in here. So, if I substitute this expression in here for Z minus.

(Refer Slide Time: 17:10)



The P is equal to expected value of the product X minus x hat times X minus x hat there is m I am sorry, I am sorry, there is a parenthesis that is missing. So, I can substitute all the four terms in here, in the, I can substitute all the four terms from the previous page in here. So, I have now sum of four expected values, in here sum of four expected values in here the, first term you can readily see is sigma X, the second term you can readily see is minus sigma X times H transpose, A transpose the third term is minus eight times, H times sigma X the fourth term, becomes ad A transpose where for simplicity the, in for simplicity notation d denotes H times sigma X plus H transpose plus sigma V, yes there is a kind of simplification.

But I think it is a worthwhile exercise. So, if you did that I ultimately get what I want. This is the structure of the covariance of the estimate. Now, look at the structure, this structure has sigma X, which is known. It has H, which is known. It has d. Let us look at the structure of d d depends on H sigma X and sigma V everything is known. So, in this every quantity, other than A is known. A is a matrix, now you can also see one more in this term. A is quadratic, in A quadratic, in A, in these two terms. It is linear in A.

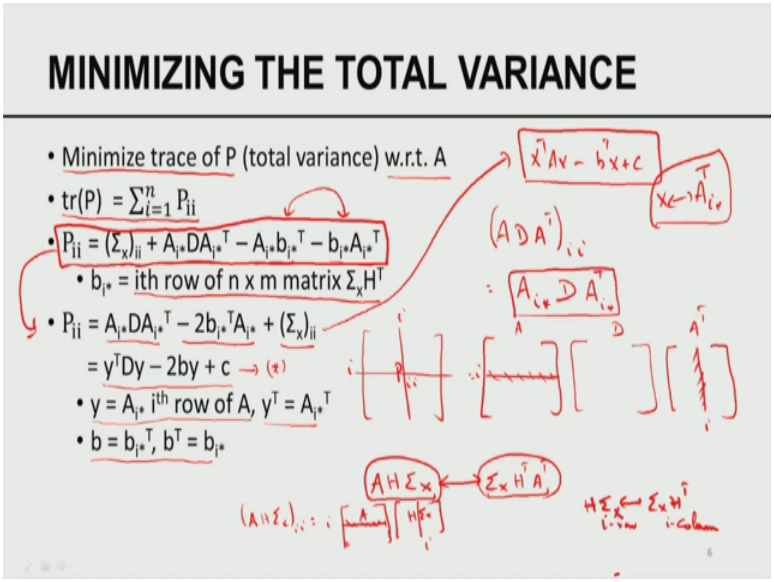
So, you can think of P is some form of a quadratic function in the elements of the matrix A is unknown, let us go back. So, what is that, we want to be able to get their meaning A linear, minimum variance estimation, we also wanted to have a linear. So, that essentially said my estimator, must be a linear function of the observation. So, we can cut that, let x hat be A A Z plus b. We forced unbiasedness that gave raise to that, gave raise to condition and B in terms of A which we substituted.

We carried through the analysis we computed the expression for the covariance of the estimate the covariance, of the estimate it is a quadratic function, in the elements of the A matrix, which is yet to be decided. So, we still have not decided what A should be. So, every parameter is the form of a control, we have it is like a knob, I can change to enable, what I want to do? We use B to force unbiasedness. Please, go back. I want my estimate to be linear unbiased linearity by assumption by structure A is Z plus B.

Unbiasedness B has been expressed in terms of any other quantity and we have eliminated B we have done. So, the only one remaining condition is minimum variance to be able to attain minimum variance, what are we going to be doing? We are going to be using the tool, the unknown, the elements of the matrix A. So, I am going to fine tune the element of the matrix. Here such that the trace of P becomes minimum y trace of P. P is a matrix the diagonal elements in the matrix are variances.

So, the total variance sum of all the variances in the components of the estimate is equal to the trace of P.

(Refer Slide Time: 21:37)



So, what is the rest of the problem? The rest of the problem is a mathematical problem, of trying to minimize the trace of P, which is the total radiance, but with respect to what? With respect to yet to be decided free parameter A.

Now, what is the trace of A? Trace of A is essentially sum of all the diagonal elements, P I I. Now, please remember A has n square elements. So, we are going to be minimizing a scalar quantity called trace of P, that depends on the elements of a there are n square of them. So, in principle, we have to do a minimization over n square variables that is the computational problem of interest, in here to simplify this compute.

So, what is that one can do? One can now, that we have expression for the trace, we can compute the derivative of the trace with respect to each element compute the condition and collate all the conditions and build the matrix A, that is one way, the another way would be an easier way, would be to determine A, not element by element, but by column, by column or row by row, that way we can simplify, the structure of this minimization problem there are several different ways, in which one can minimize, I am going to talk about one particular form of minimization.

Please recall, we have already learned how to minimize a quadratic function. We are, we have gained a good experience in the minimization of quadratic functions. So, given our understanding and knowledge about the minimization of quadratic forms, I am going to use the tool of minimization of quadratic forms to be able to determine, the elements with the optimal A to that end, I am now going to consider an i th element i th element of P.

Now, please understand this expression in slide, five gives you the entire matrix. So, if I want to consider the ith element. Please remember, P is the sum of 1 2 3 4 matrices. So, the i th element of P, which is a P i th diagonal element of P, which is P i, is the sum of the corresponding i th element in the every matrix on the right hand side, that is the basic idea here. So, pii is equal to i th element of sigma X. Now, let us look at the second. Let us look, at the second term, the second term is ad a transpose.

So, let us look at this. Now, the second term is Ad I am sorry, the second term is Ad a transpose, I would like to be able to consider the i th element of i th element of this a little reflection reveals this, i th element of this matrix is given by the product of the i th row of a times d times the transpose of the i th of A.

Please understand A is A matrix, d is A matrix and a transpose is the matrix. This is A d, a transpose, this is going to be given by the product matrix. So, I am interested in the i. This is P i i in order to get this P i I. What is that I do, I take the i th row of a i take the i th, the column of a transpose.

So, I do a quadratic form, which is given by this expression therefore, P i i is simply the sum of i i th element of sigma X, the quadratic form with the i th row of a likewise, you can also readily see for the third term, for the third term and the fourth term. We can express them as A i star A i star, look at this now, A i star bi star. Suppose, b i star a i star transpose where b i star is the i th row of the n by m matrix sigma X Ht.

Let me probably, I am, I would like to spend a couple of minutes on that, please let us go back. So, the third term is I am sorry, the third term is the third, it is time sigma X times sigma X. So, let us consider the third term times sigma X. So, I would like to be able to consider the matrix as A b where b is equal to now, I would like to do the other way, I am sorry once again.

So, let us consider the term sigma X H transpose, a transpose. So, let us look at these two terms and I am sorry, let us look at the two terms. There are two terms, I will erase this part. So, if you go back, I have two terms sigma X A H. So, let me write that down in here sigma X plus one term and if I go back there is sigma X H transpose a transpose sigma X , H transpose, A transpose.

In here this is not X, this is H sorry. This is H sigma X symmetric. So, you can see these two matrices are transposes of each other therefore, if I consider the i th element of this matrix, I can infer the i th element of that matrix very easily, which is the transpose of this. How do I compute the i th element of this matrix? So, this is A, let this matrix be, let this matrix be , H sigma X. So, this is A. So, in order to control the i th element, what is that, I now need to do?

I need to take the i th row of this, I need to control the j th column of that j th column, i th row that is that is essentially it or if I want to have the i th element, I need to consider both i and j are equal. So, so that is the basic thing. So, in here I need to be able to consider this i, this is i therefore, i th element therefore, i th element of sigma X is given by the product of i th row and the i th column of the product H sigma X the i th column of H sigma X is the.

So, sigma. So, let us look at this now, H sigma X is one matrix, sigma X H transpose is another matrix, both come in here, we also know, these two are in transpose of each other. So, what does it mean the i th row of this is the i th column of that i the row i, i row becomes the i th column of this therefore, we can readily convince ourselves A i star bi star transpose minus bi star Ai star transpose are the i th element of the matrix A H sigma X and sigma X A transpose H therefore, the entire Pi is given by this expression, I want you to understand that this is how the extract the i th element of the matrix P. It is a sum of all the P is, that gives you the transpose.

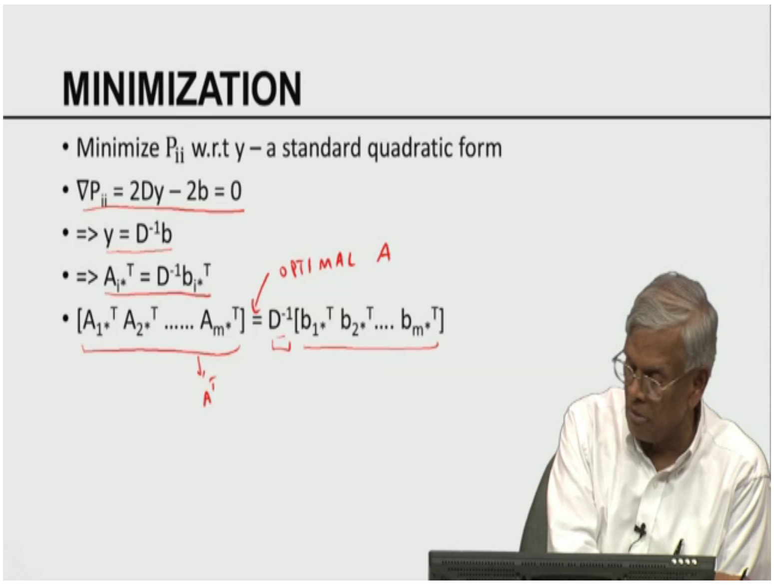
So, what is the idea here, if I want to minimize the sum, it is enough, if I minimize the individual term. So, if I minimize the individual terms, I can, in other words, we are going to minimize, the individual terms P i i to be able to minimize, the trace therefore, this P i in here, can be written as a quadratic form, please remember Ai star is A, is a row vector Ai star transpose is the column vector d is a matrix, this is minus 2 bi transpose Ai and this is a constant sigma Xii, where is the two comes, from these two terms are scalars. They are both equal, they are transposes of each other, they are scalars therefore, these two terms are combined two times bi transpose Ai star plus sigma X i i.

So, how does this looks like this term essentially looks like X transpose AX minus b transpose, X plus C that is the quadratic function X is replaced by. So, X is used in place of A i star. So, you can see the relation between X and A i star, the A i star is the i th row . So, X is a vector. In this case X is equal to Ai star transpose, because X is a column vector, A i star is the row vector.

So, with this association between X and Ai star, you can readily see, this is the quadratic function. So, I would like to be able to rewrite this in the form of a quadratic function, which is y transpose d y minus 2 b y plus C y i is the A i th row of that. So, y transpose is the column vector. So, y transpose d y is what we are concerned with; we also did know the associations.

So, this one is A, whose structure is very well known to us. Now, I am going to minimize P i i, the expression on the right hand side given by star with respect to y.

(Refer Slide Time: 33:51)

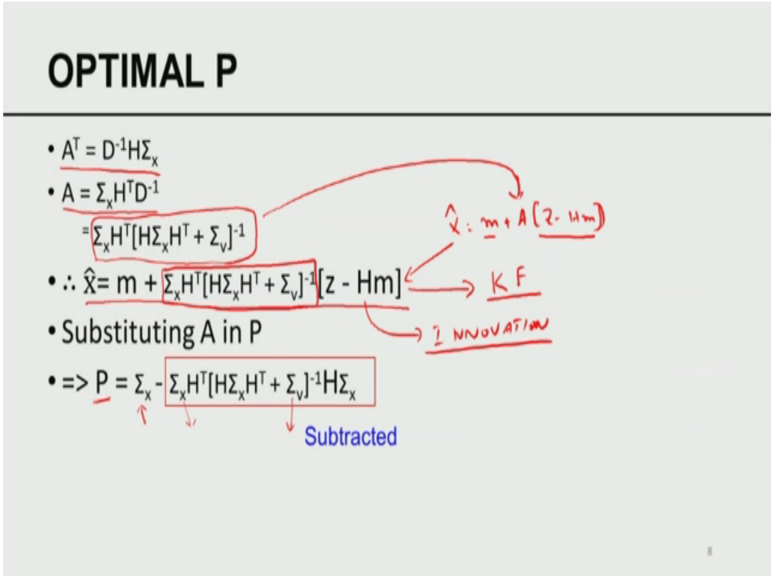


So, minimize Pii with respect to y that is a standard quadratic form, if you compute the gradient and the gradient in this case is given by 2 dy minus 2 b therefore, the optimal y is given by d inverse b i change the notation back to A. So, this is going to be equal to Ai star transpose is it going to be equal to d inverse bi star and please remember b is related to sigma X H.

Therefore, I can now construct a matrix A 1 star transpose A 2 star transpose, A m star transpose each of these have d inverse, as a common factor common factor and I have V included all the b s therefore, this matrix is equivalent to product of this matrix times, that matrix this expression provides optimal value of a optimal A. So, A 1 star is the first row of A A 2 star is the second row of A A m star is the m th row of A.

Therefore, if I consider the transposes, this matrix now becomes a transpose. So, instead of expressing A, we are trying to express it in terms of a transpose, that is a convenient thing, therfore the previous expression, look at this now the previous expression at the bottom of slide 7.

(Refer Slide Time: 35:39)



Now, can be succinctly denoted by A inverse. I am sorry, A transpose is equal to d inverse H X now, A i take the transpose of both sides, remember sigma X is a symmetric matrix d is a symmetric matrix. So, A is given by this d, we already know d is a symbol that represents sum of two.

The sum of two matrices therefore, A, the optimal A takes this particular structure, that is very important. So, we have already used the quantity minimization principle, to be able to determine A, that minimizes the trace of A. So, let me summarize, where we are right now, we started with a linear structure, we required it to be unbiased, we eliminated b, then we looked at the resulting structure of the, for the estimation error. We compute an expression for the covariance of the estimation error, then we computed an expression for the total sum of the variances of the individual components of the forecast error, which relates to a case of P. P is the covariance of the forecast error.

The covariance P is a quadric function in the elements of A, there are several different ways to do the minimization problem, we chose a particular minimization problem the reason for choosing what we did, because we already know how to minimize quadratic functions. So, we fell back on what we know and know very well, we know how to minimize quadratic forms therefore, we converted the problem of minimizing P i, we converted the problem minimize in the trace of P, trace of P is the sum of all the Pis.

So, what is that, we did? We minimize the individual terms in the summation. If you minimize the individual terms in the summation, the total variance is going to be minimum that is the line of arguments, the minimum of the individual, elements can be obtained by appropriate choice of A. So, the choice of A that relates the minimum of the i th element P i i relates to a row of Aby deciding each row of A for each element P i i. We collectively got all the elements of A.

So, the optimal A is given by sigma XX transpose d inverse, once we have that, please go back and remember what is the optimal structure, we already have the optimal structure, is equal to m plus A times H minus, I am sorry Z, this is Z Z minus. So, Z minus is known m is known. So, we have to substitute A. We substituted this value of A, in here that led to this expression in here.

So, the optimal estimate that is linear unbiased and minimum variance is given by x hat is equal to m plus A matrix, which is a weight matrix and the innovation, this is the innovation. Now, not only we have gotten the estimate, we already know the structure of the convenience of the estimate which is P, if you go back to the previous slide.

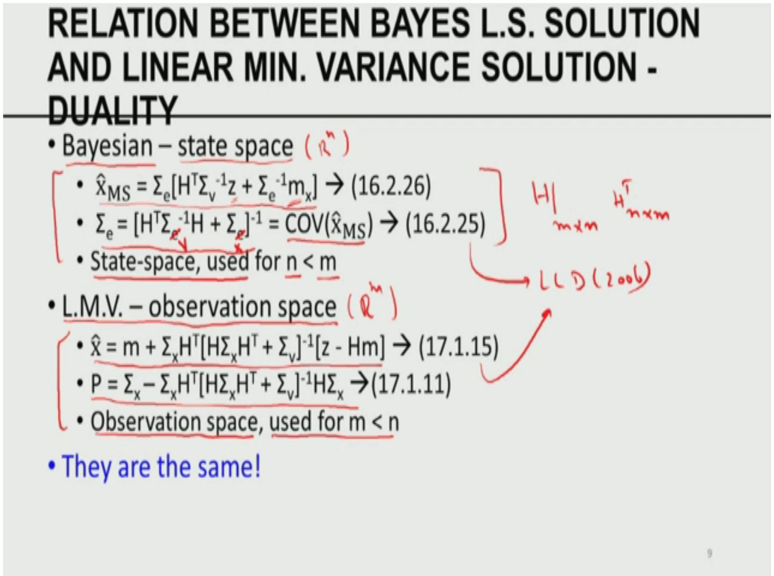
In slide 5, we have an expression for the covariance of P and there everything is known except A. Now, we have optimally, we have determined the optimal value of A. So, if you substitute the optimal value of A, in here we also get the covariance, the covariance is the optimal estimate substituting A in P and simplifying that is a good bit of algebra involved, there you get the matrix P, which is the covariance of the linear unbiased estimate.

The minimum covariance, I should not say minimum covariance, the covariance and the estimate, where the total variance is minimum, the covariance structure is it takes, this particular form. Now, look at this. Now, there is, there are two terms one is sigma X. What is it? Sigma X is the variance of the prior. I am subtracting from that a quantity. So, you can think of it, the posterior covariance is less than the prior covariance, if the posterior covariance is less than the prior covariance, means what by combining, the prior of the posterior by combining the prior on the observation, I have tried to reduce the variance of the posterior variance, of the posterior.

Now, look at this, now sigma X is the covariance of the prior sigma V is the covariance of the noise. So, this whole term is a symmetric matrix. It definitely, it is a symmetric positive, definite matrix, I am subtracting a symmetric positive definite matrix from sigma X. So, P the trace of P is less than that trace of sigma X. So, by lessening the variance, if I am going to improve the quality of the posterior mean. So, I have prior mean, I have observation, I have the posterior mean, posterior mean has a smaller variance of the prior mean and that is the result of combining two pieces of information prior and the observation.

The structure that is given in here is called the Kalman filter structure; this is the structure of the Kalman filter, which has originally derived by Kalman in 1960. So, it is an, the derivation that, if we had gone through, is the very important derivation that leads to the fundamental result in Kalman filtering techniques.

(Refer Slide Time: 42:24)



Now, I would like to build a relation between the Bayesian least square solution and the linear minimum variance solution and introduce a sense of duality between the two, to that end. I am going to now recall, what we have done the Bayesian approach, the Bayesian approach is supposed to do the calculation in what is called the state space rn. So, in Bayesian approach, you may recall from our previous analysis, this is the structure of the optimal estimate.

The optimal estimate is the linear combination observation Z and the prior mean and is given in this particular form sigma E is the matrix, that weights the sum and sigma E is given by I am sorry, this must be sigma X and this must be H transpose sigma E n by n .

This must be, I am sorry one second, I would like to correct an error here. So, this must be equal to sigma X and this must be equal to sigma V. Let me think about it, that is sigma V, no the other way.

I am sorry, , H is let me make a little calculations, in here H is m by n H transpose is n by m. So, if I am going to multiply this. So, this must be sigma V, this must be sigma X. This must be sigma X, sorry for the error. So, sigma V you can readily see from the previous calculations, we have already done. So, sigma X is the prior covariance sigma V is the observation covariance. So, that is the covariance of X, the mean square estimate coming from the Bayesian.

So, this is the estimate and it is covariance, this formulation is called state space formulation. When n is less than m it is useful to do the calculation, to the state space, which is rn, because n is less than m. So, computations in the state space are smaller of the two spaces when compared to, when you consider the observation space and the state space. So, this is the summary of the Bayesian least square solution 16.26 16.25. These are all expressions in chapter 16 of our textbook, which is the LLD 2006.

Now, linear minimum variance estimation is supposed to be working, the observation space, which is the mn, the observation space formulation is used when m is less than n in the case of linear minimum variance estimation, the structure of the estimate is given by this. We just saw the covariance of the estimate is given by this. This is again given in LLD.

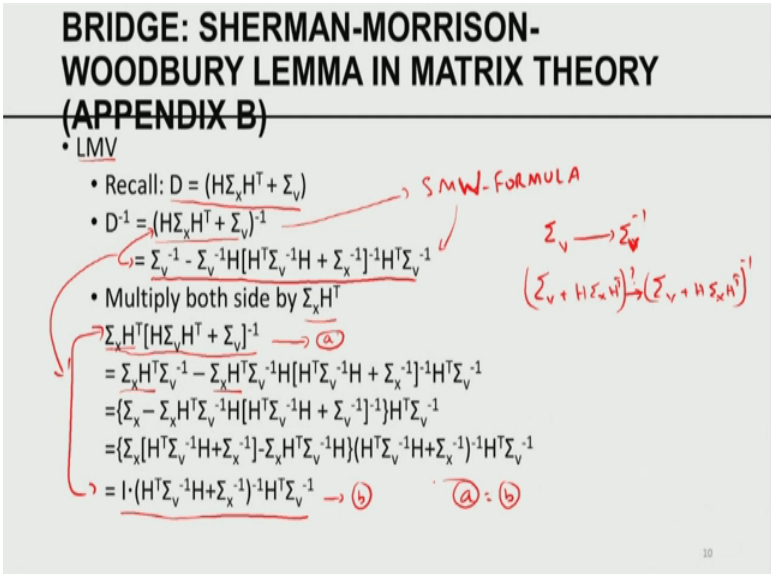
So, now you can see we have two types of results, one coming from state space formulation within their Bayesian least square set up, another one is the observation space formalism that comes from linear minimum variance estimation. At the outside it looks as though they are different, but the important part of the result is that they are indeed the same.

So, there are two different versions of the estimation problem, the same estimation problem, the results are given in two different forms, which at the first side looks very different. Now, we are going to show that one can convert one result into the other by invoking to a very simple matrix identity which is called the Sherman Morrison Woodbury formula. We have already used the Sherman. Sherman Morrison Woodbury formula in the context of recursive least squares, then we did static deterministic problems.

The same to Sherman Morrison Woodbury formula is the result from matrix theory, which are developed during the early, mid 30s and 40s is becomes very handy to see the relation between these two formalisms, the bridge between these two states. So, I would like to reinstate, one is the state spread formalism, another is the observation space formalism, one coming from the Bayesian, another coming from the linear minimum variance estimation.

So, Kalman derived using the linear minimum variance estimation, people who came after him, also derived the Bayesian formulation and then showed the Bayesian formulation and the linear minimum formulation are one of the same, they are dual to each other in one case, you do the computations observation space, another case you could do the computational state space by proving the equivalence between the two. It provides us a lot of freedom to do the computation in a space, which is smaller of the two, whichever is smaller I will adapt this formulation rather than that formulation that is the basic idea.

(Refer Slide Time: 48:18)



So, the bridge based on Sherman Morrison Woodbury formula is the formula in matrix theory, we have given in the appendix b to our book, I have also talked about extensively about the Sherman Morrison Woodbury formula in our module on matrix, matrix analysis, matrix theory, facts from matrix theory. So, cons start; let us start with the linear minimum variance analysis. We recall D is given by this, D inverse is given by this if I apply Sherman Morrison Woodbury formula Shermann Morrison Woodbury formula to this.

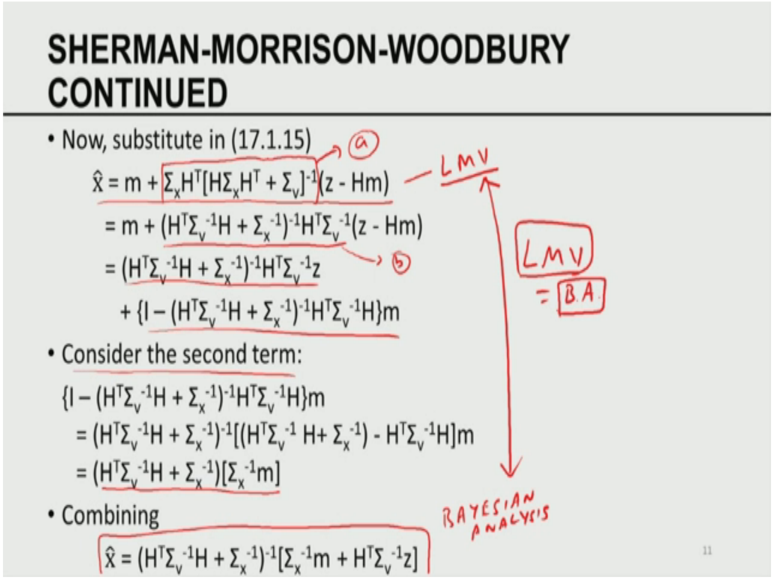
We essentially get this Shermann Morrison Woodbury formula I am not going to , quote the formula, it is already given in the module on matrices. So, what does it essentially say the Shermann Morrison Woodbury formulas says the following, if I have sigma V, I may know sigma V inverse. Suppose, I update sigma V by adding A matrix which is H sigma X H transpose.

How do I compute the inverse of sigma V plus H sigma X H sigma trans inverse. So, given this the question is how to compute this and that is given by the well known Sherman Woodbury formula, if you use that formula you will essentially get d to be this.

Now, I am going to do a little bit of a jugglery, multiply both sides by sigma X H transpose. Please understand, these two are equal, these two are equal, I multiply both the expressions on the left by this and then do a sequence of simplifications. We arrive at this formula. So, what does that tell you sigma X H transpose, the inverse in the sum is equal to the inverse of this sum times that. So, this is a matrix identity that gets out, that is a result of coming, there is a result coming from application of Shermann Morrison Woodbury formula. So, there is a lot of matrix algebra involved.

But the general steps, are should be very clear by now, yes there is a lot of checking to be done. I hope you will take a few minutes to be able to check all the details. So, in view of this fact that this matrix is equivalent to this matrix let us call this, now I am sorry. Let us call this matrix; sorry I have to let us call that matrix as A let us call this matrix as b. So, what is that, we have accomplished by using Shermann Morrison Woodbury formula, we have essentially shown the matrix A is equal to the matrix b that is the first step.

(Refer Slide Time: 51:44)



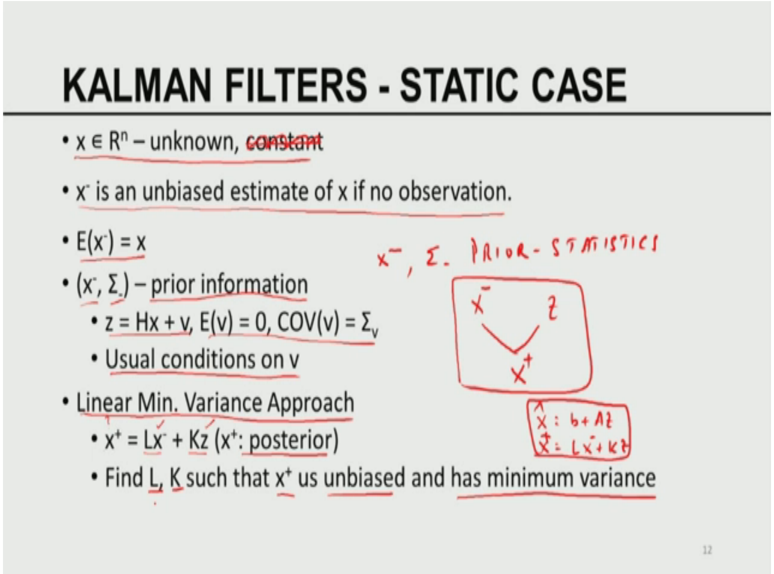
Now, I am going to import that relation, now look at this, now the structure where do I get this structure, this structure comes from the linear minimum variance estimation from Gauss to Kalman, we just talked about few minutes ago. In this I have a matrix that takes this form, this form is the matrix A. I am going to replace that matrix by matrix b, which is A, is equal to b that I derived in the last page. We know A is equal to b. So, we have earned the right to be able to replace A by b.

Now, if I do the multiplication and simplify I get that to be the sum of I get this to be the sum of two terms, this term and this term. Again I am going to do simplification, consider the second term by A sequence of simplification. It can be shown the second term, reduces to this by combining the simplification from here and substituting this in here. We get the overall structure to be X transpose. I am sorry x hat is equal to given by this.

Now, if you look back this is exactly the structure given by this, is exactly the structure given by the Bayesian sorry, this is exactly the structure given by the Bayesian analysis. So, what is that, we have done, we started with the linear minimum variance estimation technique linear minimum variance formula, for the optimal estimate, from there we have derived the Bayesian analysis by applying a matrix identity that arises out of the application of Shermann Morrison Woodbury formula by retracing this, we can readily see these two formulas are equal.

Therefore linear minimum variance analysis is equivalent to the Bayesian analysis linear minimum. Therefore, there are two equivalent ways of looking at Bayesian did assimilation or Bayesian way of estimation, one through linear minimum variance another through the classical Bayesian itself.

(Refer Slide Time: 54:23)



Now, I am going to reformulate it in the form of a Kalman filter static case, why the Kalman filter generally talked about within the kind of dynamics, but Kalman filter application is applied at a stage, where we are going to apply it at a given time. So, it is in application, if you understand the static case, we should be able to apply it repeatedly in time to get the Kalman filter equations within the context of dynamic models.

So, I am now going to quickly review some of the basic ideas. Let X be an unknown, but constant. So, x hat be an unbiased estimate of X with no observation; that means, when there is no observation that is the prior, the prior is given by minus X minus means prior X plus means posterior, the prior has a mean X. So, the prior information consists of the mean and the covariance the, I should say this is constant, I am sorry X is the unknown, that is all. What it is? X is the unknown, x hat is the unbiased estimate of X, when there is no observation Ef x hat is X, X minus is the prior covariance.

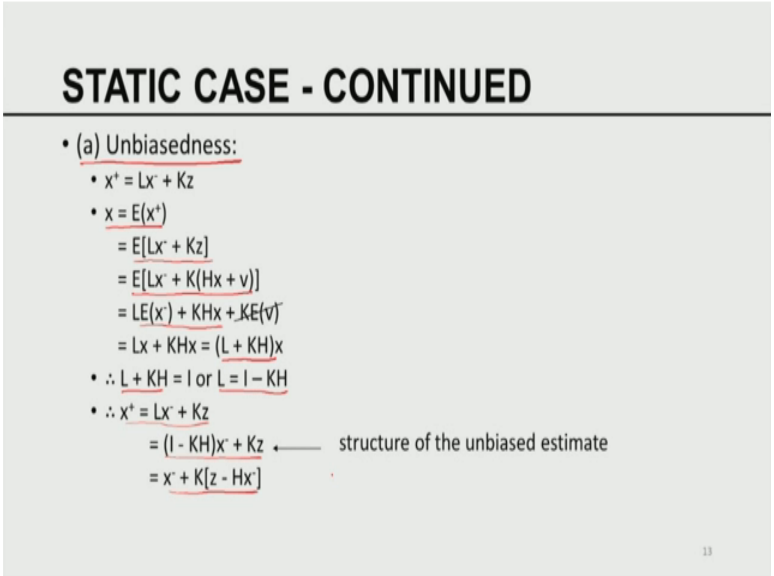
So, minus refers to the prior minus refers to X minus sigma minus r prior information or prior statistics. So, I am given the observation again you have the standard assumptions, the usual conditions on V, holds good namely the Xn and V are uncorrelated the linear minimum variance approach. So, I am now going to talk about a posterior estimate X plus the posterior estimate, is a linear function of the prior and the linear function of the function of the observation.

So, i is a matrix, k is a matrix. I would like to be able to consider the posterior estimate X L X plus is the posterior estimate. So, I am given two pieces of information X minus and Z, I would like to be able to combine them to 2 to the X plus sorry, I would like to able to combine them to get the X plus to X plus, I am going to again remain within their linear minimum variance estimation as done by Kalman. So, I am going to insist at X plus, is a linear function of the prior, on the observation i is equal to X bar plus K Z. So, that is the posterior structure.

Now, prior is given, observations are given, I would like to be able to find X plus, there are two unknowns ln K, I would like to be able to use the Ln K, such that X plus is an unbiased estimate and also it has a minimum variance. So, you can see I am just, I am trying to repeat what I did in the linear minimum variance estimation, but this is the derivation that Kalman had given originally in his paper.

So, what is the difference in here? Earlier, we assumed the estimate is b plus AX. So, earlier, we assumed x hat is equal to b plus AZ. Now, we are assuming X plus which is the posterior, which is equal to L X minus plus K Z. These two are equivalent, these two in some sense are equivalent and. So, we are now going to repeat the derivation. Our aim is to be able to guarantee express as unbiased is also minimum variance; I have to impose two conditions. The two conditions are imposed by selecting the two matrices L and K.

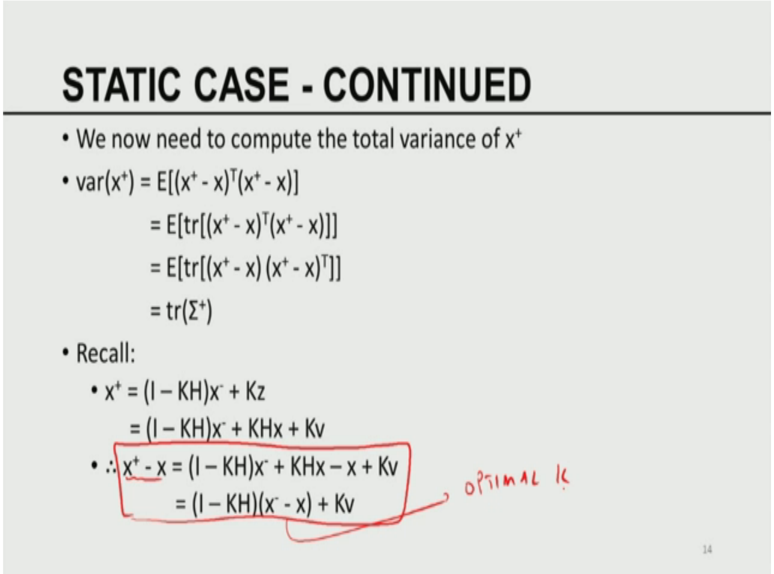
(Refer Slide Time: 59:02)



Therefore let us look at the condition for unbiasedness of this X plus is equal to L of X minus plus K, Z X must be equal to E of X plus, E of X plus is equal to E of L X minus plus KZ, Z is equal to H x plus V. You substitute that you get, here you by doing the simplification, since the expected value of E is 0 that term goes away the prior is unbiased. So, E of X minus is X k x X is X in this case therefore, I get the quantity, the X must be equal to L plus K H X. So, this is essentially tells you L plus KH, X must be identity or L must be equal to i minus K H.

So, one of the two matrices are, is expressed in terms of the other and H therefore, the structure now becomes this, which can be written like this. So, that is the structure the unbiased estimate. So, you can see we are running very parallel very much parallel to the linear minimum variance estimation, in this particular context except that the structure has been taken to be this in here; I would like to remind ourselves that.

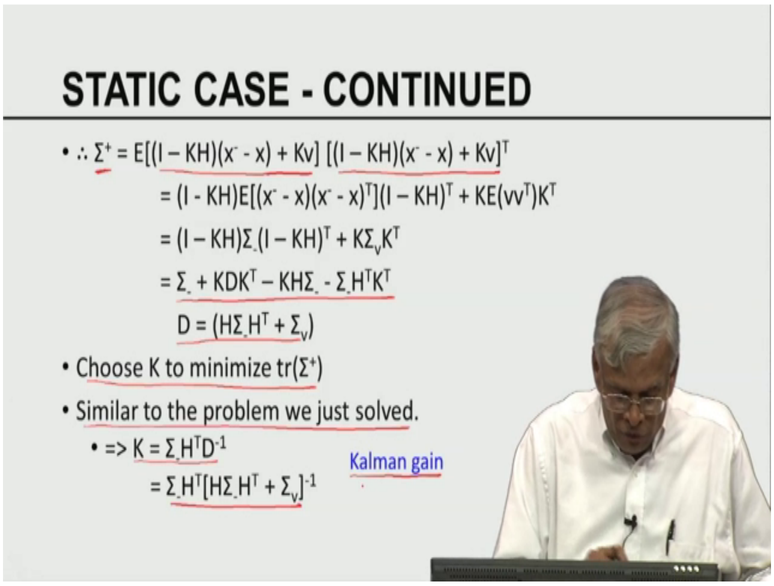
(Refer Slide Time: 60:35)



We are looking at the structure of the estimate that comes from the unbiasedness condition and the linear structure linear in both Z as well as in X prior.

So, that is what the resulting structure is. So, this is the posterior error in the estimation if I have the posterior error in the estimate.

(Refer Slide Time: 60:56)

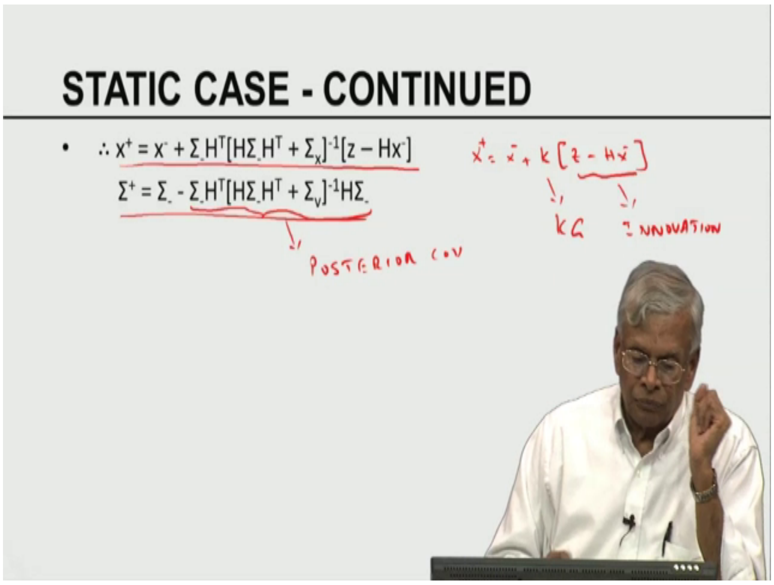


I can compute the posterior error covariance sorry, the posterior error covariance. Again, we are talking about the posterior covariance; this is the expression for the posterior error. It is transpose, you multiply, you simplify the whole expression you get this and you get d.

You can readily see the parallel between what we did and what we are trying to do. I would like to be able to choose K, such that the trace of sigma X X plus s is minimum, the problem is similar to what we just solved. So, the K that minimizes the trace of sigma plus is given by that this K takes this particular form and this K has a special name is called Kalman gain. In order of Kalman, who derived the filter for the first time in 1960.

So, we substitute this K please, go back a, this is the structure. It is into the structure, we are going to, sorry it is into the structure we are going to substitute the value of optimal K, if you substitute the value of optimal K in here.

(Refer Slide Time: 62:24)



And simplify, you get the expression for X plus, which is given here and you also remember the covariance sigma plus has K in it. So, I could also substitute the value of the covariance, in this expression, because I already know. So, I am going to submit this back in here.

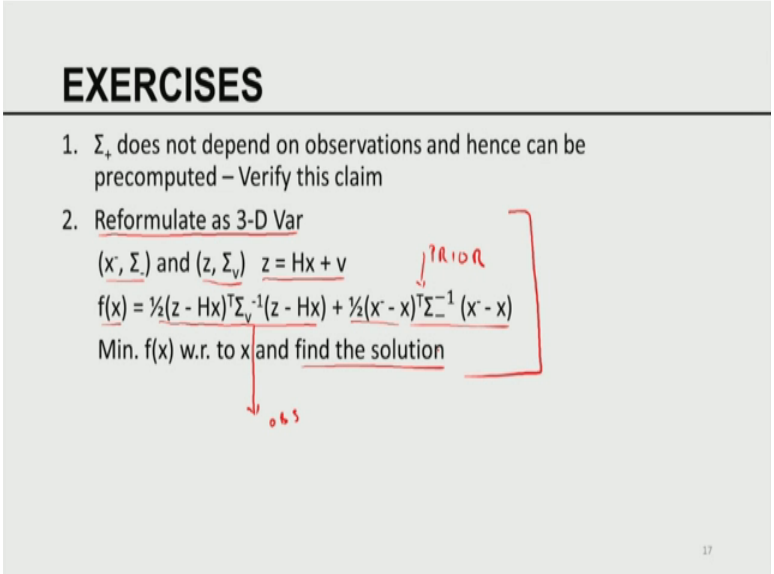
If I did this and simplify, I get the posterior being given by that and the posterior covariance is given by the posterior covariance is given by this expression and in the traditional literature, this could be written as this is generally written as X is equal to X minus plus K times Z minus H X minus K is called the Kalman gain, Kalman gain this is called the innovation. This is called the innovation, this is called the posterior covariance, this is the posterior covariance.

Again you can see the posterior covariance subtracts a matrix from the prior covariance. So, the posterior covariance is smaller than the prior; that means, after combining the prior and the and the new information, I have reduced the uncertainty in my estimate that is why posterior estimate is better and this is the reason why posterior mean is optimal, within the context of Bayes, Bayesian framework.

So, we have come to the end of discussion from guass to kalman linear minimum variance estimation. So, in the previous lecture, we talked about the structure of the Bayesian analysis and Bayesian optimal decision and as well as the resulting value of the covariance from the Bayesian structure, here we developed the theory of minimum linear, minimum variance estimation, we talked about the intrinsic relation between linear minimum variance estimation and Bayesian estimation, we built the bridge, the bridge depends on a result from matrix theory, which was independently developed by mathematicians in the 30s and 40s and that is called the Sherman Marrison Woodbury formula.

So, by using an already existing formula, we were able to build the bridge between the state space formulation and the observation space formulation and that essentially shows that Bayesian method can be interpreted in one of two ways either within the classical Bayesian or within the linear minimum, very minimum variance framework and we have now by introducing these two, we have more choices in terms of picking, which one is better from a computational perspective, we pick one or the other depending on whether m or n smaller of the two the.

(Refer Slide Time: 65:41)



So, with that I think we have provided a broad overview of the fundamentals of linear minimum variance estimation as well as a derivation of the Kalman, total equations we very strongly urge, the reader to be able to follow through the exercises and verify all the computations that are involved in here, I very cognizant in to the fact that there are, there are lot of facts, one need to verify, him in a class setting, I generally would cover this linear minimum variance estimation itself in about two and a half three classes giving all the details.

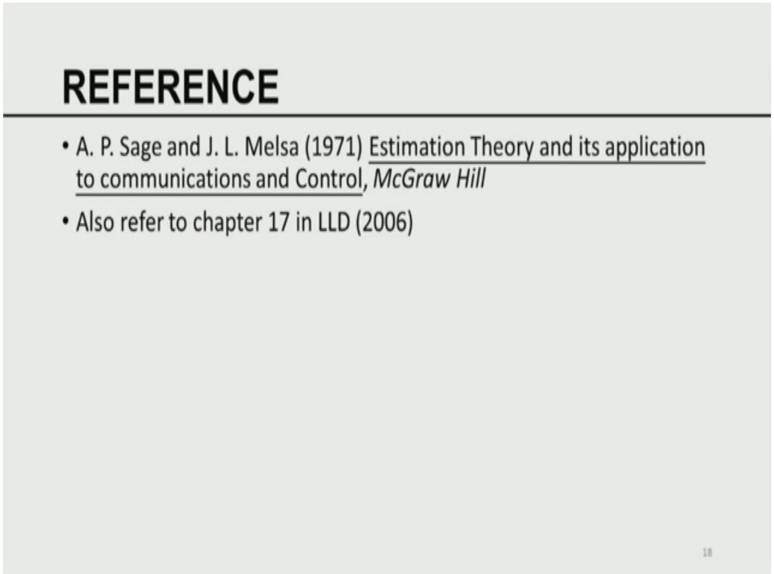
But in this compressed video form, I am giving; I am hitting all the major steps leaving behind the verification of the formulas as an exercise to you. So, please do continue, I also would like to, would like to draw your attention to a particular way of looking at what is called A 3 d var probably even though, I have not introduced them, I think it is better to anticipate that. So, I have a prior estimate and a co variance and I have a prior mean of the covariance under nominal distribution I have a prior and it is and it is covariance Z is equal to H of X plus V. I am going to consider a function H of X, which is a sum of two quadratic functions.

So, please remember there are two pieces of information one coming from prior, another coming, I am sorry, this is prior one, coming from prior one, coming from prior another coming from observations, this comes from observations. So, you can see there are two pieces of information, coming together I am joining them in the least square framework.

So, f of X is a sum of two quadratic forms, I can minimize f of X with respect to X and if you did that you will essentially get the Bayesian estimation a Bayesian results.

Therefore, you can readily see the Bayesian framework with Gaussian distribution assumptions and the 3 d var problem are essentially, one of the same. So, it is a very instructive exercise to pursue and you already know how to do the minimization of the quadratic forms. So, I very strongly urge you to be able to do the minimum and find the solution and look at the structure of the solution.

(Refer Slide Time: 68:19)



And all these derivations are further expanded in the book by sage and melsa 1971 estimation theory and application to communication and control.

Also we delve deeply into many of these in our chapter 17 in L L Ld. So, with this we have now come to the end of discussion, that relates to all the basic fundamental principles, relating to statistical estimation, starting from the properties of estimates to statistical least squares, the maximum likelihood estimates to Bayesian estimates to linear minimum variance estimation, this is simply a small sampling of results from statistical estimation theory.

The statistical estimation theory is a big ocean, I want to provide window of opportunity to be able to look at the kind of results, that statistical estimation theory provides and how some of these results are intrinsic to pursuing our goal in dynamic data assimilation.

Thank you.